Polytechnic University of Turin

Thesis for the Master's degree in

Energy and Nuclear Engineering

Solar home battery systems:

analysis of historical time series and system optimization

Supervisors

Filippo Spertino (Turin)

Sylvain Quoilin (Liège)

Candidate

Simone Caprioli

OCTOBER 2018

Abstract

Solar home battery systems: analysis of historical time series and system optimization

The installed PV capacity is growing fast worldwide and it is forecasted that will continue to grow considerably. In addition to this, home battery systems have been improved in the last years and this created high expectation in the sector, since their introduction could make the PV even more profitable for domestic use.

Analysing historical data of 19 households in Germany, equipped with these two technologies, indicators of their profitability will be calculated and discussed to better understand the potentialities of these systems and to assess if they are already economically viable or not.

Keywords: PV, battery, energy, self-consumption, SSR, SCR, Python.

Acknowledgements

A sincere thanks goes to my supervisor Sylvain Quoilin, for giving me the opportunity to work on a subject so interesting and for his precious advices.

I want to express my gratitude also towards any member of the University of Liège teaching staff that contributed to enrich my preparation and made me feel very welcome during my stay as an Erasmus student.

From my home university, the Polytechnic of Turin, a special mention to the professors Filippo Spertino and Vittorio Verda for their passionate way of teaching.

Finally, many thanks to Andrea Tomei, Guglielmo Pelino, Pauline Dupont, Roberto Nitoglia, my family, my friends and everyone that gave its contribution to the realisation of this thesis.

Table of contents

List of figures

List of tables

List of abbreviations

PV: Solar Photovoltaic

AC: Alternate current

DC: Direct current

IEA: International Energy Agency

NaN: Not a number

SQL: Structured Query Language

DSM: Demand side management

MToe: Million tonnes of oil equivalent

LCOE: Levelized cost of electricity

1 Introduction

Photovoltaic systems are an important achievement of last century: The first remarkable solar cell, which is its main component, was developed at Bell Laboratories in 1954 by Chapin et al. [1]; it was silicon based and had an efficiency of 6%.

At the beginning, solar cell technology was used mainly to power space vehicles but in few decades, thanks to important investments done to improve its performances and lowering its costs, had a huge diffusion and found application in different sectors.

1.1 General context about photovoltaic and batteries

The rise of photovoltaic (PV) is not astonishing because it, and more in general the solar energy, is a renewable source and one of the most promising one.

Renewable's growth can be explained considering the energetic-environmental-economic problems of the last decades:

- The world energy consumption has gradually increased and it is supposed to rise more: according to the International Energy Agency (IEA) [2] it amounted to 13'760 MToe[1] in 2016 and it is estimated to expand by around 30% between today and 2040, reaching 17'584 MToe.
- The high rate of fossil fuels (coal, oil and natural gas) exploitation has considerably lowered their availability.
- The pollution due in particular to the use of fossil fuels, which still are the most employed type of energy source, has been linked to several environmental issues.

[1] 1 Mtoe = 11.63 TWh

Today renewables give a consistent contribute to the electricity produced worldwide, though they are not predominant (and even less in terms of total energy consumed since the world consumes more for heat and transportation than to have electricity).

It is estimated that, at the end of 2017, the global renewable generation capacity amounted to 2’179 GW and it was composed as follow: 53% hydro, 23% wind, 18% solar, 6% others sources.

These renewables are not equally distributed across the world: 919 GW are installed in Asia, 512 GW in Europe, 348 GW in North America, 202 GW in South America and 198 GW in the rest of the world [3].

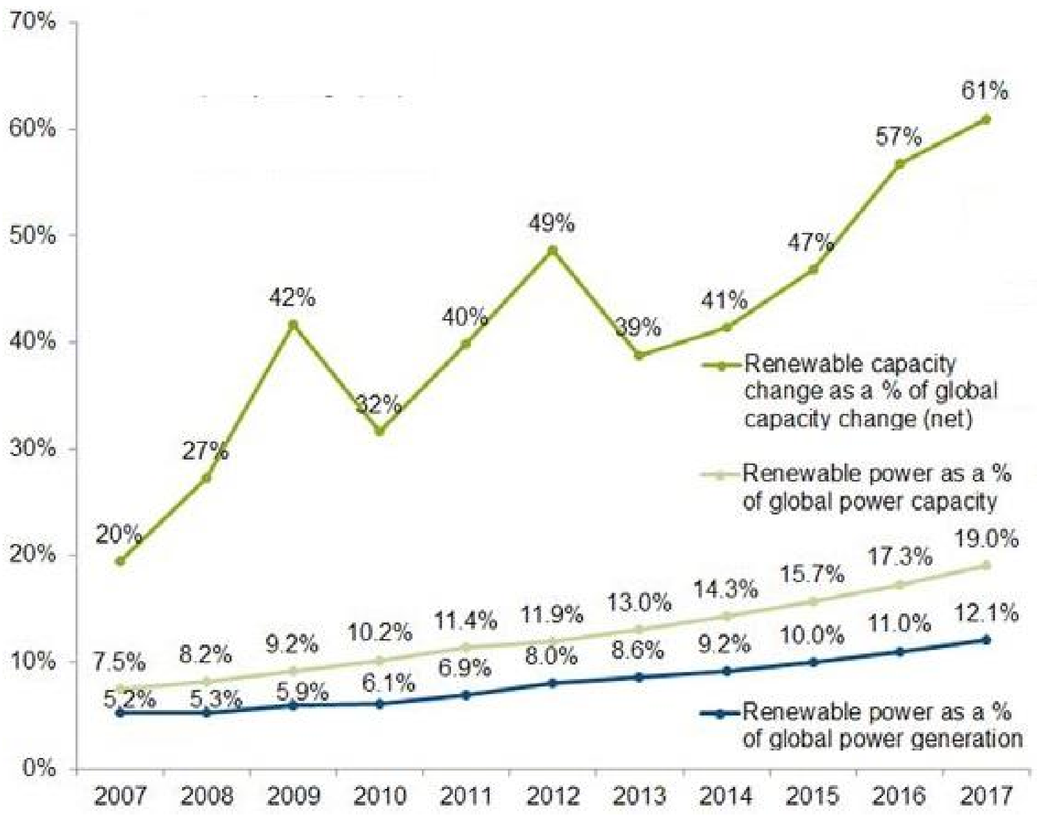


Figure 1. Renewable power generation and capacity as a share of global power [4].

