Small Project - EMHIRES dataset

Alessandro Castelli ID: 12246581

December 7, 2023

Contents

1 Introduction . . . 3

2 Some important information - 30 Years of European Solar Generation . . . 3
- 2.1 What is PVGIS? . . . 3
- 2.2 How was this data collected? . . . 3
- 2.3 List of countries . . . 3
- 2.4 How to read the dataset? . . . 3
- 2.5 Description of data, tools and methodology . . . 4

3 Studies on the dataset . . . 4
- 3.1 Study by year . . . 4
- 3.2 Study by decades . . . 4
- 3.3 Explanation of Figure 2 . . . 4
- 3.4 What these data mean and what is the next step . . . 5
- 3.5 Conclusions . . . 5

1 Introduction

EMHIRES provides generation time series for the EU-28 and neighbouring countries. The solar power times series are released at hourly granularity and at different aggregation levels. The overall scope of this dataset is to allow users to assess the impact of meteorological and climate variability on the generation of solar power in Europe and not to mime the actual evolution of solar power production in the latest decades. For this reason, the hourly solar power generation time series are released for meteorological conditions of the year 1986-2015. without considering any changes in the solar installed capacity.

2 Some important information - 30 Years of European Solar Generation

2.1 What is PVGIS?

In this study, the general approach applied to convert solar resources into power generation consists in converting satellite-based radiation data using the PVIGS model. Photovoltaic Geographical Information System is an online tool provided by the Joint Research Centre (JRC) of the European Commission. PVGIS provides data and models to estimate photovoltaic solar energy production in various regions of the world. Access to PVGIS is provided through a web interface developed to offer interactive access to data, maps, and tools.

2.2 How was this data collected?

It was generated applying using the validated and robust PVGIS model to estimate the solar electricity potential capturing local geographical information to generate meteorologically derived solar power time series at high temporal and spatial resolution, validated with transmission system operators' data.

2.3 List of countries

In the following section, you will find the list of countries for which data has been recorded and preserved:

- AT = AUSTRIA
- BE = BELGIUM
- BG = BULGARIA
- CH = SWITZERLAND
- CY = CYPRUS
- CZ = CZECH REPUBLIC
- DE = GERMANY
- DK = DENMARK
- EE = ESTONIA
- ES = SPAIN
- FI = FINLAND
- FR = FRANCE
- EL = GREECE
- HR = CROATIA
- HU = HUNGARY
- IE = IRELAND
- IT = ITALY
- LT = LITHUANIA
- LU = LUXEMBURG
- LV = LATVIA
- NL = NETHERLANDS
- NO = NORWAY
- PL = POLAND
- PT = PORTUGAL
- RO = ROMANIA
- SI = SLOVENIA
- SK = SLOVAKIA
- SE = SWEDEN
- UK = UNITED KINGDOM

2.4 How to read the dataset?

As you can see from `Figure 1`, the countries are listed in the columns and the production of one hour of energy is in the rows. In total, there are 262968 rows, and since each row corresponds to one hour of recording, there are approximately $\left(\frac{262968}{24\times365} \approx 30\right)$ years in total. Let's make an example: consider the AT column and row 9 which contains the value 0.16, it means that Austria at the ninth hour of recording produced a value of 0.16 of solar energy.

2.5 Description of data, tools and methodology

The general approach to convert solar resources into power generation involves converting satellite-based radiation data using the PVIGS model. First, meteorological data is collected, then this collected data is converted into theoretical potential, i.e., the solar electricity generation in each area given by kW generated from each kW peak of a typical PV system. This means that the value of solar energy found in the dataset is expressed in terms of the ratio between the energy actually produced and the energy that could be produced with a nominal power of 1 kW. This ratio is called the **capacity factor** or performance ratio and depends on various factors such as solar radiation, temperature, orientation and inclination of the panels, losses, and system efficiency. Finally, to obtain the power generation, the installed capacity of each region is calculated. The time series are then corrected with the actual TSO generation and statistically validated for power system analysis, evaluating power peaks and ramps, duration curves, and capacity factors.

