Geolocating Photovoltaic Systems using Geometry and Weather Data Correlation

C. Peter Naumann
Karlsruhe Institute of Technology
Karlsruhe, Germany
peter.naumann9@kit.edu

Niklas Goerke
FZI Research Center for Information Technology
Karlsruhe, Germany
goerke@fzi.de

Kaibin Bao
Karlsruhe Institute of Technology
Karlsruhe, Germany
kaibin.bao@kit.edu

Ingmar Baumgart
FZI Research Center for Information Technology and Karlsruhe Institute of Technology
Karlsruhe, Germany
baumgart@fzi.de

Veit Hagenmeyer
Karlsruhe Institute of Technology
Karlsruhe, Germany
veit.hagenmeyer@kit.edu

Abstract

With the increasing deployments of residential photovoltaic systems, successful attacks on photovoltaic inverters pose a threat to the stability of the power grid. If attackers with remote control over many inverters want to target segments of a power grid, they require knowledge of their location. To enable the development of countermeasures, we analyze how precisely inverters can be localized. We present two approaches to determine the location of an inverter, requiring only historical power generation data from each photovoltaic system. The first approach estimates times of equal solar elevation over multiple days, which uniquely identify a location through some astronomic and geometric observations. The second approach uses the correlation of daily differences in power generation with changes in local weather by measuring similarity to publicly available irradiation data. Both approaches are evaluated against real generation data from 1.428 photovoltaic systems, obtained from pvoutput.org, in a diverse set of scenarios. The first approach achieves a median spatial error of 108 km, while the second attains a median error of only 6 km. Given the broad range and rather low quality of the input data used, we achieve an accuracy that significantly exceeds existing research, even outperforming most previous results that use much higher quality input data. Multiple limiting factors are also identified, leaving opportunities for future work to further improve both approaches. The algorithms described in the present work are open-source.

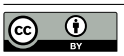
CCS Concepts

• **Security and privacy**; • **Hardware** → *Renewable energy*; • **Computing methodologies** → **Model development and analysis**;

Keywords

Photovoltaic Systems, Inverter Geolocation, Solar Elevation Fitting, Weather Data Correlation, PV Power Generation Analysis, Solar Position Algorithm, Celestial Solar Tracking, Time-Series Analysis for Geolocation, Cybersecurity, Inverter-based Resources

ACM Reference Format:
C. Peter Naumann, Niklas Goerke, Kaibin Bao, Ingmar Baumgart, and Veit Hagenmeyer. 2026. Geolocating Photovoltaic Systems using Geometry and Weather Data Correlation. In *The 17th ACM International Conference on Future and Sustainable Energy Systems (E-Energy '26), June 22–25, 2026, Banff, AB, Canada.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3744255.3811713



E-Energy '26, Banff, AB, Canada

ACM ISBN 979-8-4007-2011-6/26/06
https://doi.org/10.1145/3744255.3811713

1 Introduction

The rapid growth of grid-connected Photovoltaic (PV) system installations has transformed modern power grids. Private households switch from passively consuming electrical energy to becoming prosumers and feeding into the power grid. Central to this transformation are PV inverters as one of the most common classes of Inverter-based Resource (IBR), which convert the Direct Current (DC) generated by PV modules into Alternating Current (AC) that can be injected into the transmission or distribution grid.

Most current residential and commercial inverters are equipped with data connectivity and are connected to the local residential network, e.g. via WiFi. These connections allow home owners to connect to their PV inverter to monitor momentary and historical data, remotely control the operation, and even shut off the inverter. Power and energy monitoring is often provided via the inverter manufacturer's cloud service for convenience. This setting enables multiple attack vectors for adversaries who want to destabilize the power grid [8, 10]. An example of such an attack vector begins with an initial compromise of the manufacturer's cloud service, a third party energy monitoring service, or a home automation service. Power generation data would be readily available in the service's database. Some of these services, such as a home automation service, would allow a back-channel to control the PV system. Another exemplary attack vector uses vulnerable Internet-of-Thing devices, such as webcams, home automation devices, or routers. Infected by

botnets like Mirai [1], attackers could gain access to a large number of local networks to discover, monitor and control PV inverters.

To summarize, an attacker with control over an inverter can also access current and historical production data and control its operation [2]. This level of access poses the major threat of load-altering attacks (LAA) to power grids [6, 12]. The damage of an LAA can be amplified if the adversary knows the location of each PV system. Attacks can then be targeted towards specific grid segments, or towards multiple grid segments posed to work against each other [13]. However, inverter firmware rarely stores a geographic identifier, and conventional network-level geolocation (e.g., GeoIP) is not always reliable. This lack of explicit location information forces an attacker to infer the location of a PV inverter indirectly from the available data: the historical power-generation profile. In the present work, we deliberately restrict our analysis to only the power generation time series, discarding any auxiliary measurements and data (e.g., DC-side voltage, grid voltage, frequency, or metadata that could trivially reveal location). The research questions addressed in this work are as follows:

- How accurately can the geographic location of a PV inverter be inferred from its time series of power generation?
- How does the achievable accuracy depend on
  - the temporal resolution of the measurements, and
  - the observation window (i.e., the length of the historical record)?

This work is structured as follows: We highlight our contributions in Section 2, followed by a brief review of related work in Section 3. In Section 4, we create an abstract model for Photovoltaic systems that is relevant to our approaches. In Section 5, we describe the data we used. The data is sliced into different scenarios, which are described in Section 6. We have investigated two approaches for localization: In Section 7, we describe how PV inverters can be located using geometric observations of the sun-earth relation. In Section 8, we describe how external weather data can be used to locate PV inverters. We conclude the present work in Section 9.

2 Our Contribution

In the present work, we describe two approaches to spatially locate a PV inverter if power output data is the only data available from the inverter. We show that by estimating the time of sunrise and sunset from the power output data for at least two days, simple geometry can be used to estimate the system's location. Data for more days can be included to create a more robust estimate. In our second approach, we demonstrate that a correlation with historical weather data can result in even better location estimates. For our analysis, we use a large dataset that contains PV power output data from private households.

