Should we invest in additional solar panels ?

A study on Solar panel production.

By Denis Froment, 2025

Situation

Context and motivation

We have bought solar panels one year ago.
They produce well and we save on electricity bills.

We calculated that our investment will reach break-even after 9 years.
We're considering buying additional solar panels, because we like the feeling of self-production, but we want to keep investment aligned with a return horizon of less than 10 years.

Business question

"Should we invest in additional solar panels, given ROI and degradation rates?"

Acceptance criteria

We would buy additional panels only if the investment would beat €ster before 10 years.

Why 10 years ? because beyond this duration, investing in green infrastructure with tax exemption plan brings more than €ster, and this would be our best bet. Same green impact, better ROI.

We will assume, for the study, that energy prices increase at the €ster rate.

Investment and return up to now

How fast will the current panels bring ROI ?

There are many parameters we can think of for this calculation. We'll consider that not invested money yields €ster, and that energy bills also increase following €ster.

Figures

We bought our first solar panels for 400 EUR.
On the first year, they produced electricity reducing our electricity bill for an equivalent of 66 EUR:

Solar Panel Cost & Savings Report (12 Months)

Category	Amount / Detail	Notes
Initial Investment	400 EUR	Purchase & installation of solar panels
Energy Produced	356.83 kWh	Total production over 12 months
Cost per kWh (Daytime Rate)	0.19 EUR	Reference electricity tariff
Energy Returned to Grid	7 kWh	Not self-consumed, fed back into the grid
Self-Consumed Energy	349.83 kWh	356.83 – 7
Total Savings	66.47 EUR	Reduction in electricity bill
Payback Ratio (Year 1)	16.60%	66.47 ÷ 400

ROI calculation

If the panels keep this performance, we fully get our money back in 6 years.
The break-even for buying solar panels instead of getting €ster from 400€ would be 7 years:

Yearly prod decrease		0.0%			€ster rate	1.90%	
Year number	Opening balance	saving	Balance EOY	panels PnL	Opening balance	€ster return	Balance EOY
1	-400.00 €	66.47 €	-333.53 €	-341.13 €	400.00 €	7.60 €	407.60 €
2	-333.53 €	67.73 €	-265.80 €	-281.14 €	407.60 €	7.74 €	415.34 €
3	-265.80 €	69.02 €	-196.78 €	-220.01 €	415.34 €	7.89 €	423.24 €
4	-196.78 €	70.33 €	-126.45 €	-157.72 €	423.24 €	8.04 €	431.28 €
5	-126.45 €	71.67 €	-54.78 €	-94.25 €	431.28 €	8.19 €	439.47 €
6	-54.78 €	73.03 €	18.25 €	-29.57 €	439.47 €	8.35 €	447.82 €
7	18.25 €	74.42 €	92.67 €	**36.34 €**	447.82 €	8.51 €	456.33 €

Nota: yearly savings increase because of electricity prices going up at €ster rate

But in reality, there is a yearly decrease in produced electricity from our solar panels.
5% per year is announced by the vendor, which gives following forecast:

Yearly prod decrease		5.0%			€ster rate	1.90%	
Year number	Opening balance	saving	Balance EOY	panels PnL	Opening balance	€ster return	Balance EOY
1	-400.00 €	66.47 €	-333.53 €	-341.13 €	400.00 €	7.60 €	407.60 €
2	-333.53 €	64.35 €	-269.18 €	-284.53 €	407.60 €	7.74 €	415.34 €
3	-269.18 €	62.29 €	-206.89 €	-230.13 €	415.34 €	7.89 €	423.24 €
4	-206.89 €	60.30 €	-146.59 €	-177.87 €	423.24 €	8.04 €	431.28 €
5	-146.59 €	58.37 €	-88.22 €	-127.69 €	431.28 €	8.19 €	439.47 €
6	-88.22 €	56.51 €	-31.71 €	-79.53 €	439.47 €	8.35 €	447.82 €
7	-31.71 €	54.70 €	22.99 €	-33.34 €	447.82 €	8.51 €	456.33 €
8	22.99 €	52.96 €	75.95 €	**10.95 €**	456.33 €	8.67 €	465.00 €

So the real break-even considering panel power decrease and €ster comes after **8 years**.

Further impacts on ROI

- Would additional panels bring the same savings ?

  No, at times when the panels produce more than we consume, it is given back to the grid.
  We can see such times when the grid-consumption from our house is zero (purple line):

Around this time, some energy is returned to the grid

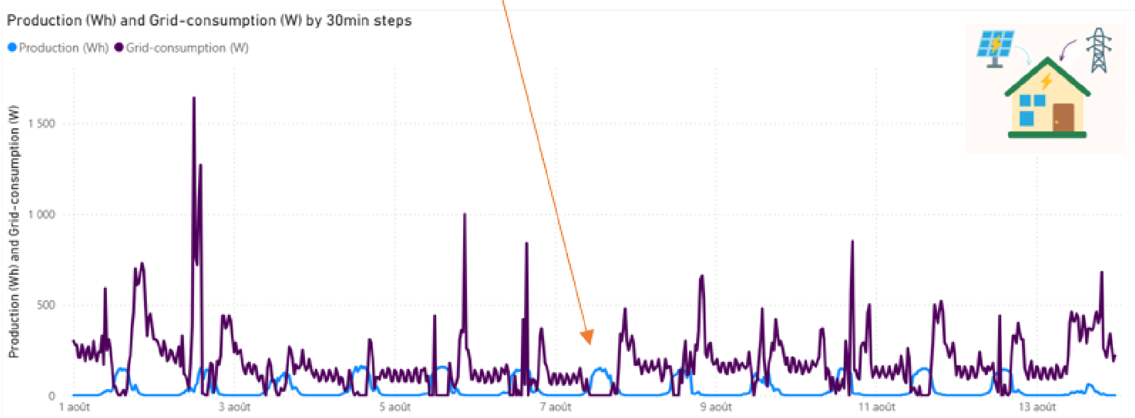


figure: Power BI. Data source: Grid provider (30 min bins) and panel box sending (Data download by app).