What was said above implies that in each cell of the dataset there will be a value between 0 and 1 which will be equal to:

$$\text{value} = \frac{\text{energy\_actually\_produced}}{\text{energy\_that\_could\_be\_produced}}$$

From Figure 2, you can see some interesting statistics about the dataset (some countries have missing data).

3 Studies on the dataset

Given the above-mentioned dataset, I have conducted some studies on the dataset with the aim of discovering and identifying the trends of the various European countries in the time span specified in the dataset.

3.1 Study by year

In order to understand how the solar energy production of each country has changed year by year, I created Figure 3 [1]. As you can see, the country that produces the most is Portugal, while the one that produces the least (because we do not have data on it) is Norway.

But what are these differences due to? There is no single answer to this question. The causes can be multiple and of different nature such as: wind direction, cloud cover, wind speed, hours of sunshine during the day, as can be seen from the dataset 1 `Solar energy power generation dataset`.

3.2 Study by decades

After conducting a study year by year, I also conducted a study decade by decade. For each country, I calculated the average for each decade and obtained what you can see in Figure 4. As you can observe, the energy production has not changed much over the years but has remained more or less unchanged.

3.3 Explanation of Figure 2

In this table, you will find:

- In the first column, the name of the country or region.
- In the second column, the installed capacity of photovoltaic energy as of December 31, 2015, measured in megawatts (MW), which are equal to one million watts of electrical power. The installed capacity (1) indicates the maximum electrical power that can be produced by a power plant or facility using a specific energy source at a given moment.

$$(\mathrm{InCap})_n = \frac{\left(\frac{\mathrm{Potential}_{\mathrm{NUTS2n}}}{\mathrm{AvgPotential}_{\mathrm{Country}(n)}}\right)^2 \cdot Size_n}{Size_{Country(n)}} \cdot InCap_{Country(n)} \quad (1)$$

Where:

- $InCap$ = Installed Capacity
- n = Geographical area identifier
- $Country_n$ = The country to which the geographical area belongs

[1] Remember that on the Y axis of the figure is indicated the ratio between energy produced and the maximum energy that could be produced

It is assumed that the share of solar generation for the n^{th} NUTS 2 depends on the geographical size of the region area valued over the solar potential on a NUTS2 level, derived from the PVGIS model.

- In the third column, you can find the Net Generation (annual report) in gigawatt-hours (GWh), indicating the amount of electrical energy produced by a photovoltaic system in a year. To calculate this value, the following formula (2) is used:

$$NetGeneration = InsCapacity(MW) \times CapacityFactor \times 8760 \quad (2)$$

  Where:

  - Ins capacity 1 (MW) is the maximum electrical power that can be produced by a photovoltaic system, measured in megawatts.
  - The value indicating the annual capacity factor is a number between 0 and 1, and usually, the average of annual capacity factors is taken.
  - 8760 is the number of hours in a year.

  Example:
  The average of the column *Austria* in the dataset is ≈ 0.11, so the annual capacity for that column will be: $404\,\text{MW} \cdot 0.11 \cdot 8760 = 406,989.6\,\text{MWh} \approx 400\,\text{GWh}$

- In the fourth column, you will find the Net Generation (hourly time series) in gigawatt-hours (GWh), indicating the amount of energy produced by a photovoltaic system measured on an hourly basis for a specific period of time. To calculate this value, the following 3 formula is typically used:

$$\text{Net Generation} = \sum_{i=1}^{n} \text{NetGen (hourly) MWh}_i \quad (3)$$

  Where:

  - n = number of hours
  - NetGen (hourly) MWh_i = the amount of electrical energy produced in the i-th hour of the time period, measured in megawatt-hours (MWh)

The table has been sorted in alphabetical order by country.