Each approach in this present work contains carefully engineered new compensation methods to target characteristics of the overall PV system, for which we provide a theoretical model. We are the first to evaluate how well both of these approaches work on a large set of power generation data acquired from over a thousand of real world PV systems.

This work is based on the master's thesis of the first author; more detailed descriptions of the approaches are presented there. The full thesis can be made available on request.

3 Related Work

The first work on localizing PV systems based on their power generation data was presented by Williams et al. [19]. They presented multiple approaches: Calculating the longitude from solar noon estimation, correlating the system's power generation with systems with known location, and correlating power generation with weather data by simulating system behavior. Unfortunately, results are only available for the first approach, with a minimum longitude error of $\sim 1^\circ$ (80 km in Europe) based on only a few PV systems.

Multiple authors [4, 11] have presented work on localizing a PV system based on determining the time of specific solar positions such as sunrise, sunset, and noon. The authors mainly evaluated against very few PV systems with high quality and high resolution data, making their results hard to compare and generalize.

Haghdadi et al. [7] estimated the longitude using the astronomical method but then employed a PV simulation model to estimate latitude, azimuth, and tilt. Similarly to this work, they evaluated their approach on PVOutput[1] data and reached a mean absolute error of $\sim 0.2^\circ$ (16 km) for longitude and $\sim 4^\circ$ (440 km) for latitude.

The work of Danner et al. [5] also employed a model fitting approach, using irradiation data to build a more complete model. Unfortunately, no comprehensive evaluation was performed, with validation on a single test system achieving 30 km of error. The high computational effort of the genetic optimization used is also unattractive for a large number of systems.

Chen et al. [3] and Papst et al. [14] presented work correlating power generation data with data from a network of weather stations. Both evaluated on very small sets of PV systems reaching 7 km and 20 km precision, respectively.

Most of these works were evaluated on a small number of systems. In the few cases where lower quality data is used, the accuracy achieved is significantly lower. In contrast, the approaches developed in the present work achieve high accuracy on a large number of diverse real-world systems with next to no pre-filtering.

4 Theoretical Photovoltaic System Model

By modeling the physical fundamentals of the power output of a PV system, we find a large number of influential factors. At the highest level of abstraction, the power output of a PV inverter can be described as a function mapping the time and spatial location to electrical power. We find that we can divide the physical system into three sub-systems which are each discussed in the following subsections:

- The **astronomical relation** between Earth and the Sun;
- **Terrestrial effects** on the Sun's radiation, such as atmospheric scattering, weather conditions, and shading effects;
- The properties of the **PV system** itself, e.g., the PV module orientation, the module configuration, cell type, and inverter.

4.1 Astronomical Relation

This system is well understood, and there are precise models [16] that, given a time and date as input, output a precise location of where the Sun is in zenith above the Earth to a precision of $\pm 0.0003^\circ$.

[1] https://pvoutput.org

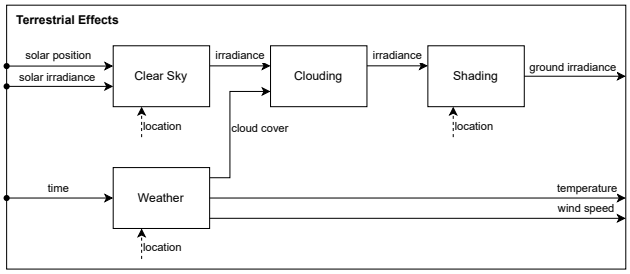


Figure 1: Information flow model that describes the physical relations between the terrestrial effects.

In 24 hours, the zenith location's longitude completes one full circle, while the latitude changes very slowly. Throughout one year, the latitude of the zenith location oscillates between ±23.4°, corresponding to the axial tilt of the Earth. Therefore, the longitude depends mainly on the current time of day, and the latitude depends mainly on the current day of year. The Sun's extraterrestrial radiation, as it hits the earth's atmosphere, also changes with time, mainly due to the eccentricity of Earth's orbit. The approximation most commonly cited [17] varies smoothly throughout the year between 1320 $\mathrm{W/m^2}$ and 1414 $\mathrm{W/m^2}$ or by $\sim$ ±3.4% [18].

4.2 Terrestrial Effects

Terrestrial effects describe the various influences on the solar radiation as it passes through the atmosphere. They depend on time, solar position, and solar intensity. In general, solar radiation at any location on Earth is a three-dimensional quantity with two angular dimensions and one spectral dimension. In practice, irradiance is given in W/m^2, disregarding the exact spectral and angular composition. The ground-level irradiance can be divided into direct and diffuse components. In practice, weather data commonly contains only the combination of both. There are three main contributors to these irradiance values, also depicted in Figure 1:

Atmospheric effects in so-called clear sky models describe the effect of air and small particulates that affect the scattering of sunlight. The details differ between models, but in general some rough location and time-dependent atmospheric data is used.

Weather effects consist mainly of clouds that obscure parts of the sky. In general, these effects are hard or impossible to model without real data at high spatial and temporal resolution, and have a possibly large effect on ground irradiance. Wind speed and air temperature are also relevant, as they influence PV cell temperature, an important aspect of cell efficiency.

Shading from objects other than clouds, e.g. mountains, can massively reduce energy outputs. Although larger objects such as mountains could be modeled, objects in close proximity to the PV modules are highly unpredictable. As shading depends only on the sun's location in the sky, the difference in influence on successive days at the same time of day is relatively small.

4.3 Photovolatic System

The Photovoltaic sub-system converts the ground irradiance into electrical power. Panel orientation in relation to the sun is the first major influence on this process. The direct irradiance scales roughly with the cosine of the angle between the sun and the panel's normal. A fraction of the incoming light is reflected by the material that covers the cell; this is also strongly dependent on the angle of incidence of the light. The process of converting light into electricity that takes place in PV cells depends on their temperature, where cooler cells are more efficient; cell temperature is dependent on ambient temperature and wind speed. Furthermore, the technical details of the inverter, such as the number of individual trackers and the quality of its Maximum Power Point Tracking (MPPT) algorithm influence the electrical energy that can be extracted from a cell array.

