ABSTRACT

Distributed solar energy is rapidly growing, and with that growth there is increasing need for tools to validate data collected on PV system configuration. Clean configuration data is the foundation for most fleet performance analytics, but large-scale deployments’ data on location, orientation, and tilt are generally subject to human error. We cover a number of techniques to automatically validate location, orientation and tilt data.

A PV system’s location can be determined via three distinct methodologies. Each method leverages the system’s production data with one approach using astronomical formulas, a second approach leveraging a network of systems with known locations, and a third approach that utilizes solar irradiance and system simulation models.

A PV system’s orientation and tilt can be determined via two distinct methodologies. Each method leverages the system’s production data with one approach using the skew of production and another approach that utilizes solar irradiance and system simulation models.

1. INTRODUCTION

Innovations in solar financing have fueled rapid growth in the distributed solar energy sector. Due to these innovations, distributed PV systems, those typically at residential or commercial sites, are increasingly owned by off-site entities. These entities are developing large portfolios of solar power systems dispersed across large geographic areas, and they generally have a strong interest in optimizing the operations and maintenance of the PV systems to ensure good performance. As distributed solar system portfolios grow in size from tens to hundreds to thousands of systems, new problems with managing these systems arise and old problems are magnified in scale.

To effectively detect and manage issues within their distributed asset portfolio, owners have turned to solar monitoring firms to track and aggregate system performance data. Individual system production data can reveal major problems with a system, such as improper system installation or inverter shutoffs, thus minimizing operations and maintenance costs by enabling owners to quickly identify and remedy the problem. In addition to providing performance and fault data, solar monitoring companies provide owners with analytics to provide a higher-level understanding of individual system and fleet performance.

While production data is valuable for detecting obvious performance issues, a more analytical approach is necessary to detect subtle problems, such as string failure or chronic underperformance. By leveraging data from a large network of monitored PV systems, regional comparative system performance diagnostics can be calculated due to the similarity of the environmental conditions driving PV production within a region. More advanced analytics can be employed that use weather stations, solar irradiance models, and PV system simulations models to calculate expected PV production in order to understand variation in system performance. These types of analytics have proven effective for understanding individual systems and entire fleets of PV systems, but require information about the PV systems’ configurations to be effectively deployed.
The primary system configuration data needed for analytics is location, orientation, and tilt, as this data determines the amount and timing of the solar irradiance received by a system. Because configuration data is manually measured and then entered through the monitoring software interface, it is subject to human errors that can be magnified in large-scale deployments. While a majority of systems in a portfolio may have correct configuration data, it is difficult to deploy analytical algorithms due to the errors caused by systems with bad data. This is particularly an issue with geospatial-based analytics. To effectively use PV system and fleet analytics, tools are necessary to validate clean configuration data, correct erroneous configuration data, and fill in missing configuration data. The required algorithmic tools and their respective methodologies are described below.

2. LOCATION IDENTIFICATION METHODS

A system’s location is the principal determinant of the amount of solar irradiance received and consequently the amount of energy produced by the system. This is because latitude and longitude determine the Sun-Earth spatial relationship. Latitude governs seasonal variation in distance from the Sun and thus the seasonal variability in solar resource received. Longitude determines when the sun will rise, peak (i.e. solar noon), and set due to Earth’s rotation. Location is of the utmost importance to PV analytical algorithms and therefore must be correct or closely approximated.

For example, relative performance diagnostics are often used in solar monitoring applications. These diagnostics compare the performance of a particular PV system to a representative set of nearby PV systems. Quality of relative performance diagnostics can quickly degrade, however, if the diagnostic algorithms have incorrect information about system location. The rapid degradation occurs because the incorrect location information affects not only the PV system with an incorrectly assigned location, but also other systems near the incorrect location. As an illustrative example, if diagnostic algorithms believe a California-based PV system is located in Colorado, they might not only compare the California-based PV system to an incorrect Colorado peer set, but also incorporate the California system into Colorado benchmark peer sets.