Figure 1 shows the trend of renewables during last ten years:

First line from the top shows the percentage of new installed renewable capacity in the corresponding year compared to all the others plants. It reaches its maximum in 2017: 61%.

Lines in the middle and at the bottom indicates respectively the trends of installed renewable capacity and power generation in the world, which are both upward. These two lines never cross each other. The share of installed capacity is generally higher indeed to compensate that renewable energy sources like solar or wind are not constant and always available, but subject to time variability.

These data, which are all referred to plants dedicated to production of electricity only, evidence that renewables still have a long way to go but are growing fast. In 2017 they beat other sources in terms of installed capacity and share of money invested (265 billion $), with a contribute given mainly by solar and wind.

Solar alone was equivalent to 38% of all the net new power capacity added in 2017 worldwide, while wind contribute was 20% [4].

This shows how important is the PV today. It could be objected that it is not the only technology existent to exploit the solar energy, but the PV is currently the mainstream one: concentrated solar power is still very limited in term of installed capacity, that amounts to 4 GW.

The leader in PV installations is China, followed by the USA, Japan, Germany and Italy. Those countries are not located in the regions with the highest potential for resource availability (Africa and Middle East), but they achieved their goals due to opportune policy and regulatory incentives (see Appendix A).

"Grid parity" has been achieved in several countries and the costs for solar power are continuing to diminish thanks to advancements in technology and the opening of new markets for solar industry in emerging and developing nations [5].

1.1.1 Future trends

IEA tries to evaluate how it will change the energy market in the future; in this paragraph are listed the projections for one of the possible scenarios [6].

Energy consumption, as already said, is supposed to rise by around 30% by 2040 in the world.

This trend is in line with the expected rise of population, that could reach and overcome 9 billion, to which will contribute developing countries in Asia, with India as first.

The growth in the energy demand takes already into account important improvements in energy efficiency, that will play a key role together with natural gas and renewables, while oil, coal and nuclear will continue to grow with a rate much lower than these.

Policies will continue to support renewable electricity, more by means of competitive auctions than feed-in tariffs, and the power sector will change due to investment in distributed solar PV done also by millions of households, communities and enterprises.

It is forecasted that renewables will receive two-thirds of global investment in power plants to 2040 and in that year the share of renewables in power generation will reach 40%. Solar, thanks to a rapid diffusion of PV, promoted by China and India, will become the largest source of low-carbon capacity.

Also in the European Union the renewables will continue to grow considerably (they will represent 80% of new capacity installed) and wind will become the first electricity source after 2030.

Another important change will be the growth of electricity share in final consumption of almost a quarter by 2040, due also to the increased employ for heat and mobility (according to IEA’s projections there will be 280 million of electric vehicles worldwide).

1.1.2 Batteries for the residential PV market

Batteries can improve the performances of PV systems that, without a storage system, cannot regulate their production to match the demand curve, but only be disconnected when it is opportune in order to stop their production.

Recently home batteries are becoming more popular, especially due to price reductions in this sector, and their further develop could support the diffusion of PV across the world, promote the micro-grids and also lead to autonomous households [7].

Despite they are already available for the residential market, these storage systems are still too expensive and need subsidies like feed-in tariffs, favourable net metering schemes or green certificates to be profitable in the European countries [8]. For this reason, their diffusion in this field it is still limited and under evaluation.

From the prosumer (term used to indicate that the owner of the PV battery system is a consumer and a producer of electricity at the same time) point of view, the benefit is given by savings to fulfil his electrical demand: they tends to become higher if higher is the share of the self-consumed energy, thanks also to exemptions (generally at least partial in the European framework) from taxes regarding the use of the electrical grid, that is financed by its users for its maintenance and development.

It is clear that this situation can put in discussion the distribution of shared grid costs into an energy market because, as the prosumers become more predominant, this change could lead to an unfair situation for the other consumers fully depending by the grid, that would be forced to pay more if the costs are not redistributed in an alternative way.

1.2 Aims and objectives of this thesis

The main goal of this work is the evaluation of Solar Home Battery Systems:

The term is referred to PV for domestic use (generally mounted on the rooftop of a residential building) coupled with batteries.

Results obtained and considerations will be based on the analysis of a database, which contains historical data of these systems, gently put at disposal by the University of Liège.

The analysis will be performed using Python, a powerful programming language that currently find application also as a data analysis tool.

Thanks to the information contained in the database object of studies, some actually existent solar home battery systems will be described, discussed and compared to two corresponding hypothetic cases:

In the first any contribution given by the battery to the system will not be considered (case without battery/stand-alone PV system).

In the second the controller that regulates the energy flows of the system is substituted with an ideal one that follows a precise dispatch strategy (case optimized).

1.3 Recent studies on Solar Home Battery Systems

The classical system considered in this thesis is illustrated in Figure 2: it consists of a DC-coupled PV and battery system that covers partially the consumption of one household and sends electricity in excess to the grid.