5 Data Sources

The development and evaluation of the methods developed in this work is based on data from **pvoutput!** (**pvoutput!**). To ensure high quality for relevant data, the 110,000 systems on **pvoutput!** were filtered by the following criteria:

- High data availability for the first 168 days of 2025[2] (see Section 6);
- The location of the system is given and is within ±65° latitude and longitude[3];
- The maximum resolution of **pvoutput!** is given as one data point every 5 minutes; systems with artificially increased frequency (repetitive data) were omitted;
- Systems not rounding their power output & systems with less than 20 different power values per day were omitted.

An analysis of the 3, 321 remaining systems showed that the distribution of these systems is heavily biased. There is a general bias towards Europe, as well as a bias within Europe towards Belgium, the Netherlands, the United Kingdom, and Italy, where more than half of the systems are located. To reduce the influence of this bias on our final results, 90% in the Netherlands, 70% in Belgium, 50% in the United Kingdom and Italy were omitted. In the end, 1, 428 systems were used for the evaluation of both approaches. Their locations are shown in Figure 2.

For the weather fitting approach presented in Section 8, irradiation data is required. The European weather satellite operator European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) provides historical irradiance data at a resolution of 5 km × 5 km × 30 min with its SARAH-3 data set [15]. This data was used and did not require any pre-filtering, though data for a few points in time were missing.

6 Scenarios

In order to compare the strengths and weaknesses of both methods presented below, the same set of scenarios is used. A scenario is meant to represent a situation of data availability that is plausible for a PV system.

For simplicity, the year was divided into 13 28-day months, data was available for the 168 days from January 1st to June 18th of 2025 (6 months). Both methods require at least a few days to be applicable, so the smallest span chosen was two weeks. Based on this, the following time intervals were chosen.

[2] missing at most 12 Days; this is equivalent to 93%
[3] These values were chosen because the data used for weather fitting was only available for this region; see Section 8

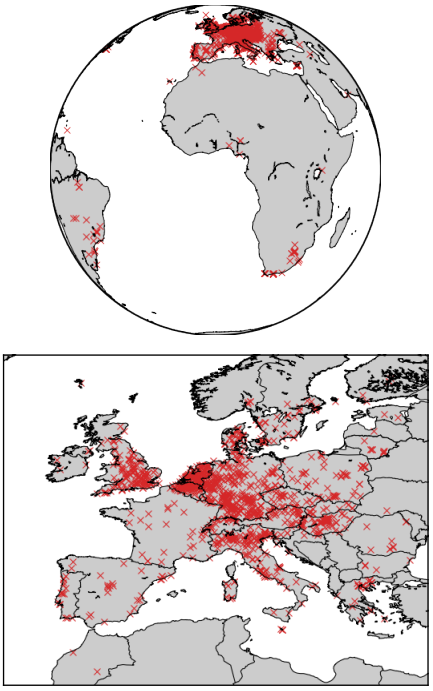


Figure 2: Map of PV systems used for evaluation. Satellite view of all systems on top, close-up of European area below.

(1) Two weeks (14 days),
(2) One month (28 days),
(3) Three months ($28 * 3 = 84$ days),
(4) Six months ($28 * 6 = 168$ days).

The time of year is relevant to potentially capture seasonal effects. Overlap between scenarios was avoided as it could introduce unwanted correlations. For the 168 days available, the following scenarios were chosen.

(1) 12 two-week scenarios,
(2) 6 one-month scenarios,
(3) 2 three-month scenarios,
(4) 1 six-month scenario.

The available date is given at a maximum of one data point per 5 minutes, giving the upper bound. Data with lower than daily resolution was not considered as the expected information at this point is very low. The chosen resolutions are:

(1) 5 minutes,
(2) Hourly,
(3) Daily.

7 Solar Elevation Fitting

This approach is based on geometric observations of the Sun-Earth system. At any point in time, the imaginary line from the center of the Earth to the center of the Sun intersects the Earth's crust at one specific location. This "zenith location" can be easily computed given the date and time. If an observer measures multiple solar positions that share the same elevation (angle above the horizon), those positions lie on a circular path centered around the observer's location. The central axis of the circle then points towards the observer's location.

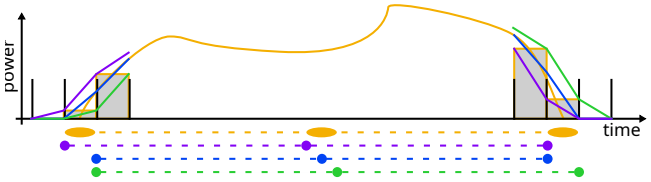


Figure 3: Solar Elevation Fitting – Possible biases depending on sampling strategy: Momentary sampling (blue) preserves expected midday but shortens duration. Forwards (purple) average and backwards (green) average introduce midday bias but preserve expected duration.

Our naïve assumption is that a PV inverter will deliver the same amount of power whenever the Sun is at the same elevation from the PV inverter's point of view. For this approach, the time of the first and last power output was chosen. The rationale is that during the morning and evening, the effects of panel orientation and clouding are the weakest because the irradiance is dominated by diffuse light. Inverters usually have a preset cutoff power, which means that cutoff times should correspond to equal panel irradiation. Knowing the time of first and last power output for a specific inverter, we can calculate the Sun's zenith location. Including data from multiple days, we receive multiple zenith locations that we can use to fit a geometrical plane that forms a circle in the intersection with the earths crust. As described above, the central axis of the circle then intersects the crust of the Earth in two locations. These can be trivially disambiguated by performing a rough longitude estimate using the mean of sunrise and sunset to find the location of the PV inverter.

