Unlocking insights with PV disaggregation: A tour under the hood for energy professionals

Ebook

Estimating Solar Production from Exported Energy

Starting with the disparity of energy management between prosumers and consumers, this ebook outlines the challenges with solar PV energy insights and how PV Disaggregation, Eliq's machine learning capability, addresses it and overcomes it.

What to expect from this ebook:

Brief background of the solar PV insights challenge and the reasons behind it and a complete tour under the hood of PV Disaggregation: What data is necessary, what challenges the data presents, and how Eliq's team solved them and processed the data to present the highest possible quality of insights for prosumers.

Strong awareness and understanding of solar PV production and consumption empower prosumers to make informed decisions and optimise solar energy production. At this point in time, prosumers don't have access to the same level of detailed insights as energy consumers. Eliq's PV disaggregation was developed to address this disparity and enable prosumers to have an equally rich experience with their energy data. PV Disaggregation paints a complete, data-driven picture of PV system performance and the individual parameters that influence it. Exploring the data requirements, analysing the data and powering advanced algorithms and artificial intelligence (AI) can unlock the full potential of PV disaggregation for accurate estimations.

Enabling Insights for PV Prosumers

Nowadays, most PV installations are of the "behind the meter" type. This means that the prosumer cannot tell apart how much energy they produced and how much of the energy their PV produced they consumed. What prosumers can see with "behind the meter" installations is only the energy exported to the grid.

Energy providers can overcome this blocker with Eliq's PV Disaggregation – a new technology that uses standard import/export data from regular utility net meters to determine and categorise the electricity produced and consumed. PV Disaggregation is tailored to prosumers and adds a pool of key insights, such as self-consumption rate. The feature is simply an extension of Eliq's Energy insights, a functionality accessible to everyone using apps powered by Eliq. Without PV Disaggregation, the energy insight puzzle is incomplete as solar energy producers are left without insights!

The PV System

There are no PV Insights without PV data – but there is no data without a PV system. So let's take it from the top and outline the physical system at hand. PV systems have 5 main components in this system. Spoiler alert: The panels themselves are just one component!

1. PV installation: Solar panels that produce energy. The installation is defined by:
    a. Azimuth
    b. Tilt
    c. Size and Efficiency
    d. Optimal temperature
2. Inverter: The device that transforms direct current electricity (DC) to alternating current (AC) electricity to make it suitable for use in location and by the electrical grid.
3. Electricity Meter: The device that measures how much electricity is imported from the grid and how much of it is exported back to the grid.
4. The Grid: The infrastructure that provides electricity.
5. The property and all electrical appliances that consume electricity.

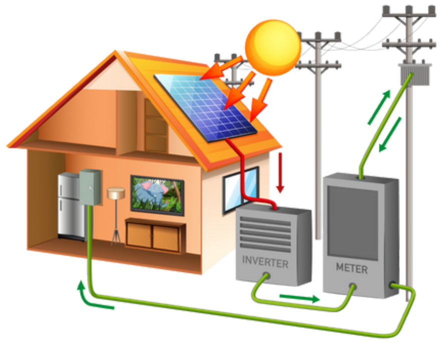


Image by brgfx on Freepik

Let's expand on the PV installation parameters. It is important to understand them, since they determine the behaviour of the PV system. There are four PV parameters in our model, but the major impact comes from the top three: azimuth, tilt and size efficiency.

The **azimuth** is the direction the PV installation is facing and it is measured in degrees from North. For example, if the PV installation is facing directly South, the azimuth will be 180 degrees. Azimuth determines the peak time of energy production – if the installation is facing directly south, the peak of the energy production will be around noon.

The **tilt** is the slope by which a PV installation faces the sky and it is measured in degrees from the ground. For example, if the PV installation is completely flat the tilt is 0 degrees; if the PV installation is vertical the tilt is 90 degrees. The optimal tilt is close to or equal to the latitude degree of your home. However, the tilt should also account for the angle of the roof.

