29th European Photovoltaic Solar Energy Conference and Exhibition,
Amsterdam, the Netherlands, 22-26 September 2014

A BIG DATA APPROACH TO THE SOLAR PV MARKET: DESIGN AND RESULTS OF A PILOT IN THE NETHERLANDS

Bhavya Kausika[1,*], Wiep Folkerts[2], Wilfried van Sark[1], Bouke Siebenga[3], Paul Hermans[4]
[1]*Copernicus Institute of Sustainable Development, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands*
[2]*Solar Energy Application Centre (SEAC), High Tech Campus 21, Eindhoven, The Netherlands*
[3]*I-Real, Stationsweg 30, Terborg, the Netherlands*
[4]*Aurum Europe, Zandsteen 6, Hoofddorp, The Netherlands*
Email: B.B.Kausika@uu.nl

ABSTRACT: In this paper we explore the value of a big data approach to the solar PV market in The Netherlands. As a pilot we choose a specific area on a scale of 200.000 inhabitants. Available data sets from different sources depicting a variety of relevant parameters were brought onto the same Geographical Information System base. We demonstrate how this approach can help in understanding, planning and controlling the energy system.
Keywords: big data, Geographical Information System, PV market, energy production, energy consumption

1 INTRODUCTION

The increasing penetration of solar electricity gives rise to a number of simultaneous challenges. The need for balancing demand and supply may lead to advanced tariff setting systems that go hand in hand with advanced monitoring and forecasting of solar electricity generation. Moreover in order to design the right investment program for the electricity grid infrastructure advanced forecasting of future new installed capacity is needed. A Big Data approach of large connected data sets can contribute in addressing these challenges. McKinsey and company demonstrated the advantages and the future of Big Data for different applications [1]. In the present study we explored the advantage of using Big Data for local electricity demand and supply matching on a scale of 200.000 inhabitants.

For this, we chose an area in The Netherlands defined by the Dutch postal codes 73 and 81 (Figure 1). This area contains one medium size city Apeldoorn and rural area with villages like Vaassen, Epe and Heerde. Ultimately the aim of the analysis is to go down to the EAN codes which show the individual demand points but due to data constraints we stick to the following: the level of spatial resolution varying between household level to 4 digit postal code; as for temporal resolution we choose 1 hour.

The present study intends to present results on the following research questions:

1. How is local generation associated with local demand over time?
2. What are possible strategies to fill in the cumulative supply-demand mismatch?
3. What are appropriate strategies for time resolved balancing?

The state of the art or the innovation lies in collecting and combining data from various sources on a scale that has not been attempted before in the solar energy sector. This would mean creating a value out of huge connected datasets which would have made less meaning when looked at individually.

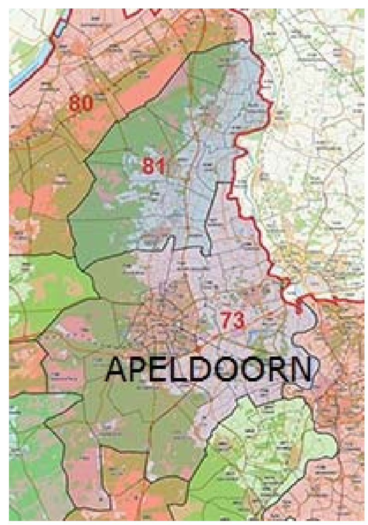


Figure 1: Pilot area; Postal codes 73 and 81 covering the region of Apeldoorn, Epe, Vaasen and Heerde.

2 Methodology

In order to get this study started we began by looking at different data sets that are available and could be used in the project. Since the backbone of the study is the Big Data approach, energy consumption, solar energy and PV capacity data are the most important datasets. Having all this huge amount of data with a spatial entity associated with it would mean much more in terms of visualization, analysis and the outputs. Therefore, a Geographical Information System (GIS) has been opted for handling the data analysis and presentation of results. The next step was to create an interface for data transfer from different systems to a common database for data storage. For this we chose to set up an Application

Programming Interface (API), which would interact among the systems and could access databases and transfer the required data. Then data analysis in terms of demand-supply mapping and study has been conducted. PV industry, consumers and utilities can make use of the results improving their business, increase efficiency of their systems.

In short the following steps have been taken in order to make the Big Data database and analysis.