This approach is strongly dependent on the assumption that the times at which the PV inverter switches the power on and off correspond to times of equal sun elevation. Depending on how the power data is sampled, it might be necessary to apply a bias correction; see Figure 3.

7.1 Refinements

The following steps are guided by the results of the 6-month scenario at 5 minute resolution. A graphical representation of the improvements can be seen in Figure 4.

Using this basic algorithm, one can observe a significant northward bias of about 150 km. Most outliers are created when the irradiation is too small for the inverter to turn on at low sun elevations, causing very late sunrise and very early sunset estimates. As most of the PV systems used for this thesis are situated in the northern Hemisphere, generally weaker irradiation in winter increases the likelihood of such outliers, leading to the northward bias. By iteratively refitting with only 90% of the points closest to the plane found previously, this bias can be reduced to about 40 km after a few iterations.

Another significant variation is between sunrise and sunset, where the estimated sunset time is at a lower elevation than the corresponding sunrise. The effect is a general westward bias of

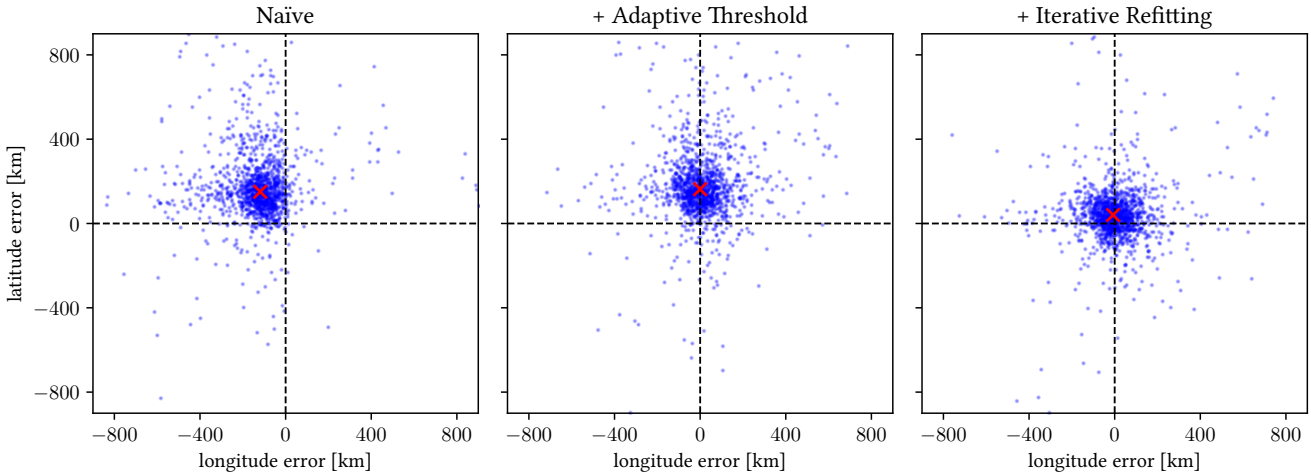


Figure 4: Solar Elevation Fitting – Improvements in estimation error from the naïve approach (left) to the final approach with both refinements (right). The red cross marks the point at median latitude and longitude error. The view is cropped to better show the resulting changes, 6% of systems fall outside the view area.

around 100 km. A possible source for this inaccuracy is the behavior of some PV inverters. At sunset, as the irradiation decreases, the inverter eventually reaches its lower operating limit and turns off. At the next sunrise, the inverter will turn on only after it can output AC voltage at a threshold above its lower operating limit, possibly to avoid rapid switching. An example of this behavior can be found in Figure 5. Important to note is that the power output at sunrise and sunset is nearly symmetric, and it is only the on-duration that is extended. The effect can therefore be compensated for by setting the sunset threshold to be the turn-on power at sunrise.

7.2 Results

The results shown are obtained with the adaptive threshold and 5 iterations of refinement as described above. Only the scenarios with 5-minute and 1-hour resolution are included as the 1-day resolution scenario is not applicable.

Table 1 shows the results of this approach for the scenarios of 2 weeks, 1 month, 3 months, and 6 months. Notable is the large difference between the errors for latitude and longitude, especially for shorter scenarios. This difference is also visualized in Figure 7. The Cumulative Distribution Function (CDF) for the same set of results can be found in Figure 6. Here, a plot per resolution shows the changes with increasing scenario length.

As can be expected, the error generally decreases for increasing scenario length and higher data resolution.

7.3 Interpretation of Results

The longitudinal error at 5-minute resolution of $\sim$ 60 km is close to the $\sim$ 45 km the Earth rotates in half a sample interval at European latitudes, where most systems are located. The data resolution might therefore be the limiting factor of this approach. This is also supported by the fast convergence to this accuracy level. Interestingly, the 1-hour intervals do not show an error that is $\sim$ 12 times

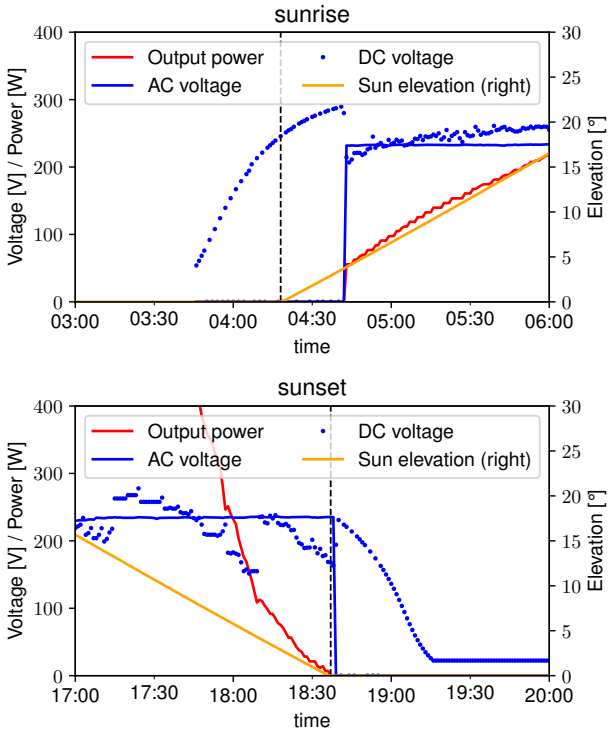


Figure 5: Solar Elevation Fitting – Example of an inverter delaying power output at sunrise. The DC voltage at sunrise and sunset is approximately equal. The power output, however, is delayed by roughly 25min.