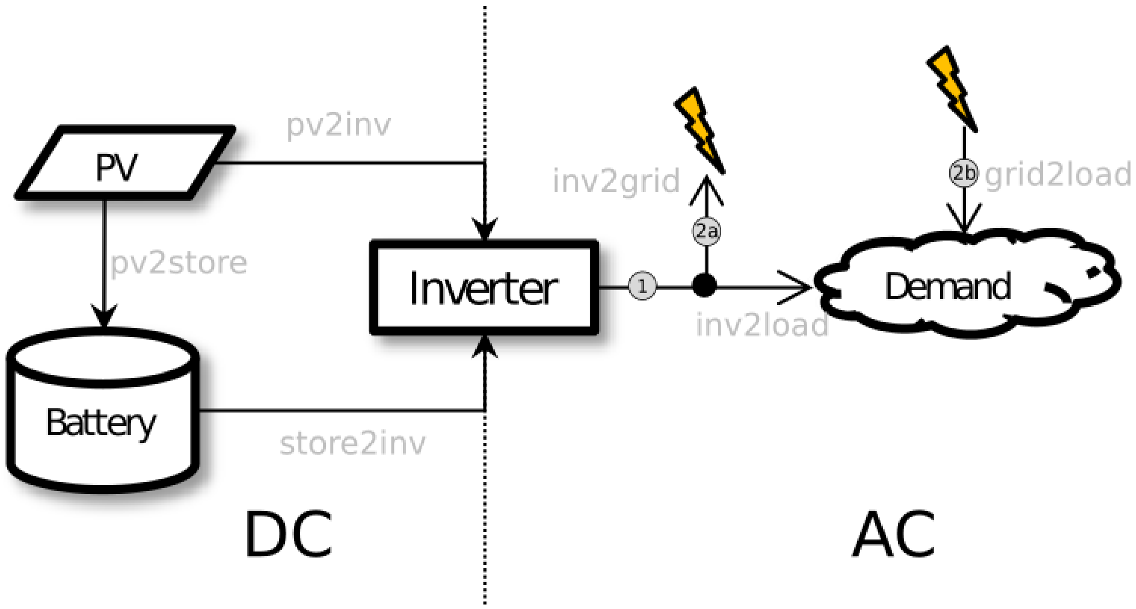


Figure 2. Energy flow chart of a typical solar home battery system [9].

All the fundamental aspects concerning components and functioning of PVs and Lithium Ion batteries (the ones most used for this kind of application) are not presented in this work because they are considered to be already known by the average reader.

In this paragraph we will instead focus on the recent studies effectuated to evaluate the profitability of these systems.

According to [10], in Germany, investments in battery storage for small residential PV systems were already profitable in 2013, without policy support, but for that study the battery investment cost was calculated by adding up the energy and power cost of 171 €/kWh and 172 €/kW respectively.

That cost assumption seems to be really low: reference [11] estimates that in 2015 the storage price was around 500 EUR/kWh and the investment resulted to be not profitable, since it would have been needed a cost below 450 EUR/kWh.

In accordance with this, another study of 2016 [12] states that these systems still needed subsidies and increasing retails price of electricity to become economically viable in Germany.

In evaluating the profitability, it is necessary to determine the volume of self-consumption. Once that is known, it is possible to calculate the self-sufficiency rate (SSR) and the self-consumption rate (SCR), two important indicators that will be described in Section 4.1.6. of this thesis.

Self-consumption depends by the system design: [13] shows that, varying the size of the battery from 0 to 16 kWh, the SSR varies from 30% to 66% in winter and from 48 to 98% in summer; in that case, a further augment of the capacity was clearly inconvenient because even with 32 kWh the SSR in winter was still 66% (while reached 99% in summer).

In order to evaluate it correctly, it is important to have data of good quality concerning the profiles of consumption and production.

Nevertheless, several previous studies are based on models and historical data that are “aggregated”:

In fact, they are relative to the profile of more household or characterized by a low time-resolution, like for instance of a value for each day, that therefore smoothen out the variability of the individual profiles or the variability over time.

This thesis instead, to obtain results as accurate as possible, will follow the same approach of [7], a study based on historical disaggregated data with high time-resolution, relative to households located in different European countries.

2 The Python language and other tools

What is Python, the main tool of this thesis? By definition: *“Python is an interpreted, object-oriented, high-level programming language with dynamic semantics”* [14].

Since its first appearance in 1991, Python has become one of the most popular programming languages, together with Perl, Ruby, and others [15]. It has become famous to create websites using its different frameworks and it is sometimes called “scripting language” since it can be used to write quickly small programs, or scripts, and it can be used also as a glue language to connect existing components together.

Adoption of this interpreted language for scientific computing in industry applications and academic research has raised considerably since the early 2000s. Actually Python is supported by a large and active scientific computing community and it is freely distributed; this gives a nice boost to its future additional diffusion and improvement.

Others reasons to justify its popularity are that its supports many scientific modules and packages, that are also distributed for free, its syntax is easy to read and, least but not last, a program written in Python is easy to debug.

In the other hand, since Python is an interpreted programming language, usually most of its code cannot be fast as code written in a compiled language like Java or C++. This is an aspect to consider if the application of interest requires high requirements in terms of process time performance at the expense of more time spent in programming.

To pursue aims of this thesis the learning of Python’s basis has been essential and additional time has been devoted to acquire the skills useful to employ it as data analysis tool.

Concerning this last application (data analysis and exploratory/interactive computing and data visualization), it is possible to remark that in last years, thanks to the improvement of its library support (mainly due to Pandas’ introduction and development), it has become an important alternative for data manipulation tasks among other dedicated languages and tools.

As mentioned Python supports numerous additional modules and packages, built to extended its basic functionalities, and it is often employed the term “library” to indicate a collection of more modules.

In this chapter will be illustrated the main libraries of Python and other tools that have been useful to reach the goals of this thesis and that in general are suitable to explore and analyse a database.

2.1 NumPy

NumPy stands for Numerical Python and it is the basic package for scientific computing in Python. It lets to use [15]:

- A fast and efficient multidimensional array object: “*ndarray*”.
- Functions for performing element-wise computations with arrays or mathematical operations between arrays.
- Tools for reading and writing array-based data sets to disk.
- Linear algebra operations, Fourier transform, and random number generation.

NumPy arrays (an array is the collection of elements of a single data type) are important concerning data analysis because they are the most efficient Python data structures to store and manipulating numerical data.

2.2 Pandas

Pandas (name derived from panel data and Python data analysis) is a Python library that adds objects and functions conceived to work easier with structured data and it makes of Python a suitable environment for data analysis.