Size and efficiency are a combination of attributes that have a multiplicative effect on energy production. Simply put, the larger the installation and the higher the efficiency, the more energy will be produced and the higher the peak wattage will be.

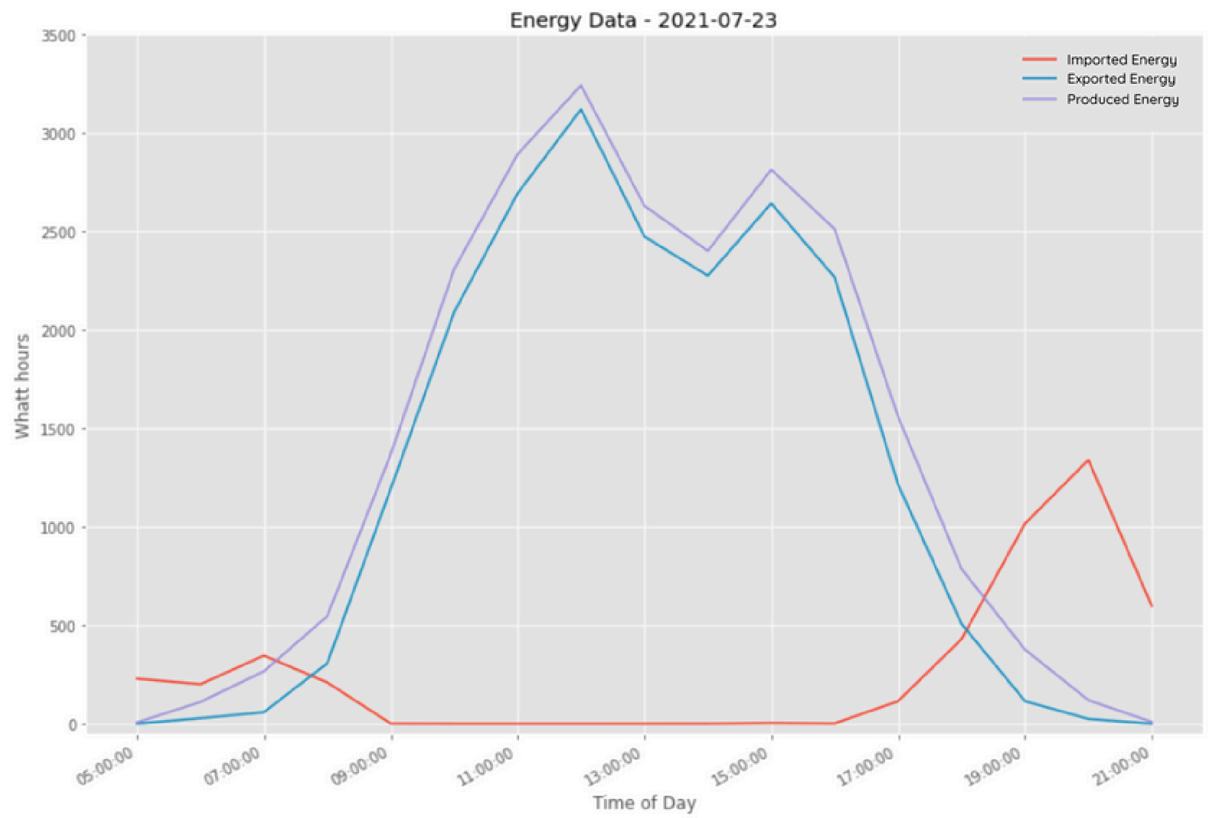
The last parameter, **optimal temperature**, is still important for data estimation, but not to the extent of the other parameters. Every solar panel has its optimal operating temperature but high or low temperature is data input required to determine the energy production.

Speaking of data input, after learning the ABCs of a PV system, let's outline what data is required for PV disaggregation.

Data Input for PV Disaggregation

There are two types of data required for PV disaggregation: **Energy** data and **weather** data.