Table 1: Solar Elevation Fitting Error Quantiles [km]. Results for scenarios with the same length but differing start times are combined into a single set.

		2 weeks			1 month			3 months			6 months		
		abs	lat	lon	abs	lat	lon	abs	lat	lon	abs	lat	lon
1h	0.5q	4149	4136	125.1	3173	3160	91.8	886.5	861.3	75.2	181.1	141.0	70.0
	0.9q	15129	15084	683.0	5430	5364	398.7	1571	1516	277.1	501.7	384.3	251.2
5min	0.5q	3739	3727	75.1	1462	1435	67.8	288.1	245.3	61.7	108.1	64.5	61.5
	0.9q	14708	14684	526.2	4469	4408	323.4	941.5	752.3	266.9	411.7	270.9	256.5

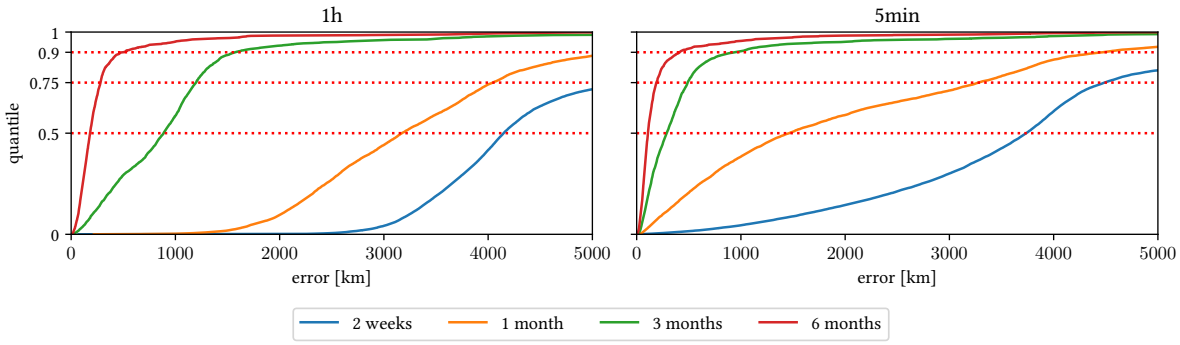


Figure 6: Solar Elevation Fitting – CDF for the evaluated resolutions and scenario lengths. Results for scenarios with the same length but differing start times are combined into a single set.

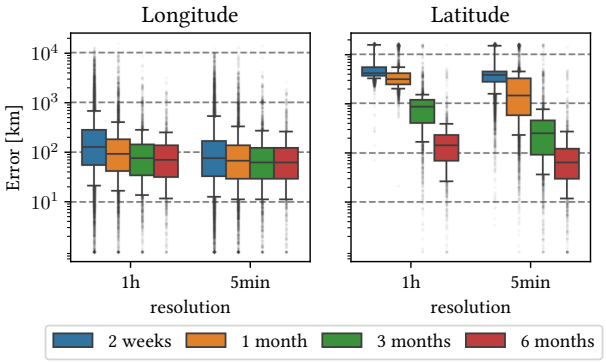


Figure 7: Solar Elevation Fitting – Overview of the estimation error for all test runs, split into latitude and longitude. Results for scenarios with the same length but differing start times are combined into a single set. The whiskers are set to the 10th and 90th percentile. Note the logarithmic scale on the y-axis.

worse. This could be explained by the fact that the 1-hour samples were obtained by averaging 12 5-minute samples instead of a true hourly average, which could have caused a similar residual error in both cases.

The main driver behind the large difference between latitudinal and longitudinal error is the movement characteristics of the zenith location. In the case of longitude, a single day provides two samples at a nearly maximum distance, resulting in a near-optimal fit; including multiple days eliminates the system noise. However, latitude changes slowly throughout the seasons, thus requiring half a year to obtain two points at the optimal distance, which itself is only $\sim 48°$ compared to the nearly $180°$ available for longitude. This also helps to explain the extreme (northward) bias for the 2-week scenarios.

A more general issue with this approach is its reliance on the low-power behavior of PV inverters. A regular user has little interest in power generation at these times as the power output is very low. It is therefore possible for an inverter to behave unexpectedly in these cases, by accident or by design.

7.4 Unexplored Refinements

During winter, when temperatures are lower, PV cells usually have higher efficiency. This might cause lower solar elevation estimations in winter, which in turn could increase the error of this approach.

Another more subtle issue might lie with the assumption of the Earth being a sphere. This assumption is broken by the position calculations using the standard WGS84 coordinate system, which deviates from a sphere by 22 km out of 6367 km or $\approx 0.35\%$ This could be an issue as the mapping between the Sun's position in the sky and the zenith location is no longer linear. Although compensating for this effect is certainly possible, the expected gains are small.

In general, a more robust plane fitting algorithm or better outlier rejection provide additional avenues for improvement. Also, inter-day changes in sunset and sunrise times are not random, a relation

that could be used as an intelligent pre-filter similar to Londono et al. [11].

8 Weather Fitting

Weather is closely related to PV plant power output and is also highly location dependent, making it an obvious candidate for geolocation. The main observation is that, excluding weather, all other effects depend almost exclusively on the position of the Sun, i.e., the current time. The main assumption now is that for the same time of day at two consecutive days, the change in solar position, temperature, and wind speed is negligible compared to the change in solar irradiance. This assumption does not need to hold for all such time pairs as long as the number of outliers is small. The second assumption is that the PV system's response to changes in irradiation is approximately linear. A holistic argument for this to be true is that the main goal of a PV system is to keep the cells as close to their optimal efficiency as possible, which is a linear property.