2.1 Astronomical Approach

As described, system longitude drives the start, peak, and end of production. Natural variation in weather conditions can influence these characteristics, so a system’s production record needs to be filtered to determine a typical production curve under good weather conditions. A typical good weather production curve allows for locationally accurate start, peak, and end production events to be identified. By treating the time of peak production as equivalent to solar noon the system’s longitude can be extracted from astronomical formulas. The implicit assumption with this approach is that the system is oriented due south or north.

However, since PV systems are not necessarily oriented due south or north, the impact of orientation bias must be accounted for to correctly estimate then longitude. Figure 1 shows the impact of orientation biasing peak production. The skew of production, or the time difference between start of production to peak production and peak production to end of production, allows for the effect of orientation to be separated from the impact of longitude on peak production. Since skew of production is almost perfectly correlated with the difference between the biased longitude and actual longitude ($R^2 > .99$), the actual longitude can be determined from the biased estimate, simply by reversing out the skew effect.

![Solar Noon Correction](image)

Fig. 1: Production curve for a PV system oriented southeast with production events labeled. The correction from biased solar noon to unbiased solar noon is also labeled.

The algorithm for approximating a system’s longitude via the astronomical approach is as follows:

1) Filter PV production data for days with clear weather conditions.
2) Identify start, peak, and end of production each filtered day.
3) Set biased solar noon each filtered day equal to time of peak production.
4) Calculate biased longitude each filtered day from biased solar noon.
5) Calculate skew of production as the time difference between start to peak and peak to end of production.
6) Calculate longitude.

Due to the variability of weather conditions, over time the algorithm determines a clearer picture of what a typical
good weather production curve looks like, allowing for the algorithm to more accurately learn the correct location. This methodology is quite robust, and with just over a month of data most PV systems can be located within ±50 miles of their true longitudinal position. Figure 2 illustrates the methodology’s learning process for 135 PV systems over 30 days in order to auto-detect longitude.

![Longitudinal Identification Accuracy](image)

Fig. 2: Mean Absolute Error over time as astronomical approach determines system longitude.

2.2 Network Approach

The environmental conditions that determine PV production are typically similar across a region, and as a result systems within this region typically have similar production behavior. This correlation typically holds on a large geographic scale, so the correlation effect can be used to find neighboring PV systems. Networks of PV systems with known locations can then be used to locate a system with an incorrect or missing location. As these networks accumulate more systems and increase regional density, this methodology becomes increasingly effective. Figure 3 shows a global network of PV systems that can be used to employ this approach.

The algorithm for approximating a system’s latitude and longitude via the network approach is as follows:

1. Calculate correlation of unknown or incorrectly located PV system’s production with all known PV systems’ production data.
2. Identify a best-fit location based on triangulating on the most correlated PV systems.

![Network of PV Systems](image)

Fig. 3: Example of a network of PV systems with known location.

This accuracy of this methodology is highly dependent upon regional network density. This means for areas with low PV penetration or no network presence the accuracy is extremely low and this methodology should not be used. However, for areas with a large network presence the methodology can frequently determine system location within 10 miles of its true location. Figure 4 shows an example of a network with high regional density where this methodology would be highly accurate.

![Regional Network of PV Systems](image)

Fig. 4: Example of a regional network with high density of PV systems.
2.3 Simulation Approach

Akin to the network approach, a simulation based methodology can be used to build a “virtual” network of systems. To create a “virtual” PV system a solar irradiance model [1], [2], is used to remotely sense the solar irradiance at a location and this estimate is then used as input to a PV system model, such as the Sandia model [3] or the single diode model [4], to create the production for the “virtual” system. The “virtual” network is built as this process is repeated for possible latitude and longitude pairs. With the network built, the correlation-based approach can then be used to locate the physical PV system.

