Estimating solar energy yields   
from citizen science and satellite data

**Keywords:** solar energy, energy transition, citizen science, satellite data

# Introduction

Climate change is a world-wide challenge. At the 2015 Paris climate agreement nations committed themselves to limiting climate heating to 1.5 degrees relative to pre-industrial levels [1]. In the European Union, the Commission has set a climate target as well. By 2030, the renewables share must stand at 32% for each individual member state [2]. To reach these goals it is important to give detailed insight into the state of affairs regarding renewable energy. In this research we present a novel approach for estimating the generated solar energy of a country on a daily and regional basis.

We combine measurement data from two ‘new’ data sources. The first one is pvoutput, an international on-line portal with real-time information of solar energy yields from photovoltaic (PV) systems (solar panels) provided by citizens or companies on a voluntary basis. The portal has data for many countries. The second data source used is modelled irradiance data derived from international satellite data.

In this research we combine these sources to infer the relation between irradiance and yield for any given day and apply it to our solar panel register to estimate daily national and regional solar energy yields.

# data sources and data cleaning

Statistical offices have a long tradition of data collection. These days survey-based data is supplemented and wherever possible replaced by administrative data and data from sensors, web data and other new data sources. Such new data sources are usually not designed for official statistics, rather they are a side-product of another business goal. The same holds for the data sources used in this approach. Hence they have to be processed, interpreted and cleaned.

## Pvoutput

Pvoutput [3] is an Australian online portal which presents near real-time energy production data of PV systems around the world. Figure 1 shows an example of the energy production profile for an entire day at a specific PV output site. After Australia, the Netherlands has the largest share of installed capacity on the website. The number of Dutch PV systems in 2016 and 2017 was on the order of 5600. In terms of capacity they represent around 0.9% of the total capacity in the Netherlands. Other countries with substantial registered capacity are the USA, Italy, UK, Germany, and Belgium.

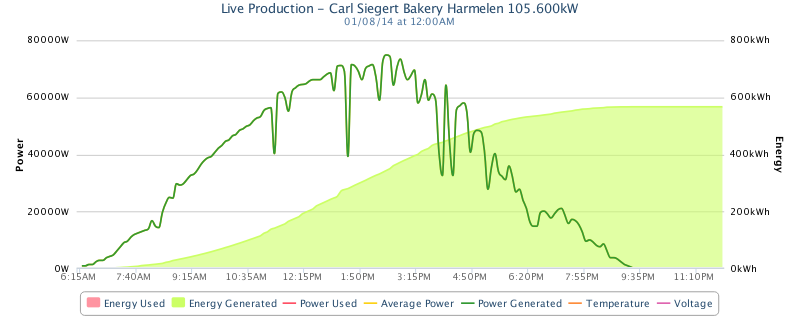


Figure 1: pvoutput energy production profile

The pvoutput data contain yields as well as system characteristics, such as orientation and tilt, panel power, number of panels, system size and inverter brand. Performance data from pvoutput appears sometimes to be unreliable due to missing measurements, misspellings in inverter brands or unlikely yields. Data cleaning was performed by specifying four quality checks applied to the daily measurements in relation to the system characteristics. We refer to [4] for a more detailed explanation of these checks. The result is that for each day we construct a set of reliable systems for which we use the data for that day. The size of this set varies around 1200 systems.

## Modelled irradiance data from satellite

The Meteosat second generation satellites operated by EUMETSAT [5] contain the Spinning Enhanced Visible and Infrared Imager (SEVIRI) instrument [6]. This instrument observes properties of the atmosphere every 15 minutes at a resolution of 3x3 km2 which results - due to projection effects - in a resolution of 3x6 km2 for the Netherlands. The Royal Dutch Meteorological institute (KNMI) developed a physics based empirical model to calculate irradiance on this resolution [7]. We aggregate these data into daily grid totals. Figure 3 shows the modelled total irradiance grid for two different days in June 2016, showing the large variety per region.