Under these assumptions, an inter-day pair of solar irradiance I_t, I_{t-1d} and power output P_t, P_{t-1d} obey the linear relationship

$$\frac{I_t}{I_{t-1d}} = \frac{P_t}{P_{t-1d}}.$$

For a single point in time, this might be true for multiple locations. However, with increasing sample size, the probability that two locations have the same irradiation profiles decreases rapidly.

This simple fraction is not well suited for noisy environments, especially as a few large incorrect values can skew the estimate because of their large contribution. This can be mitigated by using the normalized difference

$$\Delta X_t := \frac{X_t - X_{t-1d}}{|X_t| + |X_{t-1d}|}.$$

This formulation still obeys the relation that under ideal conditions $\Delta I_t = \Delta P_t$ at the system's location, with some beneficial properties: ΔX_t is always bounded by the interval $[-1, 1]$, making aggregation much easier. It is also well conditioned, that is, noise in either of the inputs is never amplified.

In every realistic case, no location will be an exact match, necessitating some error aggregation scheme. The problem can be stated as follows: Given are two time discrete functions, $dI(t, p) := \Delta I_t$ (at p) and $dP(t) := \Delta P_t$ for some set of locations p. Find the value of p for which the functions $dI(t, p)$ and $dP(t)$ are the most similar. This location of maximum similarity is the final location estimate of this approach.

Additionally, irradiation data is available only as momentary samples for every 30 minutes. We downsample our power generation data by averaging the three closest power samples for each irradiation sample.

8.1 Similarity Measures

A core part of this approach is how the similarity between $dI(t, p)$ and $dP(t)$ is calculated. Functions of this kind are known as similarity measures, and there is a wide variety of common measures available in the existing literature [9]. All measures are built to produce a result from 0 to 1, with 1 corresponding to a perfect match. Of those presented by Levy et al. [9], the Inner Product and

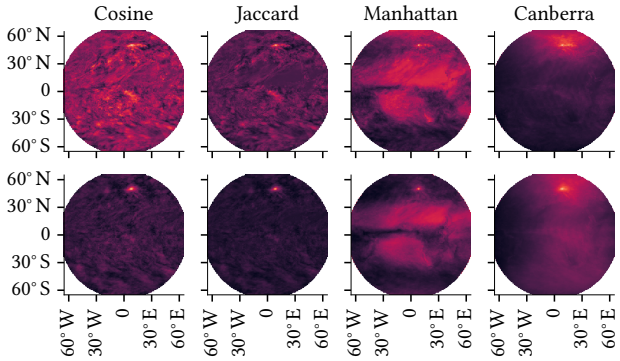


Figure 8: Weather Fitting – Sample of the base similarity measures for a single system. The first row shows the result for 4 weeks of data, the second row for 6 months. The heat maps show the similarity for all locations covered by the available weather data. Each plot is normalized separately.

Minkowsky families are applicable to this approach. From these, the following base measures were selected:

$$Sim_{cos}(I, P) = \frac{\langle I, P \rangle}{\|I\| \|P\|} \quad \text{(Cosine)}$$

$$Sim_{jac}(I, P) = \frac{\langle I, P \rangle}{\|I\| + \|P\| - \langle I, P \rangle} \quad \text{(Jaccard)}$$

$$Sim_{L1}(I, P) = \left(1 + \sum_i |I_i - P_i|\right)^{-1} \quad \text{(Manhattan)}$$

$$Sim_{can}(I, P) = \left(1 + \sum_i \frac{|I_i - P_i|}{|I_i| + |P_i|}\right)^{-1} \quad \text{(Canberra)}$$

For a test system, using any of the described measures, one can find a global maximum at the correct location. However, the measures vary widely in their characteristics for the remaining (false positive) locations, as can be seen in Figure 8.

The different characteristics of the base measures shown allow for a simple optimization. The true location tends to score high on any metric, with false positives varying between them. When multiplying two measures with different characteristics point-wise, the non-overlapping noise should vanish, while the true location remains. Upon evaluation of various combinations of these metrics, the most consistent measure was found to be the Cosine-Canberra combination, which is the main measure used in the evaluation.

$$Sim_{cos,can}(I, P) = Sim_{cos}(I, P) * Sim_{can}(I, P)$$

Details of this evaluation are described in the masters' thesis on which this work is based, see Section 2.

8.2 Results

An overview of the 50th and 90th percentile of estimation errors can be found in Table 2. An overview of the absolute estimation error is also shown in Figure 9. Due to the fact that this approach does not depend on the different characteristics of latitude and longitude, the errors for both are very similar and thus omitted

Table 2: Weather Fitting Error Quantiles [km]

		2 weeks	1 month	3 months	6 months
1d	0.5q	50.7	20.0	10.0	7.4
	0.9q	8150	3043	34.3	20.6
1h	0.5q	12.8	8.9	6.7	6.0
	0.9q	328.3	40.6	18.1	13.3
30min	0.5q	20.0	13.3	11.1	9.9
	0.9q	946.4	55.0	24.3	19.4

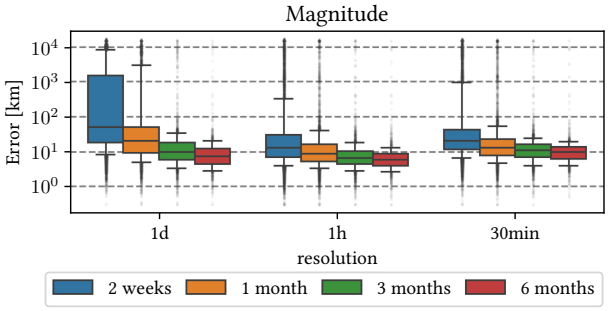


Figure 9: Weather Fitting – Overview of the estimation error for all test runs. Results for scenarios with the same length but differing start times are combined into a single set. The whiskers are set to the 10th and 90th percentile. Note the logarithmic scale on the y-axis.

here. The CDF for the same set of results can be found in Figure 10. Here, one plot per resolution shows the changes with increasing scenario length.