The primary object in pandas is the "*DataFrame*": a two-dimensional, tabular, column-oriented data structure.

Pandas combines the high performance array-computing features of NumPy with the flexible data manipulation capabilities of spreadsheets and relational databases. It includes sophisticated indexing functionality to facilitate classical data analysis operation such as to reshape, slice and dice, perform aggregations, and select subsets of data [15].

2.3 Matplotlib

Matplotlib is the Python library most commonly used to create plots and other 2D data representations. It is possible to choose among different types of plots and styles and the quality of figures so created is good enough for scientific publications.

2.4 Spyder

Spyder stands for "Scientific PYthon Development EnviRonment" and it is one of the possible IDEs (integrated development environments) for Python. It is compatible with Python libraries and it provides advanced editing, interactive testing, debugging and other useful features, such as the "variable explorer".

2.5 MySQL

MySQL is an open source Relational Database Management System (a particular type of software that makes possible the interaction between the user and the database of interest) based on Structured Query Language (SQL).

SQL is the most popular language for adding, managing and retrieving data in a database. It lets to perform these operations quickly and it is reliable.

3 Data management of the database Speicherdata

3.1 Speicherdata's content and structure

The database to analyse, own by the University of Liège, is called Speicherdata ("*Speicher*" means memory in German) and it contains measures (mostly electrical measures), regarding the performances of 19 houses equipped with photovoltaic panels and electrical batteries.

Together with these data were not given additional information, except that all these houses were located in Aachen (a city in the west of Germany, close to the borders with Belgium and Netherlands).

To access to the database and its content was quite immediate since it was stored on MySQL; the access was effectuated through remote desktop connection to exploit a computer of the university already configured to communicate with the database.

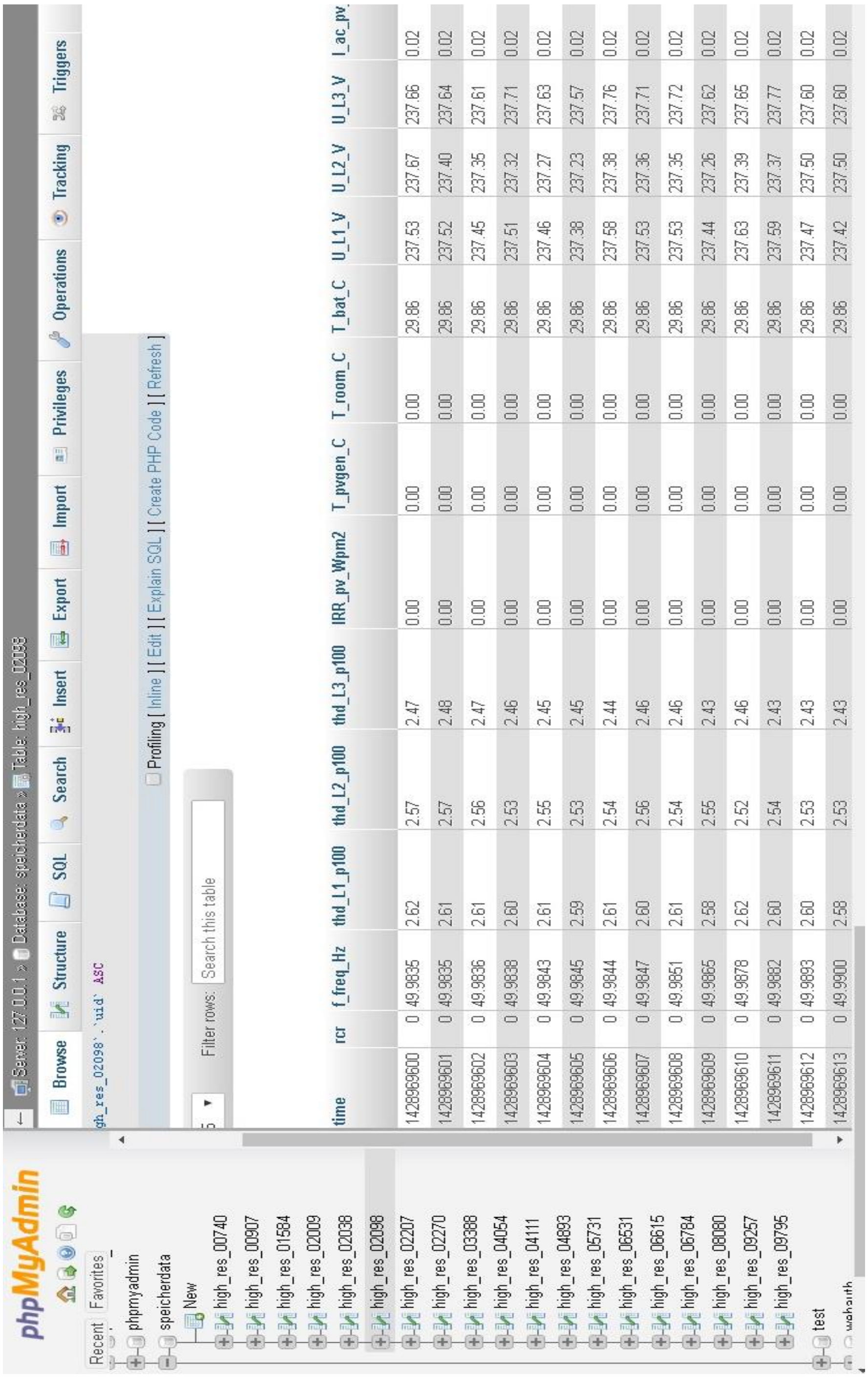
It was immediate, once logged into MySQL, to realize how big it is the database Speicherdata:

It occupies in total 80.5 Gigabytes and it is divided in 19 tables. Each table represents one household, it is distinct by a numeric code and has a size that varies from a minimum of 0.4 to a maximum of 8.5 Gigabytes.

This difference in size among tables is especially due to the fact that the period between the first and last measure effectuated is not the same for each table, while the period between two consecutive measure is always the same, as it will be better clarified in this chapter.