  With additional panels, more energy would return to the grid, without triggering any saving for us (we do not stock energy, cost for a battery is not acceptable).

  Of course, the progression of this is exponential for small values. We lost 7/357=2% of produced energy in 12 months, and we can consider that 4% of the additionally produced power would feed the grid.

- How much is the decrease of power panel capacity

  This is the most important question: The vendor indicates a yearly loss of production power of 5%.
  If it happens to be more, the ROI would be impacted even for our first panels.
  Let’s dig into the data to find out.

Methodology

Comparing Year-on-Year monthly production values makes no sense: the amount of sunlight is not the same every year.

We need data of sun radiation from an environmental/weather service to compare them to the production.
Once we have these values for sun radiation and corresponding panel production, we can find regression functions for a month of previous year and for this year and compare them to know the power production capacity was reduced.

But we must mitigate following risks:

Risk of error	**Mitigation**
Shadows from neighborhood on the panel	Check hourly distribution and identify shadowless time. Then consider only this time to compare YoY
Time measurement from panels vs weather service	Align the timestamps by postprocessing
Dust on the panels	We clean them every month
Measurements in the town vs single clouds on our premises	This is a tricky point. Weather service might record full sun when we locally still have cloud shadow, and vice versa. To mitigate this, we shall keep only values of production & radiation that are high enough to be supposed cloud-free.

Extract Data

Consumption data

For our own information, and to estimate grossly spots where additional electricity from the panels would be fed to the grid, we got data from the grid provider.

Format: Excel. Some lengthy headers processed in Power Query

Saved on github in:

https://github.com/dfrome/solar_panel/tree/main/data/raw/consumption

Once loaded, we can draw both lines consumption + production on same Date.

Production data

Created by the module attached to the power cells. Might be optimistic, because it is in their advantage to show more than reality. But first checks show plausible values (through addition of consumed+produced compared to usual consumption before installation of the solar cells).

Downloaded via the Beem android app : this sends file by e-mail. Then extract the file from the e-mail and copy it to:

https://github.com/dfrome/solar_panel/tree/main/data/raw/production

Copernicus weather data

Get the data manually

From https://cds.climate.copernicus.eu/datasets/sis-ecde-climate-indicators?tab=overview ,
chose download, then search solar. You can download files manually but it will be more industrialized/reproductible to send API request.

Get the data from API

We create a python script:

import_cdsapi.py, in https://github.com/dfrome/solar_panel/tree/main/src/

and API credentials in a separate file (gitignored), see https://cds.climate.copernicus.eu/how-to-api

It downloads zipped files. In the same Python, we unzip and copy to the folder scrutinized by Power BI.
Due to Copernicus API quotas, I requested over several days to get all months (07/2024 to 09/2025).

Import data downloaded from Copernicus in Power BI

We create a dedicated python environment with necessary packages, and configure Power BI desktop to use it.

Then we load all files from the folder with a python script embedded in Power Query:

```
= Python.Execute("import xarray as xr
import pandas as pd
import os

# Dossier containig .nc files
folder = r'F:\<your local path>\fromAPI01'

# .nc file list
nc_files = [f for f in os.listdir(folder) if f.endswith('.nc')]

# List to store DataFrames
df_list = []

# Loop on each file
for file in nc_files:
    path = os.path.join(folder, file)
    ds = xr.open_dataset(path)
    df_all = ds.to_dataframe().reset_index()
    df_all['source_file'] = file  # Optionnel : pour tracer l'origine
    df_list.append(df_all)

# final concatenation
dataset = pd.concat(df_list, ignore_index=True)")
```

We can show daily solar radiation values:

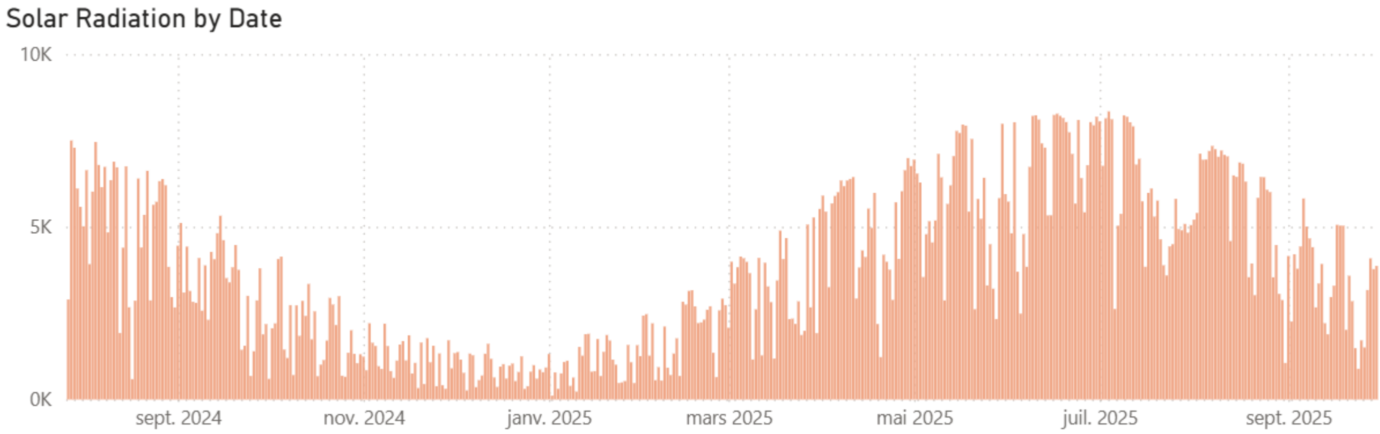


figure: Power BI. Data source: Copernicus API, value="ssrd" (Surface Solar Radiation Downwards).

Time synchronization

Copernicus time

It is supposed to be UTC according to discussion at:

https://forum.ecmwf.int/t/time-zone-and-time-stamp-in-era5-hourly/12478/11

UTC is the time at Greenwich and is never adjusted for daylight saving time.

Solar panel production time

According to documentation, the data timestamps are supposed to be given in CET/CEST

https://en.wikipedia.org/wiki/Central_European_Time

From the last Sunday in March to the last Sunday in October, countries within the CET area (like mine) switch to Central European Summer Time (CEST, UTC+02:00) for the summer.

Adapting data

Given the time shift, it is better to add a column “Shifted Hour” to one of our fact tables and use it to match Data for regression model in next steps.

Production:

- last Sunday in March to the last Sunday in October, is +2 compared to Radiation
  => apply -2 to Production time (Hour) before our measures compare time with Radiation,
  (or apply +2 to Radiation time)
- other dates

⇨ apply -1 to Production time (or -1 to radiation)

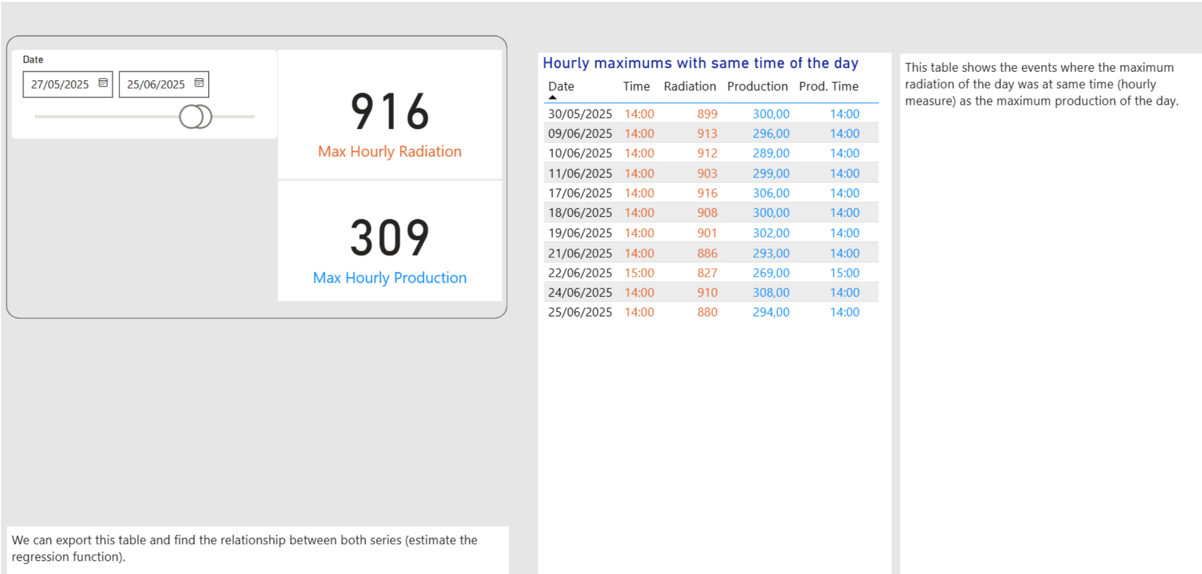
The M snippet I created for Shifted Hour for Radiation was :

```
#"Added Shifted Hour" = Table.AddColumn(#"Changed Type", "Shifted Hour", each
let
    currentDate = [Date],
    currentHour = [Hour],
    currentYear = Date.Year(currentDate),
    startDST = Date.StartOfWeek(#date(currentYear, 3, 31), Day.Sunday),
    endDST = Date.StartOfWeek(#date(currentYear, 10, 31), Day.Sunday),
    myShift =
        if currentDate >= startDST and currentDate < endDST
        then 2
        else 1,
    shifted =
        if currentHour < 24 - myShift
        then
            currentHour + myShift
        else
            currentHour - 24 + myShift
```

This works, we can see the switch from CEST to CET here:

3 distinct, 73 unique	1 distinct, 0 unique	1 distinct, 0 unique	44 distinct, 43 unique	2 distinct, 1 unique	4 distinct, 1 unique	24 distinct, 0 unique	24 distinct, 0 unique
29/03/2025 00:00:00	47.25	0.75	0	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	0	1
29/03/2025 01:00:00	47.25	0.75	0	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	1	2
29/03/2025 02:00:00	47.25	0.75	0	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	2	3
29/03/2025 03:00:00	47.25	0.75	0	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	3	4
29/03/2025 04:00:00	47.25	0.75	0	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	4	5
29/03/2025 05:00:00	47.25	0.75	0	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	5	6
29/03/2025 06:00:00	47.25	0.75	1,7721156	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	6	7
29/03/2025 07:00:00	47.25	0.75	81,890396	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	7	8
29/03/2025 08:00:00	47.25	0.75	257,81174	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	8	9
29/03/2025 09:00:00	47.25	0.75	427,53067	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	9	10
29/03/2025 10:00:00	47.25	0.75	581,1762	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	10	11
29/03/2025 11:00:00	47.25	0.75	648,08136	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	11	12
29/03/2025 12:00:00	47.25	0.75	613,99146	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	12	13
29/03/2025 13:00:00	47.25	0.75	685,18695	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	13	14
29/03/2025 14:00:00	47.25	0.75	625,8366	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	14	15
29/03/2025 15:00:00	47.25	0.75	569,7196	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	15	16
29/03/2025 16:00:00	47.25	0.75	422,85165	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	16	17
29/03/2025 17:00:00	47.25	0.75	250,63	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	17	18
29/03/2025 18:00:00	47.25	0.75	97,94825	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	18	19
29/03/2025 19:00:00	47.25	0.75	3,5908659	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	19	20
29/03/2025 20:00:00	47.25	0.75	0	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	20	21
29/03/2025 21:00:00	47.25	0.75	0	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	21	22
29/03/2025 22:00:00	47.25	0.75	0	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	22	23
29/03/2025 23:00:00	47.25	0.75	0	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	29/03/2025	23	0
30/03/2025 00:00:00	47.25	0.75	0	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	30/03/2025	0	2
30/03/2025 01:00:00	47.25	0.75	0	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	30/03/2025	1	3
30/03/2025 02:00:00	47.25	0.75	0	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	30/03/2025	2	4
30/03/2025 03:00:00	47.25	0.75	0	H_ERA5_ECMW_T639_GHI_0000m_Euro_025d_S202503010000_E20...	30/03/2025	3	5

And it works for summer: for the time selected below, we had only 6 common peak hours before our time synchronization, we now have 11.



Date	Time	Radiation	Production	Prod. Time
30/05/2025	14:00	899	300,00	14:00
09/06/2025	14:00	913	296,00	14:00
10/06/2025	14:00	912	289,00	14:00
11/06/2025	14:00	903	299,00	14:00
17/06/2025	14:00	916	306,00	14:00
18/06/2025	14:00	908	300,00	14:00
19/06/2025	14:00	901	302,00	14:00
21/06/2025	14:00	886	293,00	14:00
22/06/2025	15:00	827	269,00	15:00
24/06/2025	14:00	910	308,00	14:00
25/06/2025	14:00	880	294,00	14:00

figure: Power BI. Data source: Solar anel box (Data download by app) and Copernicus environmental data.

Looking at the charts for winter and summer, I finally opted for a one hour shift valid for the whole year (myShift = 1) instead of the CET/CEST differenciation. It seems the Beembox does not apply CEST 😊.

I sent a question to the panel providers about this.

A visualization also helps us check what time of the day should be considered for calculating regression parameters:

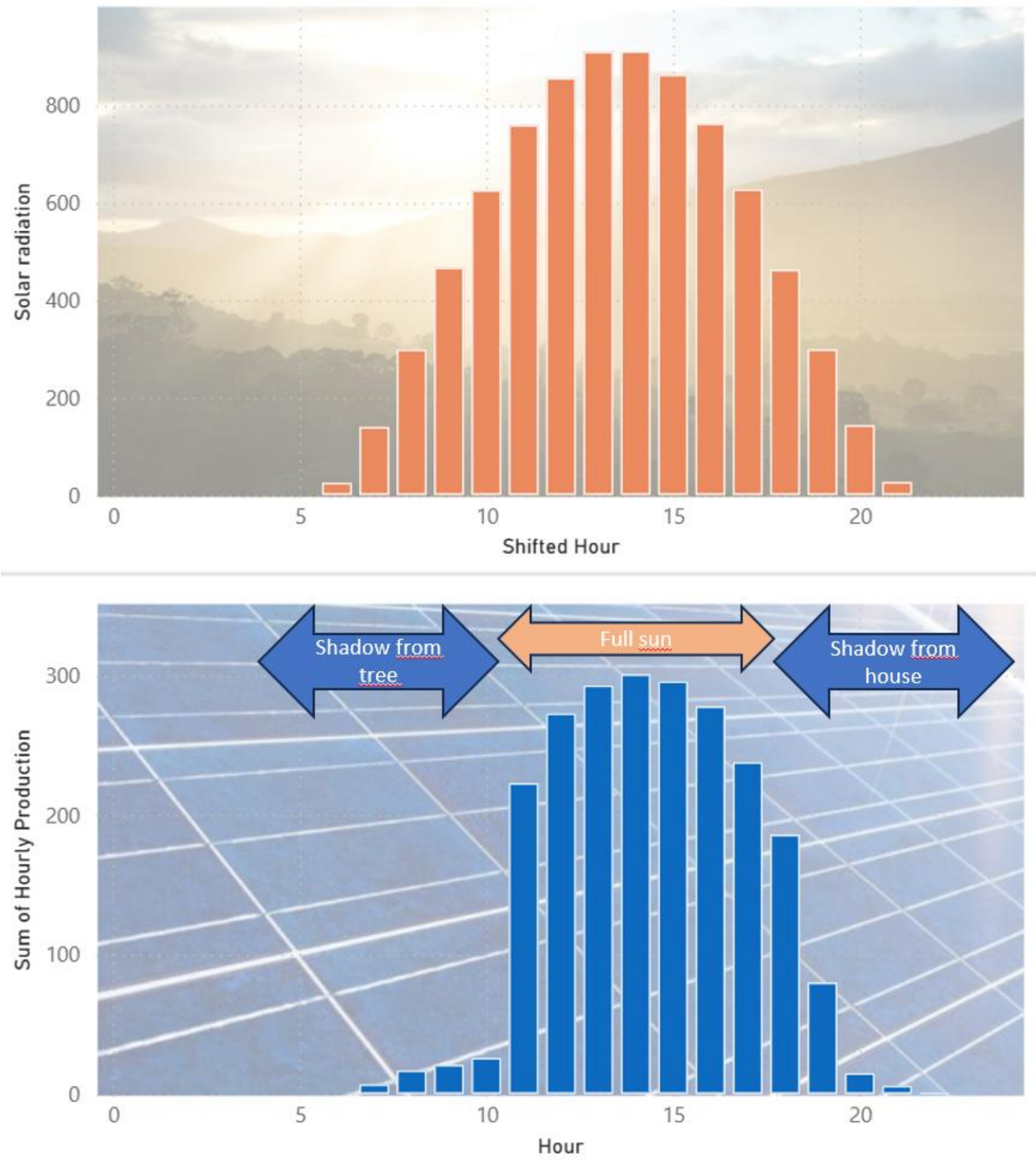


figure: Power BI and arrows added in powerpoint

This typical daily distribution shows that:

1. The time is correctly synchronized
2. The hours when the panels receive full sun are between 11 and 17 (included). So we should consider those for regression calculation.

Although there will always be discrepancies because of local conditions (moving clouds in the area, birds or cats sitting on the panels...), we can see that the best time alignement has been achieved:

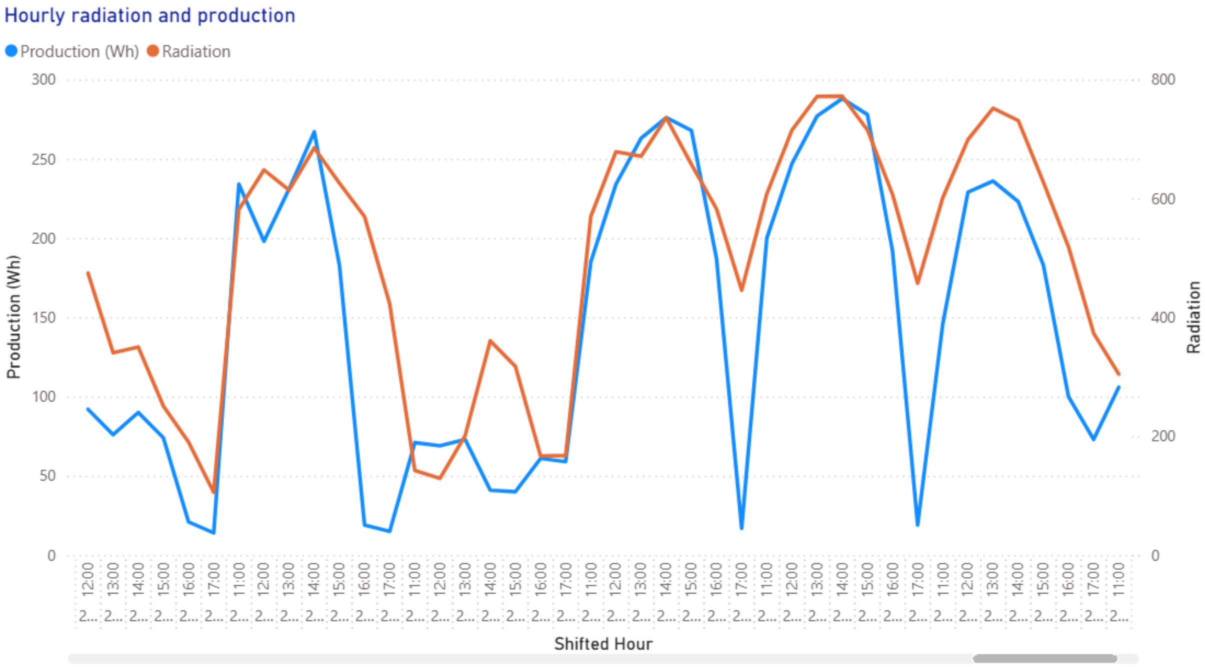


figure: Power BI. Data source: Solar panel box (Data download by app) and Copernicus environmental data.

Now we are ready to apply data science models to check the regression between Radiation and Production.