8.3 Interpretation of Results

The general similarity between latitude and longitude is expected as the method operates on an equal-resolution grid with no clear dependency on latitude and longitude. For the same reason, a small bias of 3 km appearing along the longitudinal axis is unexpected, with no satisfactory explanation available at the time of writing.

In case of 6 months of data, the resulting median error is, when compensating for the longitude bias, near or even below the grid resolution. Therefore, no significant improvement can be expected with more data, as even with 6 months most estimates are in the correct or at least a neighboring 5 km × 5 km grid cell. Higher resolution weather grids, however, may allow for better estimates even for similar time spans.

Interestingly, the optimal temporal resolution seems to be at 1 hour, consistently outperforming the 30-minute scenarios. This inferior performance for 30-minute data can be attributed to the quality of the input data. For very noisy input data, either due to measurement error or highly variable shading, the 1-hour down-sampling acts as a noise-reduction filter. This improves the efficiency of the following similarity matching, resulting in a lower error.

The accuracy for 1-day resolution is surprisingly high, especially at longer scenario lengths. Comparing the number of data points, 6-months of daily data has fewer samples than even one month of hourly data. This suggests that, when normalized differences are taken, a mean value over a longer time span contains more information.

Also of note is the stark contrast between the median and 90th percentile, with the 90th percentile being far worse for short scenario lengths.

9 Conclusion and Future Work

We present two approaches for spatially localizing PV systems based on power output data.

The first approach, solar elevation fitting, uses power cutoffs near sunrise and sunset to find times of equal solar elevation. Over multiple days, the solar position at these times can be used to fit a circle with the inverter location estimate at its center. It is applicable globally without any external data and, given its low complexity, could be useful as a rough filter to guide other approaches.

The second approach is based on the comparison of day-to-day differences in power generation with day-to-day differences in solar irradiance. The location for which this difference is the smallest is the estimated location of the system. It is dependent on high-resolution irradiance data and allows us to locate 50% of the systems to within 20 km given only 1 daily sample for 28 consecutive days.

Both approaches are evaluated on power generation data sourced from **pvoutput!**, a third party website where users voluntarily upload their PV system's data. Multiple scenarios are used in the evaluation, varying in start time, length, and temporal resolution.

The first approach reaches a median accuracy of 108 km, with a 90th percentile of 412 km. The second approach performs much better, reaching a median and 90th percentile of 6.0 km and 13.3 km, respectively. As can be expected, the accuracy for both approaches improves with longer time spans of input data and, lesser so, with higher sampling frequency.

Coming back to our motivation, this shows that the second approach could be used by an attacker to localize compromised PV inverters to a high-enough accuracy to target the power grid of individual cities.

In the future, additional evaluation with longer data time spans, more diverse resolution selection, and higher data quality could provide further insight into the accuracy of the approaches presented. Similarly, evaluating the weather fitting approach against other weather datasets, possibly considering temperature as an influence on PV cell efficiency, could yield further results. Re-evaluation with higher resolution irradiance data, either by upsampling existing data or finding higher resolution sources, may produce more accurate estimates.

Improving location estimates by other means should also be investigated, as many options have not been explored yet. For example, the weather fitting approach uses a simple maximum for the final estimation step. A more sophisticated method that includes neighboring data-cells might be able to achieve higher, possibly sub-cell-size, accuracy. The estimation speed and success rate might also be improved by estimating at multiple temporal and spatial resolutions. Comparison against satellite images to find exact locations, with disambiguation via system attributes such as peak power or orientation, is also worth investigating. Developing a

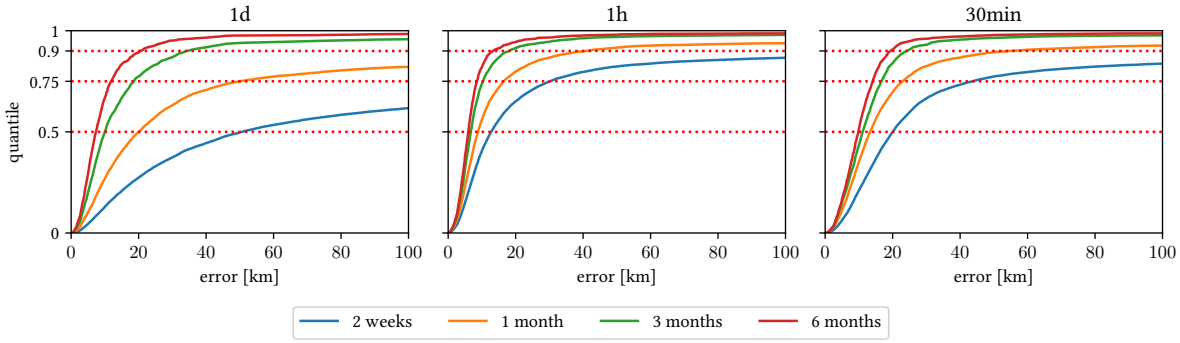


Figure 10: Weather Fitting – CDF for the different resolutions and scenario lengths. Individual scenarios for each length are stacked together.

confidence measure for the estimated locations will also be required for both approaches to be useful in most cases.

Finally, we provide the code for both approaches online[4].

Acknowledgments

This work was supported by the Helmholtz Association under the programs "Energy System Design (ESD)" (topic number 37.12.01) and "Engineering Digital Futures (EDF)" as part of the KASTEL Security Research Labs, Karlsruhe (topic number 46.23.02).