Energy data encompasses both imported and exported data, which is available in some countries via open energy data hubs populated by data straight from the energy providers. However, data from the hubs are complementary to the actual solar production data. On the bright side, the solar production data doesn't have to be all-encompassing. Of course, it's better if it is but it is sufficient if it comes from a subset of locations.



This graph delineates complete energy data from a PV system in a single day – energy produced, exported, and imported. Notice the gap between energy production (in purple) and the exported energy (in blue): It is rather consistent throughout the day. Meanwhile, during this time, the imported energy closes in on zero. This translates to high production throughout the daytime. Imported energy rises in the evening as the property still consumes energy but doesn't produce it and then dips again at night when the property assumingly consumes much less energy.

The dip in energy produced in the early afternoon is a result of the weather conditions – this is why weather data needs to be taken into account. This weather data is consistent data input in Eliq's PV Disaggregation. PV Disaggregation fetches a variety of weather parameters but the most critical ones are cloud coverage, temperature (windchill) and humidity. These 3 parameters either obscure sunlight or substantially change the atmospheric conditions (moisture in the atmosphere that reflects light) thus impacting production. Additionally, the season and time of day are helpful data inputs that explain variations in produced energy.

We have covered the PV system and the data required to make an estimate of PV production. It is time to learn how we combine this knowledge to produce our solar production estimate. Let's look into the algorithm.

The Algorithm that moves PV Disaggregation

Solar energy estimation is a broad research domain. As previously mentioned, there are several blockers for energy providers, hence researchers dive into finding effective ways of estimating energy production. This 2020 study by Chen & Ardakanian outlines and categorises a spectrum of methods to estimate energy production.

Eliq's PV disaggregation uses the SunDance algorithm from D. Chen and D. Irwin, (2017). The SunDance approach is highly appropriate for energy estimation in the real world, as it requires only imported/exported energy data and weather data from a small set of locations with actual solar production.

This is how the algorithm crunches the data into actionable, personalised insights:

1. PV Parameters for each location are estimated using imported or exported energy data. The solar energy conversion coefficient is calculated for the location, using the location-specific dataset, weather parameters, and other factors.
2. A machine learning model is trained based on the calculations above. Note: When trained with a substantial dataset, this step does not have to be done frequently as ML models won't change a lot.
3. The maximum solar curve is produced for each location with estimated PV parameters. The machine learning model is then applied to output the final solar energy production estimate. This is the part that mostly runs in the production environment.

Estimating PV Installation Parameters

The most resource-intensive part of the whole process is the PV system parameter estimation. This happens in three steps:

1. Generating the total available solar energy (under ideal weather conditions) at a particular location with a particular set of values of PV parameters.
2. Calculating the forecasted accuracy metric (Mean Average Percentage Error, Median Absolute Percentage Error or any other relevant metric) for the particular PV parameter set.
3. Finally, selecting the PV parameter set with the lowest probability of error.

Steps 1 and 2 are continuously iterated until all the available parameter combinations are processed and all the corresponding metrics are calculated. Only after this is complete, PV Disaggregation moves to Step 3, selecting the parameter combination with the lowest error probability.

The Challenges

This iteration is the most resource-intensive part for two reasons:

- The combinations of parameters that need to be evaluated are many.
- The data involved in each combination is plentiful.

These two factors in themselves present additional challenges.

The first challenge is **locating the multiple parameters necessary for the evaluation**. This is due to the four parameters that define a PV system. These parameters interact with each other in non-linear ways, making optimisation highly complex. The second challenge is a positive one: Eliq sits on **an abundance of high-resolution data**. In particular, when the data set includes 15-minute resolution data, it amounts to approximately 35 thousand data points in a year for a single location. The 12-month span is necessary to cover the full cycle of weather conditions. Moreover, solar irradiance data must be generated for each of these data points and for each PV system parameter combination.

This can slow down the iteration time between Steps 1 and 2 very quickly – but it is reversible! Thankfully each challenge comes with a strategy to overcome it.