We export the complete data (filtered in Power BI to show only sunny hours) to handle by python:

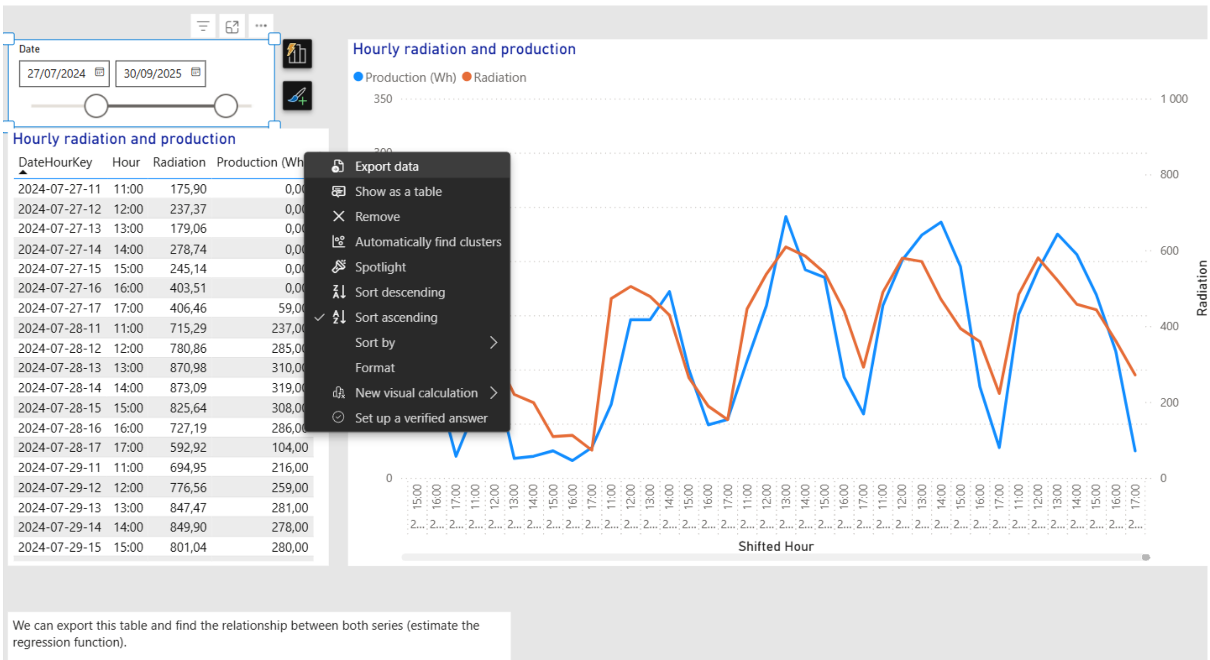


figure: Power BI. Data source: Only sunny hours, Solar panel box (Data download by app) and Copernicus environmental data.

(an alternative would be to call a python directly from Power BI, but we want to use several steps in a Jupyter Notebook, so it is simpler to save the data and process it outside power BI)

Determine correlation and regression

Data exploration in Python shows outliers.
Cleaning data:

- Data with small numbers are meaningless: We delete data where Production or Radiation is below a given threshold.
- We calculate a linear regression on target=Production and Feature=Radiation.
  Then, we delete data points that are too far from the prediction, thus keeping hours not impacted by local discrepancies between Copernicus environmental sensor and solar panels.

With: df = df[(df['Radiation'] > 500) & (df['Production'] > 130)]