- Collect data and make meaningful conclusions
- Create clean and useable datasets
- Add spatial entity for visualization and analysis in GIS.

Figure 2 shows the methodology adopted for the study. A detailed description of the type of data used in this research is given below. In addition the methods of acquisition of the data and the challenges with the data have also been discussed.

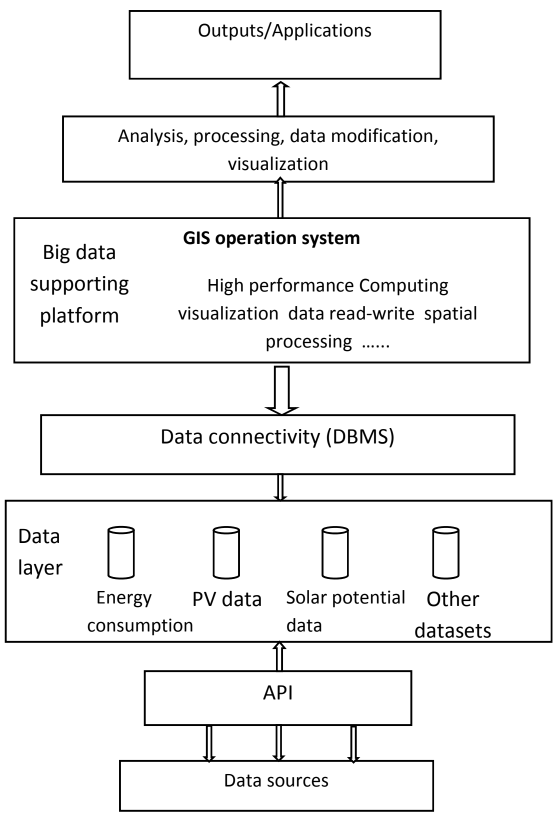


Figure 2: Methodology adopted for the study

2.1 Energy consumption data

The companies involved in the project work on solutions for the balance between the generated renewable energy and the local consumption of energy. They also deliver products and services that help people understand and manage electricity usage. Therefore the energy consumption data for residential and industrial sectors in the pilot area has delivered by them. The challenge here was in integrating the data from different sources onto a common platform and format for analysis. In addition, the data was huge and accumulating by the hour. This also demands a strong and reliable method for collecting, storing, analysis and visualization of these huge databases.

2.2 PV generation data

Installed capacity of PV systems distributed over the Netherlands is provided by many companies freely or the data is published on the internet by some service providers. This data includes installations per postal code, average yield per month, capacity of the PV installation, in few cases the address of the PV installation, orientation of the panel etc. This data base is necessary to keep track of current installations and to manage future progress. The data has been acquired from PIR (production installation register) of Rijkswaterstaat (Ministry of Infrastructure and the Environment), AURUM and SOLAR CARE. Due to privacy issues acquiring this data was a difficult task. There was a mismatch in the level at which this data was available. Neverthelss, it has been made sure that the combined dataset was unique and realistic.

2.3 Other relevant data bases

In addition to the above databases the study also required information on buildings, geographic locations, electricity grid capacity and solar potential information. This information is available from different sources and bringing it all onto one database is a huge task. Solar potential data has been generated with the help of AHN (*Actueel Hoogtebestand Nederland*), the Dutch height map. The necessary building information has been acquired from the Dutch Base Address Register (BAG) and the geographic locations through extensive use of Google maps and ArcGis. A special challenge was to convert all the available data into a spatial database. A spatial database allows visualization in a GIS environment which aids in better analysis.

3 RESULTS

Electricity consumption data for 94 households with hourly resolution has been analysed along with electricity generation with the same resolution for one household. In addition, we also looked into monthly PV generation for 20 households and about 4 households with a daily resolution of electricity generation. The supply demand matching or mismatch has been analysed with the help of above mentioned data.

The total installed capacity in the area was found to be 17.55 MWp according to the data gathered from PIR Rijkwaterstaat. The area has 3828 PV installations. Figure 3 shows the map of the pilot area covering the Dutch postal codes 73 and 81. The map illustrates the total installed PV capacity per postal code in the area.

Mapped locations of PV installations have also been displayed at postal code level.

Figure 4 shows the electricity consumption and the production for a household in the pilot area with solar panels. The peaks in the consumption clearly indicate the use of a heavy appliance while the production line corresponds to the availability of sunlight which is the highest around mid-day for a day in December. A close look at these consumption patterns could give an insight into grid management.