The Solutions

Let's hone in on the solar irradiance data that needs to be generated for parameter estimation. Since solar irradiance depends on the PV parameters, each set of parameters requires its own set of solar irradiance data in turn generated for each export data point available. This means that **the fewer data points crunched from the original data, the faster this iteration goes**.

Our chosen approach for PV disaggregation enables us to employ this solution. PV parameters depend mainly on low consumption / clear sky days – days when energy export is at its peak and is the closest to the actual solar production. This enables the selection of days with the highest solar energy export and the estimation of PV parameters only with that data. This way secures the high quality of the estimation and reduces the computational resources. Additionally, we ensure that peak days across the year are selected so that the changing conditions for all seasons are fairly captured. This estimation strategy enables PV disaggregation to reduce the computation time in half.

When it comes to the challenge of locating the parameters needed for the evaluation, the slimmed-down data approach that solves the previous challenge wouldn't work as positively on this one. Since this was a known challenge, the joblib library that provides tools for parallel computation was employed from the get-go. Mid-process, the Eliq team discovered the potential for huge improvements with the Ray library, which was immediately put to the test but didn't demonstrate significantly improved results. A diligent investigation revealed that the data was sent to each actor in the Ray library creating an unnecessary bottleneck for inter-process communication. The team's solution to this was to serialise the function with the data before sending it to the workers. The result was an impressive 2x speedup!

The Results

This was a detailed tour under the hood of Eliq's Solar PV Disaggregation – now it's time to take it for a drive and witness the results. Below are the data accuracy results of the process. The data set includes 69 locations checked for data quality, from a single, West European country with a year-round worth of energy import, export, and solar energy data.

MdAPE rate bucket	Number of locations // **Hourly**	Cumulative % // Hourly	Number of locations // **Daily**	Cumulative % // Daily
(0%, 10%]	-	0	23	33
(10%, 20%]	31	45	37	87
(20%, 30%]	31	90	6	96
(30%, 40%]	6	99	2	99
(40%, Inf)	1	100	1	100

Table 1. Medium Absolute Percentage Error (MdAPE) for hourly and daily granularity of solar production data

Regarding the rather unconventional choice of metric with median APE (Absolute Percentage Error) instead of the more conventional mean APE, this was a conscious choice. Mean APE is a notoriously misleading metric when there are outliers. Given the dataset, which includes a full year of data for each location with 8,700 hourly observations and 365 daily observations, some outliers are expected and should be accounted for. Therefore median APE is a better choice that allows for more impactful improvements.

Regarding the rather unconventional choice of metric with median APE (Absolute Percentage Error) instead of the more conventional mean APE, this was a conscious choice. Mean APE is a notoriously misleading metric when there are outliers. Given the dataset, which includes a full year of data for each location with 8,700 hourly observations and 365 daily observations, some outliers are expected and should be accounted for. Therefore median APE is a better choice that allows for more impactful improvements.

When it comes to the results, the algorithm shows an overall adequate performance. The majority of the locations are in the 10% to 30% error rate range. The daily-level data meets and exceeds Eliq's quality standard for end users, with 33% of locations in the 0-10% error bucket and a total of 87% of locations with an error rate below 20%. In correlation with how slim the data input for solar production estimation is, this is a rather impressive result. The performance on hourly-level data ranks at a 20-30% error rate, which is not up to par with Eliq's quality standard for end users. In this iteration, no location places in the best-performing bucket of 0-10% error rate. However, with the rise in PV installations, we expect this metric to improve over time.

Summary

Solar installations are bound to become a standard feature in most locations that can support an installation. Given the slim quality of insights and management today's prosumers have, this presents an urgent challenge for energy providers to supply the same standard of care for this expanding client segment. Eliq recognised this development early and decisively created the solution with PV disaggregation, aiming to quickly and effectively enable energy providers to offer high-quality customer experience for prosumers without having to labour over data quality and insight extraction.