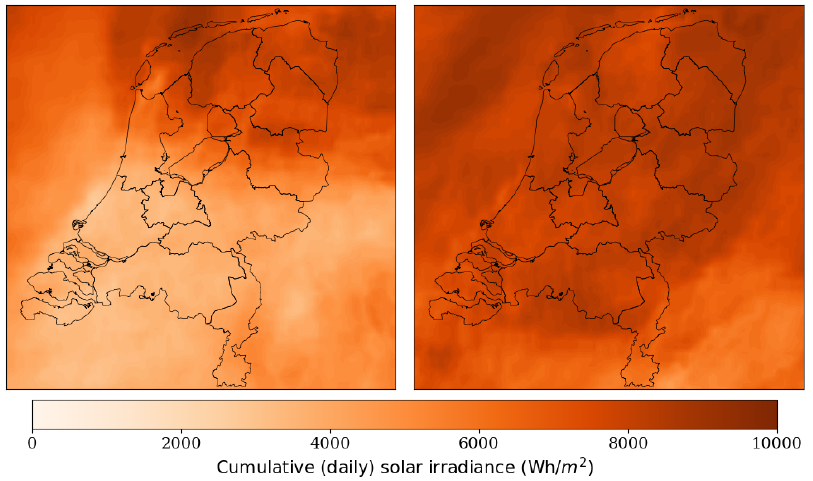


Figure 3: Cumulative daily irradiance on two different days in 2016

# Methodology

## Linking data sources

The first step is to link the two data sources based on geographical location. For the precise linking strategy we refer to [4]. Figure 3 shows the quarter hourly irradiance and energy patterns for two different pvoutput sites. While the patterns are well correlated for the system on the left, they are less so for the system on the right. At the location on the right hand size, other local factors are presumably at play such as shade or impeding objects, resulting in lower yields than one might expect, given the irradiance.

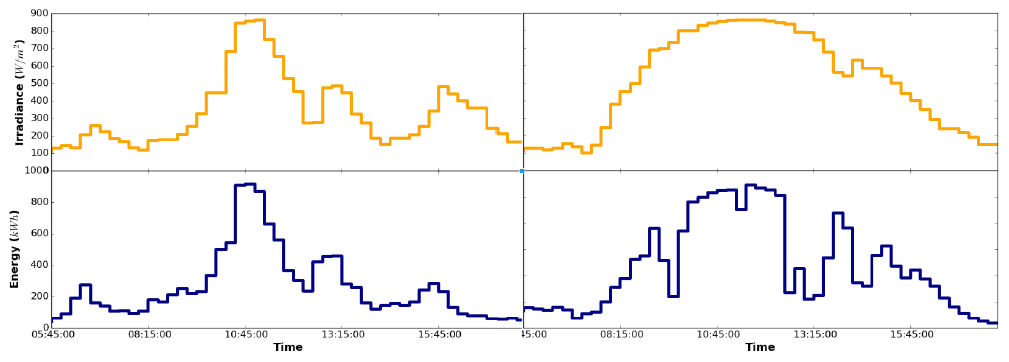


Figure 2: Irradiance (top) and yield (bottom) for two different systems

## Modelling irradiance versus yield

Combining the two aforementioned data sources, we create 2D probability density functions, linking irradiance and yield with each other. Figure 3 shows this function in the irradiance-energy yield plane for two different days under cloudy (left) and clear-sky (right) conditions. From these functions, it is apparent that the relation between energy yield and irradiance may not be expressed in a linear fashion. As confirmed by literature, temperature effects and other more local effects influence this relation. Using this probabilistic approach, we determine the most likely yield for all the PV systems in our register, given the irradiance observed at their locations. The result of this may be seen in Figures 4 and 5.

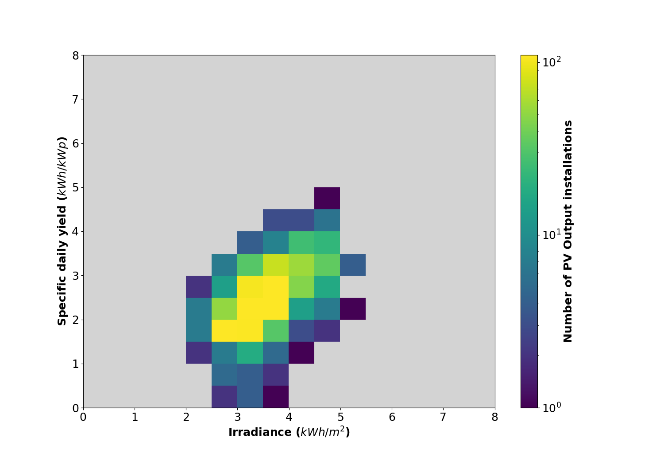
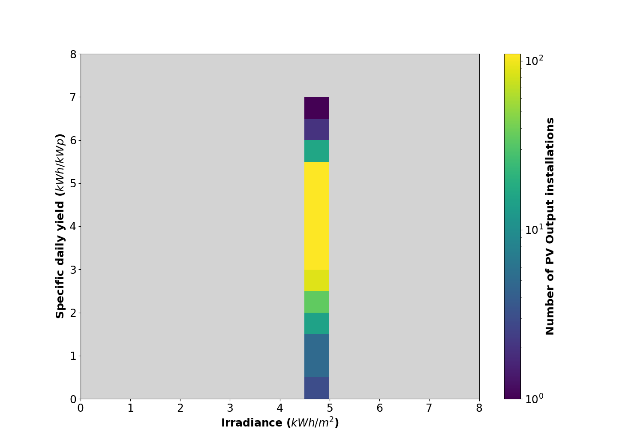
 