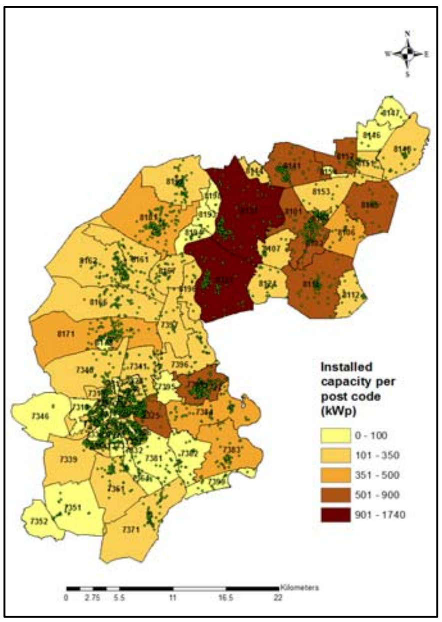


Figure 3: Map showing the total installed capacity per post code in the pilot area.

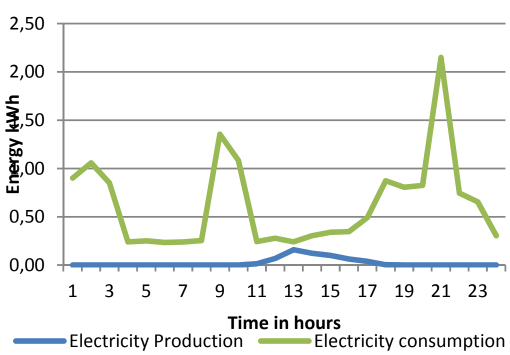


Figure 4. Energy consumption and production in hourly resolution on 12th December, 2013 for a household in Post Code 8131.

Figure 5 shows the production pattern for 3 different capacity PV systems in the pilot area for the whole year. They follow a similar pattern but the amount of power generated differs due to the installed capacity.

Figure 4 shows a scenario: how much electricity could have been generated if a particular house had a PV system installed. The blue line shows consumption pattern for a household with annual electricity consumption for the year 2013, which was 4216 kWh. The red line shows the predicted generation if the household had a PV system of 2.25 kWp capacity. It would have generated about 1767.6 kWh, which is about 42% of the electricity consumed by the household.

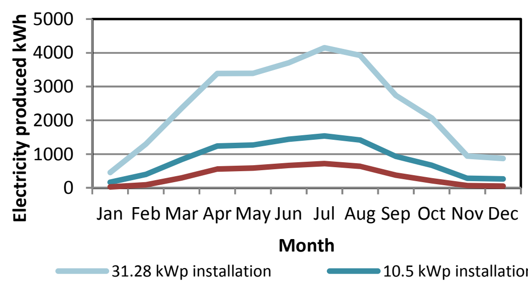


Figure 3: Comparison of 3 different capacity systems in the pilot area for the year 2013.