References

[1] Manos Antonakakis, Tim April, Michael Bailey, Matt Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J. Alex Halderman, Luca Invernizzi, Michalis Kallitsis, Deepak Kumar, Chaz Lever, Zane Ma, Joshua Mason, Damian Menscher, Chad Seaman, Nick Sullivan, Kurt Thomas, and Yi Zhou. 2017. Understanding the Mirai Botnet. In *26th USENIX Security Symposium (USENIX Security 17)*. USENIX Association, Vancouver, BC, 1093–1110. https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/antonakakis

[2] Kaibin Bao, Sid Chi-Kin Chau, Ghada Elbez, Qi Liu, and Veit Hagenmeyer. 2025. Advanced Persistent Threats on Consumer Energy Resources in Decentralized Energy Systems. In *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*. ACM, Rotterdam Netherlands, 838–845. doi:10.1145/3679240.3734653

[3] Dong Chen and David Irwin. 2017. Weatherman: Exposing weather-based privacy threats in big energy data. In *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, Boston, MA, USA, 1079–1086. doi:10.1109/BigData.2017.8258032

[4] Dong Chen, Srinivasan Iyengar, David Irwin, and Prashant Shenoy. 2016. SunSpot: Exposing the Location of Anonymous Solar-powered Homes. In *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments (BuildSys '16)*. Association for Computing Machinery, New York, NY, USA, 85–94. doi:10.1145/2993422.2993573

[5] Philipp Danner and Hermann de Meer. 2021. Location and solar system parameter extraction from power measurement time series. *Energy Informatics* 4, 3 (Sept. 2021), 14. doi:10.1186/s42162-021-00176-2

[6] Niklas Goerke, Alexandra Märtz, and Ingmar Baumgart. 2024. Who Controls Your Power Grid? On the Impact of Misdirected Distributed Energy Resources on Grid Stability. In *The 15th ACM International Conference on Future and Sustainable Energy Systems*. ACM, Singapore Singapore, 46–54. doi:10.1145/3632775.3661943

[7] Navid Haghdadi, Jessie Copper, Anna Bruce, and Iain MacGill. 2017. A method to estimate the location and orientation of distributed photovoltaic systems from their generation output data. *Renewable Energy* 108 (Aug. 2017), 390–400. doi:10.1016/j.renene.2017.02.080

[8] Xiangyu Hui, Samuel Karumba, Sid Chi-Kin Chau, and Mohiuddin Ahmed. 2025. Destabilizing Power Grid and Energy Market by Cyberattacks on Smart Inverters. In *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems (E-Energy '25)*. Association for Computing Machinery, New York, NY, USA, 136–151. doi:10.1145/3679240.3734613

[9] Avivit Levy, B. Riva Shalom, and Michal Chalamish. 2025. A guide to similarity measures and their data science applications. *Journal of Big Data* 12, 1 (Dec. 2025), 188. doi:10.1186/s40537-025-01227-1

[10] Yuanliang Li and Jun Yan. 2023. Cybersecurity of Smart Inverters in the Smart Grid: A Survey. *IEEE Transactions on Power Electronics* 38, 2 (Feb. 2023), 2364–2383. doi:10.1109/TPEL.2022.3206239

[11] Alejandro Londono-Hurtado, Bennet Meyers, Elpiniki Apostolaki, and Robert Flottemesch. 2021. Estimation of Photovoltaic System Location and Orientation from Power Signals. In *2021 IEEE 48th Photovoltaic Specialists Conference (PVSC)*. IEEE, Fort Lauderdale, FL, USA, 1807–1812. doi:10.1109/PVSC43889.2021.9518783

[12] Sajjad Maleki, Shijie Pan, Subhash Lakshminarayana, and Charalambos Konstantinou. 2025. Survey of Load-Altering Attacks Against Power Grids: Attack Impact, Detection, and Mitigation. *IEEE Open Access Journal of Power and Energy* 12 (2025), 220–234. doi:10.1109/OAJPE.2025.3562052

[13] Ehsan Naderi, Samaneh Pazouki, and Arash Asrari. 2023. A Coordinated Cyberattack Targeting Load Centers and Renewable Distributed Energy Resources for Undervoltage/Overvoltage in the Most Vulnerable Regions of a Modern Distribution System. *Sustainable Cities and Society* 88 (Jan. 2023), 104276. doi:10.1016/j.scs.2022.104276

[14] Franz Papst, Naomi Stricker, Rahim Entezari, and Olga Saukh. 2022. To Share or Not to Share: On Location Privacy in IoT Sensor Data. In *2022 IEEE/ACM Seventh International Conference on Internet-of-Things Design and Implementation (IoTDI)*. IEEE, Milano, Italy, 128–140. doi:10.1109/IoTDI54339.2022.00015

[15] Uwe Pfeifroth, Steffen Kothe, Jaqueline Drücke, Jörg Trentmann, Marc Schröder, Nathalie Selbach, and Rainer Hollmann. 2023. Surface Radiation Data Set - Heliosat (SARAH) - Edition 3. doi:10.5676/EUM_SAF_CM/SARAH/V003 Version Number: 3.0.

[16] Ibrahim Reda and Afshin Andreas. 2004. Solar position algorithm for solar radiation applications. *Solar Energy* 76, 5 (Jan. 2004), 577–589. doi:10.1016/j.solener.2003.12.003

[17] J. W. Spencer. 1971. Fourier series representation of the position of the sun. *Search* 2, 5 (May 1971), 172+. http://www.mail-archive.com/sundial@uni-koeln.de/msg01050.html

[18] Joshua S. Stein, Clifford W. Hansen, and Matthew J. Reno. 2012. *Global Horizontal Irradiance Clear Sky Models: Implementation and Analysis*. Technical Report SAND–2012-2389. Sandia National Laboratories (SNL-NM), Albuquerque, NM (United States). doi:10.2172/1039404

[19] Matthew K Williams, Shawn L Kerrigan, Alexander Thornton, and Locus Energy. 2012. Automatic detection of PV system configuration. In *World Renewable Energy Forum*. American Solar Energy Society Denver, Colorado, Denver, Colorado, 1933–1937. http://documents.locusenergy.com/automatic-detection-of-pv-system-configuration-wref-reproducible.pdf

[4] https://codeberg.org/naumann-kit/cloud-nein.git