In Appendix B it is possible to see the first screen of Speicherdata; it can be noted also that there is correspondence, almost proportional, between the numbers of rows contained in a table and its dimensions.

In the next page instead it is shown the screen that appears when one of the tables is selected, that provides simply a partial view of its measures:



time	rcr	f_freq_Hz	thd_L1_p100	thd_L2_p100	thd_L3_p100	IRR_pv_Wpm2	T_pvgen_C	T_room_C	T_bat_C	U_L1_V	U_L2_V	U_L3_V	I_ac_pv
1428969600	0	49.9835	2.62	2.57	2.47	0.00	0.00	0.00	29.86	237.53	237.67	237.66	0.02
1428969601	0	49.9835	2.61	2.57	2.48	0.00	0.00	0.00	29.86	237.52	237.40	237.64	0.02
1428969602	0	49.9836	2.61	2.56	2.47	0.00	0.00	0.00	29.86	237.45	237.35	237.61	0.02
1428969603	0	49.9838	2.60	2.53	2.46	0.00	0.00	0.00	29.86	237.51	237.32	237.71	0.02
1428969604	0	49.9843	2.61	2.55	2.45	0.00	0.00	0.00	29.86	237.46	237.27	237.63	0.02
1428969605	0	49.9845	2.59	2.53	2.45	0.00	0.00	0.00	29.86	237.38	237.23	237.57	0.02
1428969606	0	49.9844	2.61	2.54	2.44	0.00	0.00	0.00	29.86	237.58	237.38	237.76	0.02
1428969607	0	49.9847	2.60	2.56	2.46	0.00	0.00	0.00	29.86	237.53	237.36	237.71	0.02
1428969608	0	49.9851	2.61	2.54	2.46	0.00	0.00	0.00	29.86	237.53	237.35	237.72	0.02
1428969609	0	49.9865	2.58	2.55	2.43	0.00	0.00	0.00	29.86	237.44	237.26	237.62	0.02
1428969610	0	49.9878	2.62	2.52	2.46	0.00	0.00	0.00	29.86	237.63	237.39	237.65	0.02
1428969611	0	49.9882	2.60	2.54	2.43	0.00	0.00	0.00	29.86	237.59	237.37	237.77	0.02
1428969612	0	49.9893	2.60	2.53	2.43	0.00	0.00	0.00	29.86	237.47	237.50	237.60	0.02
1428969613	0	49.9900	2.58	2.53	2.43	0.00	0.00	0.00	29.86	237.42	237.50	237.60	0.02

Figure 3. Table “high_res_02098” in MySQL.

Each table has rows and columns; each columns has a label that refers to the type of measure contained. For example, the column with label "*T_bat_C*" contains the measures of the battery's temperature in Celsius degrees.

The number of columns/measures is not exactly the same among tables and it is usually around 50. Of these, several measures won't be interesting to analyse and therefore a selection will be done.

The fact that the measures contained in each table are not exactly the same it is not a big problem since the most important measures are contained in almost every table and the convention used to write labels is always the same (name of the measure in English abbreviated plus unit of measure).

A key measure, also useful to understand better how are structured these tables, is the time, contained in the column labelled *"time".*

As displayed in figure, the time is recorded in a numeric format: The Unix Epoch.

According to this format, every number (of 10 digits in our case) represents the number of seconds that have elapsed since the midnight of 1th January 1970.

The time conversion is immediate with Python or other programs such as Epoch Converter (freely available on the web) therefore we know, for example, that the first measure of the table considered, was collected the 14th April 2015, at 12:00:00 AM (GMT).

Looking the "*time*" column we note that each number written in the Unix Epoch format is equal to the previous value plus one and so it is straightforward to conclude that measures were collected with a time step of 1 second, or that in other words each row contains the values collected in a precise instant of time and that the database has a resolution of 1 second (even if data are not perfect as it will be explained).

3.2 Data processing: download

In order to work with a database of these dimensions, the first preliminary step was to arrange the data in a more suitable way:

In fact, even if Python, through a toolkit called SQLAlchemy, can directly interact with the database on the MySQL platform to read and process the values there contained, this was highly inadvisable since it would have slowed down considerably the next steps.

The best way to proceed, that made unnecessary a further communication between Python and MySQL, was to download the data on the computer, in a suitable format.

Before to download the database, some considerations were done:

According to common sense, it looked useless in this case to keep unaltered the original high resolution of the database that, as already said, contains measures effectuated with a time step of 1 second.

This especially for two reasons:

- In this analysis we are interested in results evaluated on time intervals much higher of one second, usually comprised between one hour as minimum and one year as maximum, so considering a lower resolution should not affect much the results.
- The precision of each measure is unknown and this discourage from evaluating the errors and following standards of precision really strict.

About the first point, it find confirm in reference [7]: this article demonstrates that a time step of 15 minutes is good enough to carry out data analysis of this kind.

Another concern before to proceed with the download was the following: all the 19 tables contain values of interest or not?

To answer it was useful to run a script on Python (results are displayed in the following figure).

	high_res_00740	high_res_00907	high_res_01584	high_res_02009	high_res_02038	high_res_02098
first	2015-11-12 14:49:34	2016-09-13 02:00:00	2015-10-06 02:00:00	2015-10-09 02:00:00	2016-08-17 02:00:00	2015-04-14 02:00:00
last	2016-10-24 01:59:59	2016-10-24 01:59:59	2016-10-24 01:59:59	2016-10-20 01:59:59	2016-10-24 01:59:59	2016-10-17 01:59:59
delta	346 days, 11:10:25	40 days, 23:59:59	383 days, 23:59:59	376 days, 23:59:59	67 days, 23:59:59	551 days, 23:59:59
delta_s	29934625	3542399	33177599	32572799	5875199	47692799
Nrows	29350983	2299190	26782815	26020538	5874540	44949404
Completeness	0.980503	0.649049	0.807256	0.798843	0.999888	0.942478
Completeness*	0.980503	0.649049	0.797806	0.818137	0.999888	0.956617