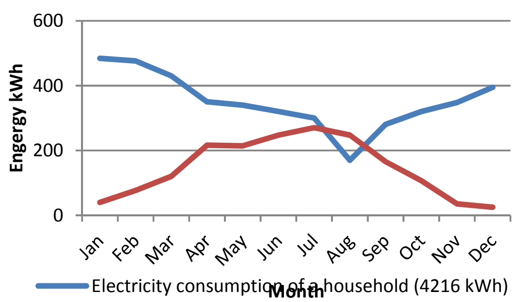


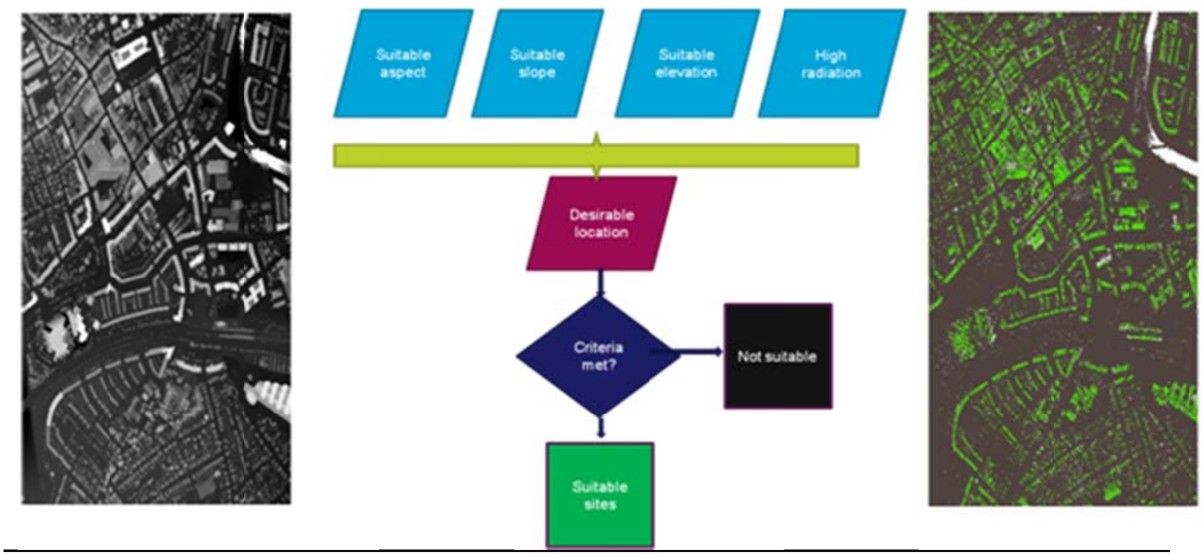
Figure 4: Scenario: if the household had a PV system**.**

In addition we also looked at the potential PV capacity in the area. Figure 5 shows the model used in locating potential PV sites adopted from Chaves and Bahill [2] and (a) shows the AHN image used as an input in the model and (b) shows the final result after the model has been implemented. The green areas in the image depict the suitable sites for PV. The model needs refinement as small patches of unfeasible areas have also been selected like parts of roads, which have high elevation, or areas of roof with gables or chimneys.

4 DISCUSSIONS AND CONCLUSIONS

We arrived at the conclusions presented in Table 1**,** by performing a few calculations based on the data acquired directly and from running the above mentioned model. The calculations were performed on a test tile of 1.25 sq. km. The area feasible for PV has been calculated from the image on the right in Figure 5. Then the potential capacity has been estimated at the rate of 100kWp/m^2. The second row of the table shows the estimation of PV capacity for the whole pilot. We considered only the building rooftops for the calculations. There could be some mismatch as this is only an estimate.

Figure 5: Methodology adopted for selecting potential PV sites. The image on the left is the input and the image on the right the output after model implementation.



	No of buildings	Buildings with installations	Present Installed capacity (MWp)	Potential capacity (MWp)	Present Electricity consumption (MWh)	Potential PV production (MWh)
Test Run on 1.25 sq km	4000	21	0.06	**3.6**	14,000	**3,150**
Total Pilot Area	29,2764	3,828*****	17.55	**430**	950,848.5	**376,250**

Table 1: Results of calculations for finding PV potential (*data acquired from PIR (Rijkswaterstaat, Ministry of Infrastructure and the Environment, Netherlands), AURUM and SOLAR CARE).

We were able to bring for the pilot area data sets from different sources onto the same GIS base i.e. create BIG DATA. This helped us the arriving at the following conclusions. It was found that about 1.61% of the total electricity consumption is covered by PV in the pilot area for the domestic sector. Therefore, there is scope for filling this gap. Grid operators would have to keep track of all the ongoing installations and be able to manage the feed–in and the consumption and consumers will be able to understand and forecast the generation of solar power as a function of place and time which will enable smooth management of supply and demand. We performed a new analysis of PV potential based on the AHN elevation data. The roof potential is not large enough for 100%

coverage of the electricity demand. About 39% coverage can be achieved by rooftop PV. Time resolved balancing will require a combination of demand steering (variable tariff setting) and storage.

We are currently working on running the PV potential model with a few modifications for the whole pilot.

ACKNOWLEDGEMENTS

The authors would like to thank Panos Moraitis for his help with a few python codes and comments in setting up the database. We are also grateful to SOLARCARE for providing data on PV production in the postcode area. This project is financially supported by Netherlands Enterprise Agency RVO.

5 REFERENCES

[1] Manyika, J., Chui, M., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H., 2011. Big data: The next frontier for innovation, competition, and productivity

[2] Chaves, A., Bahill, T.A., 2010. Locating Sites for Photovoltaic Solar Panels - solarsiting.pdf