3.4 What these data mean and what is the next step

This study can provide several interesting insights for the model we intend to build in the near future. First of all, we can expect a solar energy production that is correlated with the trend of the data we have seen in the dataset. We have observed that over 30 years, the production data have remained more or less the same and the reason why there are some countries that produce much more than others. This trend could continue more or less in the same way, unless there are significant changes in environmental conditions or energy policies.
Another very important thing to consider when talking about energy production of this kind are the uncertain variables. Many things, in addition to physical variables such as: wind direction, presence of clouds, etc., influence solar energy production. Consider climate change, which is so drastically altering the weather in the world and will continue to do so in an unpredictable way even in the coming decades. Another factor is government choices, which encourage or discourage each country to push towards renewable energies, therefore to invest in building and modernizing renewable energy production infrastructures and in particular solar energy infrastructures. All these factors are difficult to predict and model in a machine learning algorithm, and require constant updating and validation of data and models.

3.5 Conclusions

This dataset contains a lot of information that will be useful to us after we have created the model with the other datasets. From the data observed and from the fact that these data do not have significant variations over the years, it makes sense to expect that the values that will be predicted after the creation of the model will be more or less consistent with the data present in this dataset. For the creation of the actual Machine Learning model we will use the datasets that have features such as: wind direction, cloud cover, time of year, etc. and after we have obtained the data it will be possible to compare them with those studied in this dataset.

```
-----------------------------------------------------First 20 rows of the dataset-----------------------------------------
          AT        BE        BG        CH  CY        CZ        DE        DK  \
0   0.000000  0.000000  0.000000  0.000000   0  0.000000  0.000000  0.000000
1   0.000000  0.000000  0.000000  0.000000   0  0.000000  0.000000  0.000000
2   0.000000  0.000000  0.000000  0.000000   0  0.000000  0.000000  0.000000
3   0.000000  0.000000  0.000000  0.000000   0  0.000000  0.000000  0.000000
4   0.000000  0.000000  0.000000  0.000000   0  0.000000  0.000000  0.000000
5   0.000000  0.000000  0.000000  0.000000   0  0.000000  0.000000  0.000000
6   0.000000  0.000000  0.000000  0.000000   0  0.000000  0.000000  0.000000
7   0.000000  0.000000  0.052450  0.000000   0  0.000000  0.000000  0.000000
8   0.117483  0.000000  0.135141  0.068950   0  0.069215  0.019261  0.000000
9   0.166434  0.064906  0.206412  0.152344   0  0.127238  0.069999  0.044591
10  0.148252  0.052604  0.211925  0.222754   0  0.135462  0.106351  0.077809
11  0.165035  0.055391  0.195369  0.281452   0  0.207547  0.159913  0.089586
12  0.153846  0.057430  0.179980  0.283257   0  0.201180  0.157862  0.089498
13  0.132867  0.073741  0.143731  0.239572   0  0.138849  0.129450  0.072395
14  0.111888  0.092431  0.049545  0.200476   0  0.094823  0.083720  0.042892
15  0.151049  0.155978  0.000000  0.132959   0  0.026125  0.031445  0.000000
16  0.000000  0.000000  0.000000  0.000000   0  0.000000  0.000000  0.000000
17  0.000000  0.000000  0.000000  0.000000   0  0.000000  0.000000  0.000000
18  0.000000  0.000000  0.000000  0.000000   0  0.000000  0.000000  0.000000
19  0.000000  0.000000  0.000000  0.000000   0  0.000000  0.000000  0.000000
```

Figure 1: First 20 rows of the dataset

COUNTRY	Installed capacity as of 31/12/2015 (MW)	Net Generation (annual report) GWh	Net Generation (hourly time series) GWh
Austria	404	400	820
Belgium	3068	3000	3038
Bulgaria	1041	1400	NA
Switzerland	756	0	NA
Cyprus	85	0	NA
Czech Re	2067	2200	2198
Germany	38411	35200	34746
Denmark	781	600	NA
Estonia	6	0	NA
Spain	6967	8263	4232
Finland	11	0	NA
France	6192	7400	7175
Greece	2444	3665	3583
Croatia	44	0	NA
Hungary	29	0	NA
Ireland	1	0	NA
Italy	19100	23900	19673
Lithuania	69	100	NA
Luxemburg	116	100	NA
Latvia	2	0	NA
Netherlands	1429	100	981
Norway	14	0	NA
Poland	87	0	NA
Portugal	429	800	759
Romania	1249	2000	1239
Slovenia	263	200	375
Slovakia	532	500	NA
Sweden	104	0	29
United Kingdom	9000	7500	7655