	high_res_02207	high_res_02270	high_res_03388	high_res_04054	high_res_04111	high_res_04893
first	2015-07-11 02:00:00	2016-01-31 19:11:10	2015-12-21 14:58:26	2015-07-17 12:54:35	2015-07-11 02:00:00	2016-05-24 02:00:00
last	2016-10-24 01:59:59	2016-10-24 01:59:59	2016-10-24 07:59:59	2016-10-24 01:59:59	2016-10-22 16:19:54	2016-10-24 01:59:59
delta	470 days, 23:59:59	266 days, 6:48:49	307 days, 17:01:33	464 days, 13:05:24	469 days, 14:19:54	152 days, 23:59:59
delta_s	40694399	23006929	26586093	40136724	40573194	13219199
Nrows	30115014	23002718	26567902	36505517	31107623	12527924
Completeness	0.740028	0.999817	0.999316	0.909529	0.766704	0.947707
Completeness*	0.908616	0.999817	0.999316	0.972624	0.741891	0.947707

high_res_05731	high_res_06531	high_res_06615	high_res_06784	high_res_08080	high_res_09257	high_res_09795
2015-07-18 02:00:00	2015-12-08 18:32:00	2015-10-29 14:21:09	2016-07-15 02:00:00	2016-08-12 02:00:00	2016-07-07 02:00:00	2016-06-22 02:00:00
2016-10-10 01:59:59	2016-10-24 01:59:59	2016-10-22 15:53:29	2016-10-24 01:59:59	2016-10-24 01:59:59	2016-10-24 01:59:59	2016-10-17 08:53:48
449 days, 23:59:59	320 days, 7:27:59	359 days, 1:32:20	100 days, 23:59:59	72 days, 23:59:59	108 days, 23:59:59	117 days, 6:53:48
38879999	27674879	31023140	8726399	6307199	9417599	10133628
28069929	22659699	30697623	8726383	4406306	9417600	10133018
0.721963	0.818782	0.989507	0.999998	0.698615	1	0.99994
0.644628	0.818782	0.989507	0.999998	0.698615	1	0.99994

Figure 4. Information about Speicherdata's tables.

Here each column contains 7 values relative to one table; in order there are:

- "First": starting date of the measures.
- "Last": ending date of the measures.
- "Delta": difference of time between these two extreme dates (*last* - *first*).
- "Delta_s": the previous value (*delta*) expresses in seconds.
- "Nrows": total number of rows in the table.

- “Completeness”: The ratio *Nrows*/*Delta_s*. If it is equal to 1 the completeness of the table is considered maximum (100%) since in the ideal case it should contain one row (set of measures) for each second.
- “Completeness*”: It is like the *Completeness* but it is evaluated only on the last 365 days if *delta* it is higher than one year. We are considering the last 365 days and not the first because the ending date doesn’t change much from table to table (making the last measures more interesting for our purposes), at difference of the starting date.

Looking at these values it emerges that tables are different in terms of quantity of data and also of quality.

Of 19 tables, 7 were discarded, according to this preliminary analysis, because they could provide information about a period shorter than 200 years.

Since we would like to calculate yearly indicators in our analysis, it was in fact advisable to select tables characterized by a period not too shorter than one year and a percentage of completeness as high as possible, in order to obtain good results.

The data contained in the remaining tables of Speicherdata were instead elaborated and then saved on the computer:

Using Python in combination with MySQL, the values contained in each table of interest were read at groups of 900 at a time, then averaged and saved to obtain 12 corresponding tables (Panda’s dataframes to be more precise) with a resolution of 15 minutes instead of 1 second (so much smaller but still good enough as previously explained).

This process took hours and it was the only one expensive in term of process time. Done this, it was possible to collect all the information in one file:

As format pickle (a serialization format of Python) was considered the most convenient choice. The file resultant was easily accessible through Python and had a size of 195 Megabytes.

3.3 Further data's manipulations

At this point, even if data were easily accessible (less than 1 minute required to load all the rearranged database) and apparently ready to be used, it was still necessary to pay more attention to the quality of the data and to do additional steps before to start the true analysis.

3.3.1 NaN values

In the tables, due to the processing done to transfer data from MySQL to Python, is now present a DateTimeIndex: a column with dates in chronological order, in which the first corresponds to the time of the first set of measures available for the selected table/house, the next is equal to the first plus 15 minutes (the new time step) and so on until the last measure available for that table.

Nevertheless, as we already could expect when we noted that the numbers of the rows in the tables is generally lower than the expected value, for some dates that compose the DateTimeIndex, instead to have the corresponding numeric values, it appears repeatedly the acronym "NaN" that stands for "not a number".

This since even in the original version of the data there are not measures associated to certain periods: it is like if the recordings were interrupted for some days (some weeks in the worst cases) before to start again.

To compensate this, instead of leaving holes, the best thing to do was to fill all the NaNs using values as realistic as possible.

It was chosen to fill the NaNs with the values of the previous day and, in case that it was missing more than one day of measures, half of the missing days were built using the first day available going backward and the other half with the first day available going forward.

This modify was effectuated for all the tables running a Python script, that was also useful to check if the number of NaNs was well described by the complementary percentage of the completeness calculated before.

It was unavoidable to introduce an error by proceeding with the filling but, since the system object of study usually do not change behaviour considerably during few consecutive days, we expect that the error introduced is not relevant at condition that the number of missing data was not excessive.

Considering this, it was chosen to discard all the tables with a completeness lower than 80% and consequently the number of tables to study decreased to 10.

3.3.2 Null values

Observing the data, it was noted that some particular measures had only null values. This means that zeros were used to indicate the lack of information.

It was important to check if it happened also for the measure of interest, and therefore through a script was calculated the percentage of null values for each measure, for all the tables.

For some measures (like for instance the power of the battery or the power of the PV) it was normal to have a lot of zeros and so it was not trivial to make the distinction.