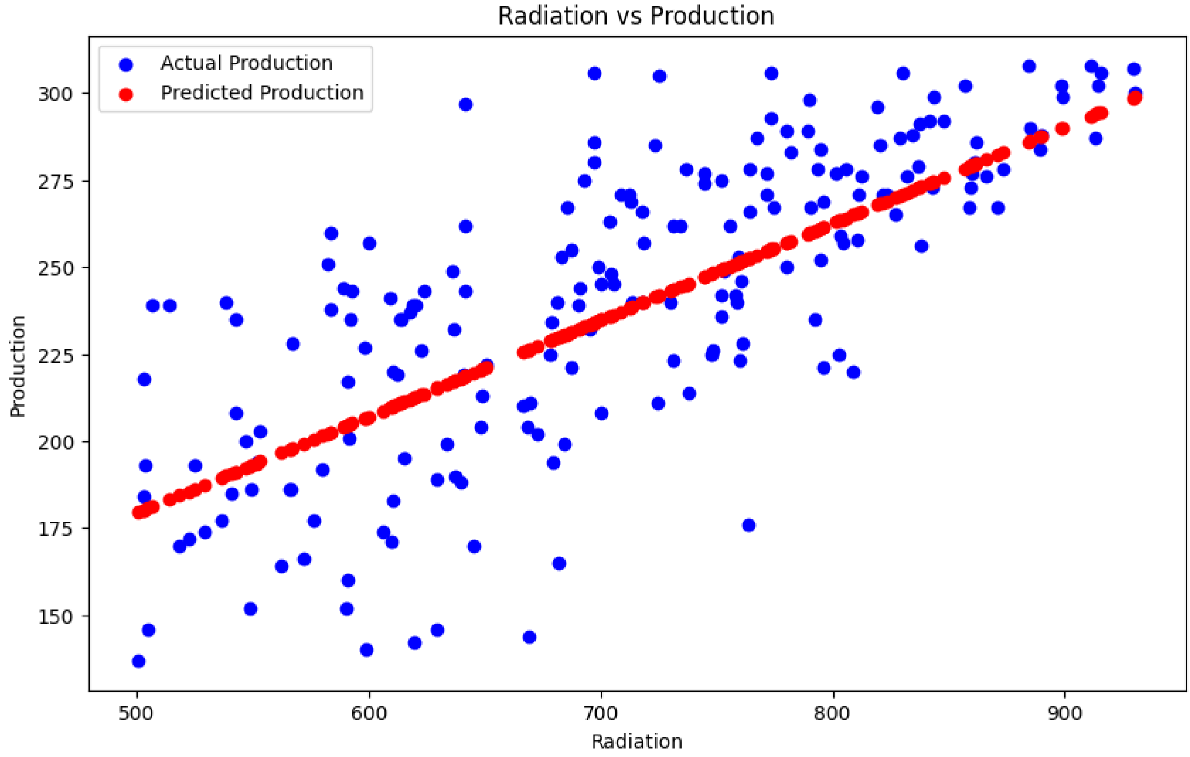
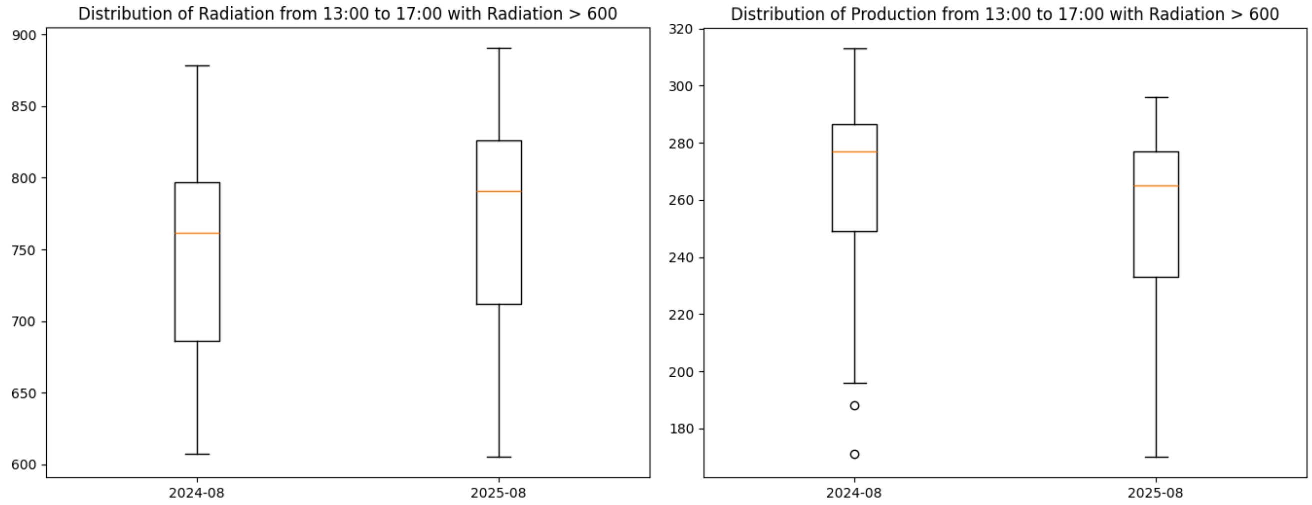


figure: Python matplotlib. Predicted production is according to linear regression

The linear regression starts to make sense. It might not be the best regression model, but we'll see later that it is a valid approximation for our purpose.

We have the full month of August in common between 2024 and 2025.



figures: Python matplotlib, units as given by Copernicus for radiation(=ssrd) and Wh for production

Although there was more radiation in 2025, we had the solar panel producing less power than in 2024. First sign of panels capacity loss, but we still need scientific proof.

We also have complete September 2025 but radiation was much higher than previous year, making comparison impossible (climate change?).

Methodology to compare Year-on-Year drift

For the same month (August) for 2024 and 2025, we will:

1/ Keep only meaningful figures
2/ Apply regression models
3/ Observe the drift

1/ Keep only meaningful figures

We focus on hours which had a reasonable amount of radiation and power produced (we eliminate again the noise of small numbers), thus ensuring that we keep the cleanest hours in the data.
This is pure Sun->power without other interference.

Those points are scarce but enough to get an insight:
68 measures for 08/2024, and
73 measures for 08/2025.

2/ Apply a regression model on data from 08/2024 and 09/2025

We calculate regression intercept and coefficient separately for each year.

3/ Observe the typical drift between those 2 regressions.

Luckily, the regression lines have a very similar coefficients, so looking at a middle value of radiation allows us to calculate easily the discrepancy of power produced with same amount of radiation:

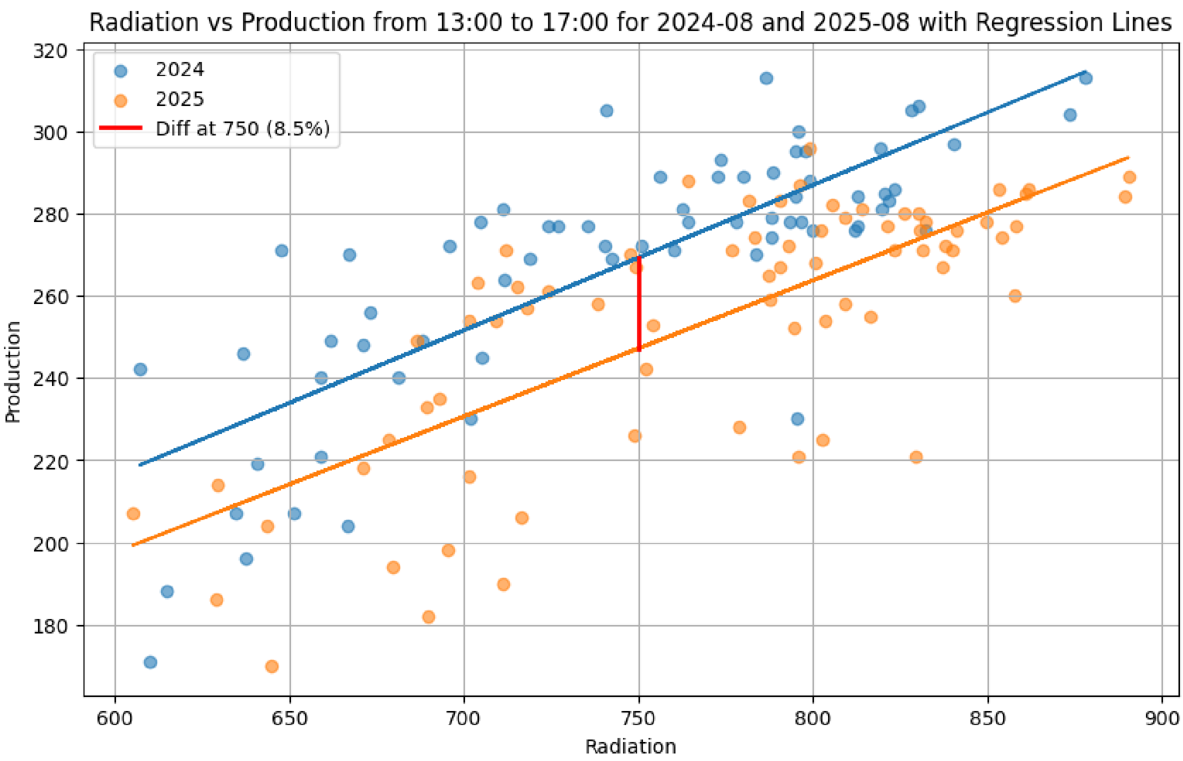


figure: Python matplotlib. Data source: Solar panel box (Data download by app) and Copernicus environmental data. Linear regression.

The result is: 8.5% less power produced for same amount of radiation Year-On-Year.
For our investment calculation, we'll assume this percentual loss would be the same every year (study suggestion for next year!).

Regression model used: Graphically, we notice that the linear regression may not be the best fit. But polynomial regression of SVM would show roughly the same delta between 2024 and 2025.

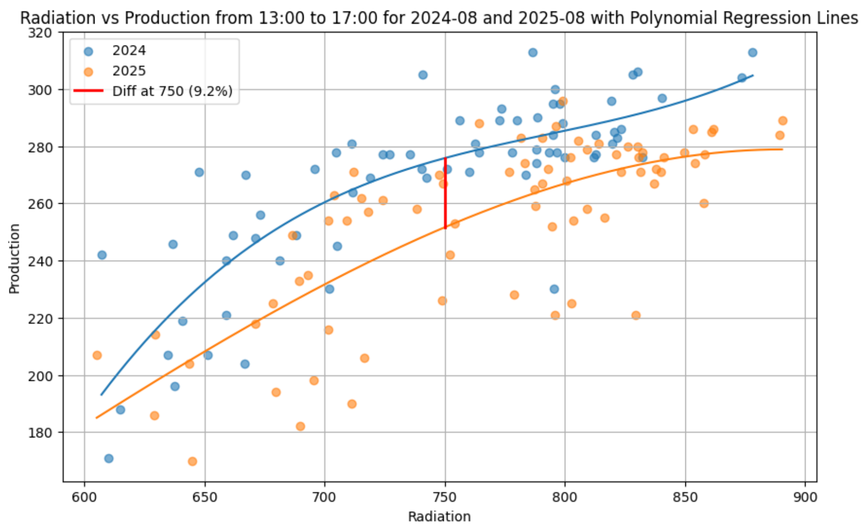


figure: Python matplotlib. Same as above but with polynomial regression, 3 degrees. No additional findings with this small number of samples.
=> We stick by the more robust figures of linear regression.

Conclusion

Applying the 8.5% annual loss to our investment calculation, together with first year saving being 63.81€ instead of 66.47€ (4% unused additional power fed to the grid) :

Yearly prod decrease	8.5%				€ster rate	1.90%	
Year number	Opening balance	saving	Balance EOY	panels PnL	Opening balance	€ster return	Balance EOY
1	-400.00 €	63.81 €	-336.19 €	-343.79 €	400.00 €	7.60 €	407.60 €
2	-336.19 €	59.50 €	-276.69 €	-292.04 €	407.60 €	7.74 €	415.34 €
3	-276.69 €	55.47 €	-221.22 €	-244.46 €	415.34 €	7.89 €	423.24 €
4	-221.22 €	51.72 €	-169.50 €	-200.78 €	423.24 €	8.04 €	431.28 €
5	-169.50 €	48.22 €	-121.28 €	-160.75 €	431.28 €	8.19 €	439.47 €
6	-121.28 €	44.96 €	-76.31 €	-124.13 €	439.47 €	8.35 €	447.82 €
7	-76.31 €	41.92 €	-34.39 €	-90.72 €	447.82 €	8.51 €	456.33 €
8	-34.39 €	39.09 €	4.70 €	-60.30 €	456.33 €	8.67 €	465.00 €
9	4.70 €	36.45 €	41.15 €	-32.69 €	465.00 €	8.84 €	473.84 €
10	41.15 €	33.98 €	75.13 €	-7.71 €	473.84 €	9.00 €	482.84 €
11	75.13 €	31.68 €	106.81 €	**14.80 €**	482.84 €	9.17 €	492.01 €

We need 11 years until break-even for the investment.

Thus, we answer our business question, with constant electricity price:

It is not worth for us to buy additional panels for our house.