Figure 2: PV installed capacity(MW) ass of 31st of December 2015 and annual generation (GWh) by country

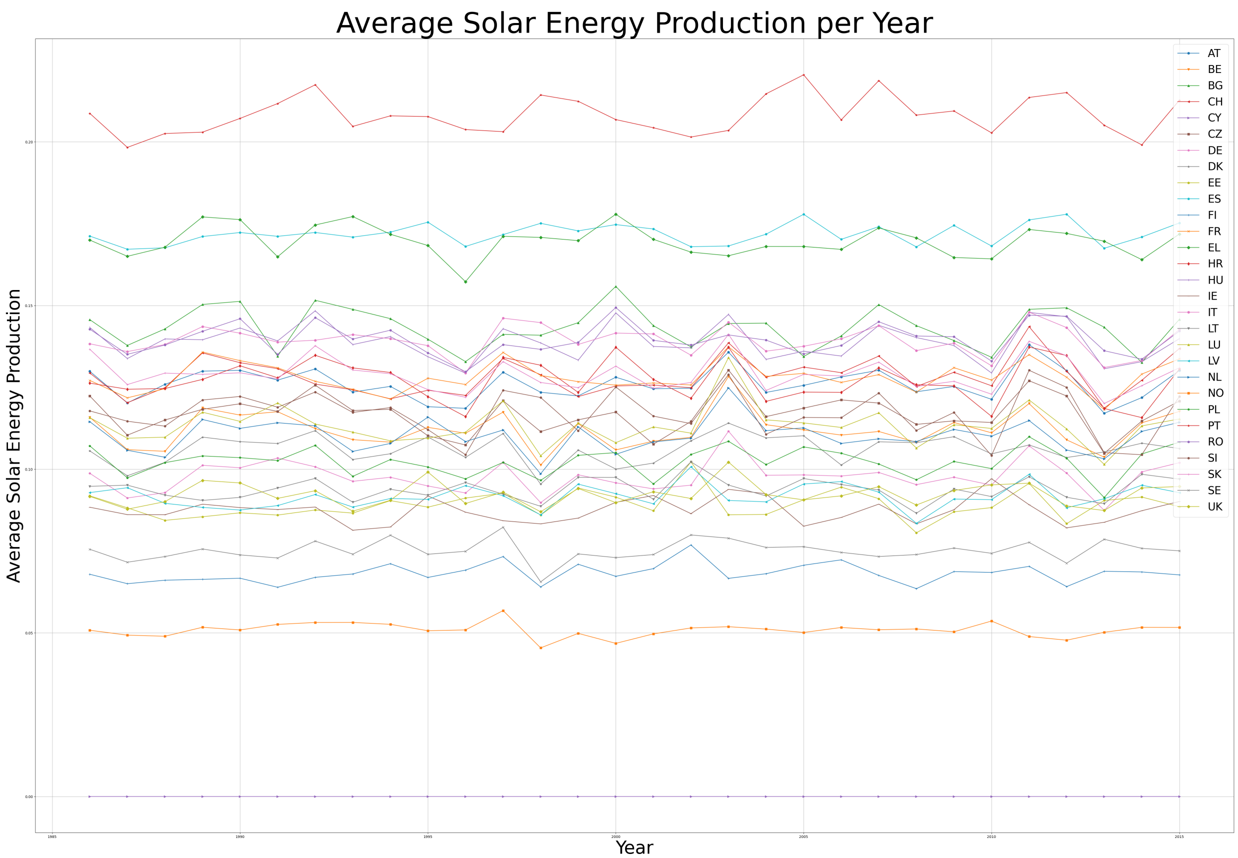

Figure 3: Graphic

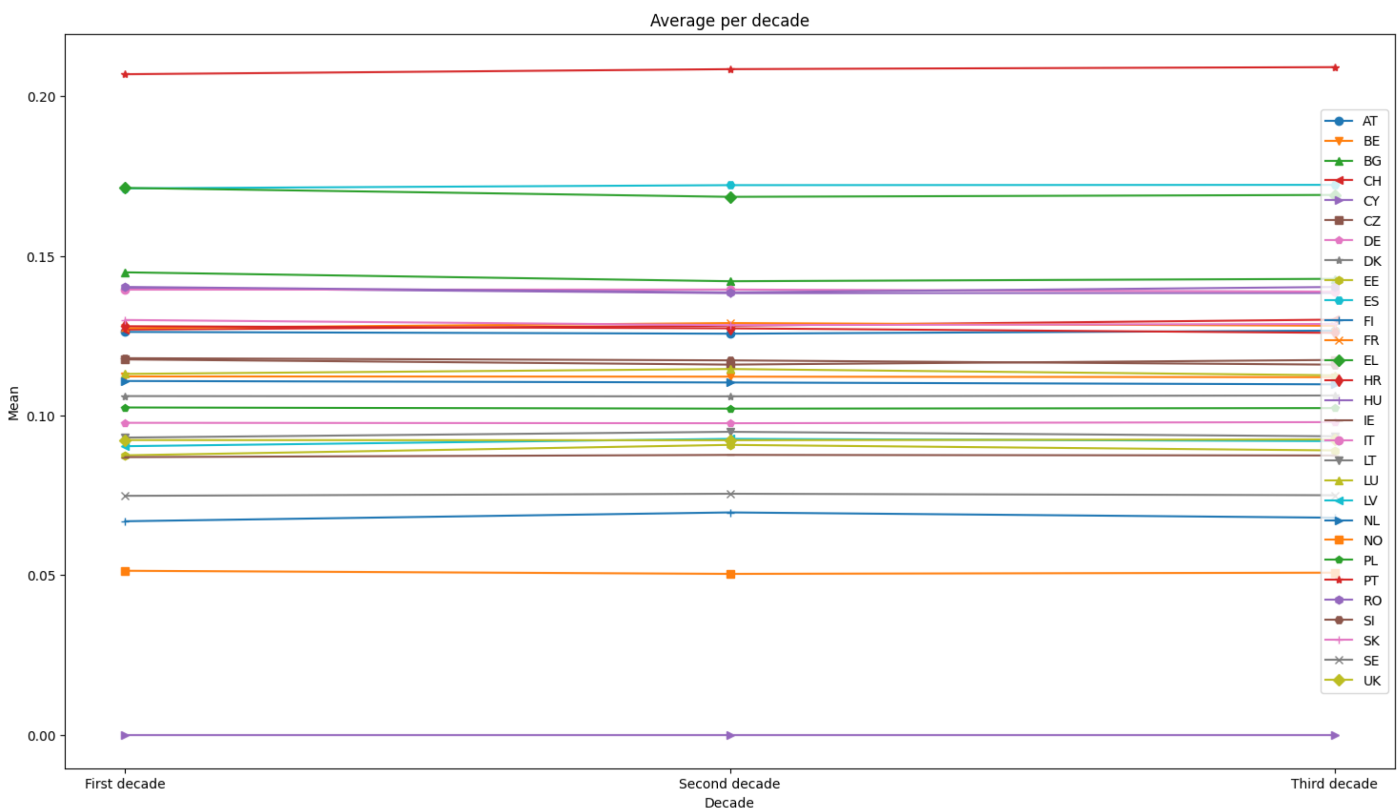

Figure 4: Graphic per decade