Figure 3: 2D plane irradiance x yield for a variable (left) and clear day (right)

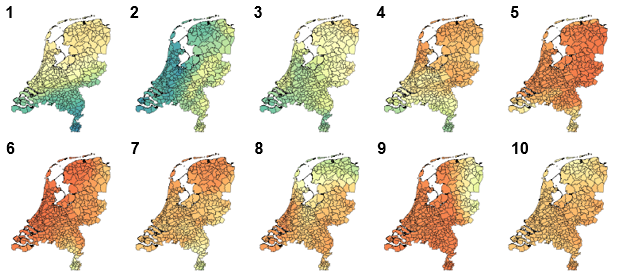
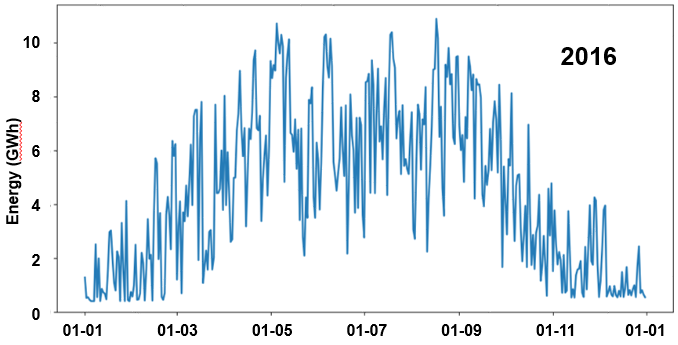


Figure 4: daily national solar energy yield 2016 (left) and daily regional yields for 10 consecutive days (right)

## Representativeness: scenario analysis

Up until now, we have assumed that the pvoutput population is representative of the total Dutch PV population. However we do not know if this is the case. We examined how calculated yield varies by choosing different subpopulations in pvoutput based on their PV specifications. We chose a few factors which are of substantial influence on the yield such as the orientation and tilt. By repeating this method a band width on the yield can be calculated. We calculated 15 scenarios in total and the results confirm the validity of the model. Moreover comparing the results with data on large PV systems (solar farms) and other research confirm the results. For a more detailed explanation of the analyses regarding representativeness we refer to [4].

# Conclusion

New data sources such as data from on-line portals and data from satellites can be used to improve official statistics. In this research we presented a method to combine measurement data from the international pvoutput data portal with modelled irradiance data from satellite grid data to estimate solar energy yields on a daily and regional level. The method could be applied in other countries where pvoutput data (or any other type of PV measurement data) and satellite data on irradiance are available.

# References

1. UN, United Nations. (2019). The paris agreement. United Nations, Bonn. https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement
2. EC, European Commission. (2018). Renewable Energy – Recast to 2030 (RED II) – EU Science Hub - European Commission. Brussels.
3. PVOutput (2020). Latest Outputs. PVOutput. <https://pvoutput.org>
4. Laevens, ten Bosch, Pijpers, van Sark, Observationally daily and regional photovoltaic solar energy production for the Netherlands, Royal Statistical Society (submitted), <https://arxiv.org/abs/2003.01728>
5. Eumetsat (2019, Aug). Welcome to EUMETSAT | EUMETSAT. Accessed 28. Aug. 2019, <https://www.eumetsat.int/website/home/index.html>
6. Schmid, J. (2019, Aug). the SEVIRI instrument. Accessed 28. Aug. 2019,
7. KNMI, Koninklijk Nederlands Meteorologisch Instituut. (2020). <https://www.knmi.nl/over-het-knmi/about>