The algorithm for approximating a system’s latitude and longitude via the simulation approach is as follows:

1. Create a grid of possible latitude and longitude pairs, and for each location:
   a. Estimate or observe environmental conditions.
   b. Simulate PV system production.
2. Calculate correlation of unknown or incorrectly located PV system’s production with all simulated systems’ production data.
3. Identify a best-fit location based on triangulating on the most correlated renewable energy systems
4. Repeat with successively tighter geographic grids of virtual networks to locate systems.

3. ORIENTATION AND TILT IDENTIFICATION METHODS

The orientation and tilt of a PV system is the secondary determinant of solar resource received by the system. Orientation is the angle, on a 360° scale from north, of which a system faces. The orientation angle typically impacts the time of peak production. Tilt is the angle, in degrees, of which a system is inclined relative to a horizontal plane. The tilt angle influences the winter and summer seasonal variation in PV production. These factors affect the angle at which sunlight reaches the PV system and consequently influence the amount of direct, diffuse, and ground reflected irradiance reaching the system. Thus to accurately model PV system production, knowing the system’s orientation and tilt is necessary.

3.1 Astronomical Approach

As was previously mentioned, the orientation of a PV system will influence the timing of peak production. The orientation angle will consistently cause peak production to deviate from solar noon, unless the system is oriented at 180° (due south) or 360° (due north). Figure 1 in section 2.1 illustrated this orientation caused deviation. Natural variation in weather can also impact the timing of peak production, so a system’s production record needs to be filtered to determine a typical production curve under good weather conditions. By determining typical start, peak, and end of production events, the skew of production can be calculated. Figure 5 illustrates how a long-term production curve can be separated into production start/peak/end in order to determine a system’s skew of production. Based on the amount of skew the PV system’s orientation angle can be computed.

![Skew Separation](image)

**Fig. 5:** A single day production curve for a PV system oriented South-East showing the skew of production. Skew calculations are in the upper right corner.

The algorithm for approximating a system’s orientation via the astronomical approach is as follows:

1. Filter PV production data for days with clear weather conditions.
2. Identify start, peak, and end of production each filtered day.
3. Calculate skew of observation as the time difference between start to peak and peak to end of production.
4. Calculate orientation.

Over time the algorithm corrects for the variability in weather conditions, and refines the typical good weather production curve, allowing the algorithm to more accurately calculate the correct system orientation. This methodology “learns” quickly, and within a month of data the orientation of most PV systems can be determined to within ±7° of their true orientation angle. Figure 6 illustrates the methodology’s improved accuracy in detecting orientation angle for 149 PV systems over 30 days.
Fig. 6: Mean Absolute Error over time as astronomical approach determines system orientation.

3.2 Simulation Approach

Building upon the logic of the location identification network approach and simulation approach, a “virtual” network of systems can be built at the location of the system with missing or incorrect orientation and tilt data. The “virtual” PV system is created via the same methodology described in the location identification simulation approach, but instead of iterating through locations to build the “virtual” network, orientation and tilt angle pairs are iterated through. With the network built, the correlation-based approach can then be used to determine the physical PV system’s orientation and tilt angles.

The algorithm for approximating a system’s orientation and tilt via the simulation approach is as follows:

1. Search through all possible orientation and tilt angle pairs, and for each pair.
   a. Estimate or observe environmental conditions.
   b. Simulate PV system production.
2. Calculate the correlation of the target system’s production with all simulated orientation and tilt pairs.
3. Identify a best-fit orientation and tilt from the most correlated simulation.

4. CONCLUSIONS AND FURTHER RESEARCH

This paper presents a number of approaches for automatically validating and correcting system location and configuration information for PV systems. These approaches can be applied to improve the overall accuracy of a large PV fleet database.

Further work is needed to fully quantify the tradeoffs and precision possible with the simulation-based approaches.

Additionally, further work is needed to combine the different approaches into a single integrated solution.

5. ACKNOWLEDGEMENTS

This work was done under funding from Locus Energy. Thanks to Asaf Peleg for assistance in data acquisition for testing methodologies. Thanks to George Kalogeropoulos for comments and feedback.

6. REFERENCES