In these uncertain cases it helped to graphically evaluate the distribution of the null values:

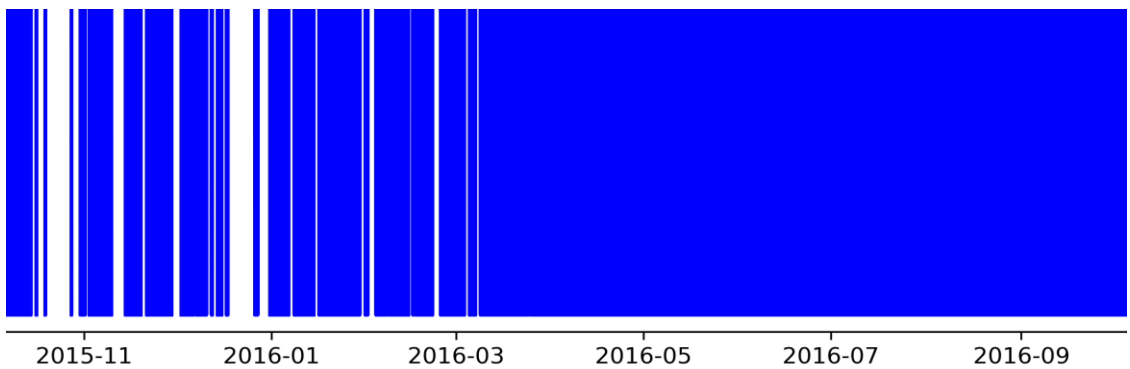


Figure 5. Power battery null values distribution in “high_res_01584”.

This simply plot gives a representation of the periods in which the power of the battle is predominantly null for the considered table, that appear as not coloured areas.

It was immediate to conclude that the null values are more frequent during the winter months, as we expected since during winter the battery is less used (because the PV produces less electricity). In addition to this the null values were never concentrated on periods longer than 10 days and so this distribution was considered acceptable.

In general, the problem of null values was not so relevant since it was more related to not useful measures. Anyway other tables were discarded because of that, like the table high_res_06615 that had null values of consumption for a whole month.

3.3.3 Reindexing and final adjustments

Checked the quality of the data during the previous steps, the tables still had a variable length in term of time: the average was similar to one year, with some tables much shorter and a few of them closer to one year and half.

In order to obtain better results, it looked convenient to be a little more restrictive in the selection and so at the end it was chosen to work only with the following 5 tables:

- High_res_01584
- High_res_02009
- High_res_02098
- High_res_02207
- High_res_04054

These tables were considered the most suitable to be analysed since they have a completeness equal or higher to 80%, a length of at least 365 days and they contained all the most important measures.

It was possible to select a period of one year common to all these 5 tables to work with:

The period comprised between the midnight of the 15 October 2015 (included) and the midnight of the 15 October 2016 (not included).

Therefore, the tables were reindexed to take this into account. This scheme can be useful to better understand the new structure common to all 5 the tables:

Index	(label of measure 1)	(label of measure 2)	…	(label of measure n)
2015-10-15 00:00:00	(value of 1 at that time)	(value of 2 at that time)	…	(value of n at that time)
2015-10-15 00:15:00	(value of 1 at that time)	(value of 2 at that time)	…	(value of n at that time)
…	…	…	…	…
2016-10-14 23:45:00	(value of 1 at that time)	(value of 2 at that time)	…	(value of n at that time)

Table 1. Structure of one table in Python (Type of object=Pandas' Dataframe).

To end tables were renamed keeping the last four digits of their name, preceded by the letter "t" (ex: "t1584" instead of "high_res_01584"), and about half of their columns were deleted because they contained measures not useful.

At this point each table had 35136 rows (corresponding to 366 days because during that period is comprised the 29th February 2016) and a number of columns comprised between 22 and 28.

All that information, that will be used to perform the true data analysis, was stored in a pickle file of 39 Megabytes. A really small one if compared with the huge dimensions of the database Speicherdata.

4 Data analysis

The five selected tables, containing data (gathered during one year) of 5 different households, equipped with PVs and batteries, could finally be analysed.

The procedure is more or less the same for each table and, to avoid repetitions, the table t4054 has been chosen as reference. Computations and plots are done using Python and its libraries.

The most important measures at disposal are:

- Power consumption of the house (AC) [W]. It's always positive.
- PV production before the inverter (DC) [W]. It's always positive.
- PV production after the inverter (AC)[2] [W]. It's always positive.
- Power entering or exiting the battery (DC) [W]. It's positive when the battery is discharging and negative during the charge.
- Irradiance [W/m^2].

4.1 Computations and estimations

4.1.1 House consumption

Since we know the average power consumed by the house and its corresponding date, it is easy to evaluate the energy consumed on a period of choice (daily, monthly, etc…).

In general:

$$E = \int_{t_i}^{t_f} P(t)\, dt$$

Where *E* is the energy, *P* the instantaneous power and *t* the time.

[2] In few cases this measure is not available. To overcome this the inverter's efficiency is considered equal to the average value of the other tables.

In our case the integral becomes a summation and it is important to keep in mind that it does not refer to 1 second (SI unit) but to 15 minutes (time step for each data after the data management phase).

Considering this, to calculate the energy consumed during a period *X*, expressed in kWh, we have that:

$$E = \frac{\sum_{i=1}^{n} P(i)}{4}$$

Where *n* is the number of measures contained in X.

If we want to evaluate the yearly energy consumed, n is equal to 4 (4 measures for hour), multiplied by 24 (24 hours for day), multiplied by 366 (days contained in the period considered). So in this case n is equal to 35’136.

For table t4054 the result is 6’135 kWh. This and other results will be shown and discussed in the next chapter.

4.1.2 PV production

The yearly energy produced by the PV of each house is calculated in the same way and for the reference it amounted to 5’781 kWh (value before the inverter).

If it is present also the measure of PV power after the conversion in AC (missing for some tables), the inverter’s efficiency (μ_{inv}) can be calculated considering the simple relation:

$$P_{pv_{ac}} = P_{pv_{dc}} * \mu_{inv}$$

It is also interesting to evaluate the relative PV size, given by the ratio of the energy yearly produced by the PV and yearly consumption of the house.

In average it was equal to 1.2 for the five households.

4.1.3 Peak power

The peak power has been estimated for each PV installation, despite the fact that we do not have proper information, such as the tilt angle and the orientation of the panels, and we do not know neither how it was exactly taken the measure of irradiance.

For this estimation it was taken the maximum power registered with a temperature of the panels equal to 25°C or higher, divided by its respective irradiance and multiplied by the value of irradiance in the standard conditions (1000 W/m^2).

We obtained values comprised between 6.7 and 12.5 kW_p.

As next step it was interesting to calculate the ratio between the energy yearly produced by the PVs and their peak power. In average the result was equal to 869 kWh/kW_p, with a maximum value of 942 kWh/kW_p.

Confronting these values with a reference [16], they look realistic for the area around Aachen where the correspondent optimal value should be slightly lower than 1'000 kWh/kW_p.

Among the PVs, it was especially one to keep the average low with a value of 766 kWh/kW_p, and this can be caused by different reasons like errors in the data, not optimal installation of the PV or low efficiency of the solar cells (for example due to ageing or scarce maintenance).

4.1.4 The battery

About the battery we know the corresponding power, positive when the battery is discharged and negative when it is charged. Its maximum value is about 3 kW for each battery, except for t4054's one that has a maximum power of 2 kW.

With these data is possible to calculate the energy exchanged by the battery during a certain period, as seen before for the consumption.

The ratio of these two values corresponds to the efficiency of the battery in that period (since in an ideal case all the energy stored in the battery can be released, while in reality it is unavoidable to have some losses, like losses due to Joule's effect for instance).

For three tables the battery's efficiency so evaluated is high (between 94 and 97.5%) and for the other two is considerably lower (about 76 and 81%). This difference could be due to errors in the measures, hypotheses supported by the fact that one of these batteries results to be in state of charge also when the PV production is null (it happens for t2098).

Going back to energy yearly discharged, it amounted to 1'672 kWh (of DC) in average.

Another important data of the battery that can be evaluated is its available capacity:

It is considered to be equal to the energy associated to the maximum consecutive power flow of discharge (situations in which the battery discharges a little before to be charged again are not considered), during the whole year of measurements.

The average value of capacity its about 9.5 kWh, with a minimum of only 2.2 kWh for table t4054.

A data correlated to this is the relative battery size that is the ratio between the battery's capacity in kWh and the yearly house consumption expressed in MWh.

Also in this case the minimum corresponds to table t4054 with 0.4 kWh/MWh against an average of 1.3 kWh/MWh.

Even if that table is equipped with the smallest battery, it will be the one of reference to create plots because it is considered to have the most trustworthy measures.

4.1.5 The grid

Like it happens in vast majority of cases, about solar home battery systems, the five households are not independent from the energetic point of view: despite they can produce energy with PVs, they still exploit the local electric grid to fulfil their electric consumption.

Generally, the house-owner try to cover its consumption exploiting as much as possible the electricity that his PV can produce, to save moneys that are also required to recover the cost of his investment.

Energy is bought from the grid only when it is necessary:

Thanks to the presence of the battery he can store the surplus of electricity produced into the battery that can behave like a secondary source of power, to exploit when the demand exceeds the production.

At the same time if the production is higher than the consumption, since the amount of energy that can be stored is limited by the capacity of the battery and since it does not exist a storage system that does not comport losses, he can send his surplus to the grid. He usually receives a compensation lower than the average price that pays to buy electricity and, if in that moment the line is congested, he could instead be forced to pay for this choice.

In the database Speicherdata are not included measures of the power extracted or injected into the grid (that would have been also useful to evaluate the precision of other measures, calculating the residual of the following balance), but these values can be calculate by difference, considering this balance of power:

$$P_{grid} = P_{consumed} - \left(P_{produced(dc)} + P_{battery(dc)}\right) * \mu_{inv}$$

Where all terms respect the convention seen before; they are all positive except the power of the battery and the power of the grid. This last term will be positive if the electricity goes from the grid to the house, and negative in the reverse situation.

Once that also the power of the grid was known (along the full period with a time step of 15 minutes), it was straightforward to consider separately its positive and negative values to calculate, respectively, the electricity received from the grid during one year (2’917 kWh in average) and the electricity sent to the grid (4’051 kWh in average).

The same procedure was repeated to consider the case without battery (considering the power of the battery always null in the balance).

As it was expected, in this case the energy flows exchanged with the grid raise: even if consumption and production do not change, the system loses flexibility and as consequence it becomes more dependant from the grid to balance the mismatch between production and consumption during each period.

4.1.6 Self-consumption, SSR and SCR

We use the word “self-consumption” to put in evidence the amount of energy consumed by the house that was made available by the PV or the battery.

By definition it is equal to the consumption when the house does not receive energy from the grid and it is lower in the other cases.

It can be evaluated for each time step using this expression:

$$P_{self\ consumed} = \min[P_{consumed}\ , \left(P_{produced(dc)} + P_{battery(dc)}\right) * \mu_{inv}]$$

In which “min” is the operator “minimum” that selects the smallest value between the power consumed and the other term separated by the comma.

The energy yearly self-consumed is calculated in analogy with the energy consumed through a summation; the result is 3’003 kWh for the table t4054.

To visualize better the concept of energy self-consumed we can use two important indicators:

- The self-sufficiency rate (SSR) is the ratio between the self-consumed energy and the energy consumed by the house in a period (usually of one year, and n is the number of time steps contained in the period selected):

$$SSR = \frac{\sum_{i=1}^{n} P_{self\ consumed}}{\sum_{i=1}^{n} P_{consumed}}$$

- The self-consumption rate (SCR) is the ratio between the energy self-consumed in a period and the energy produced by the PV in that period, evaluated before the inverter:

$$SCR = \frac{\sum_{i=1}^{n} P_{self\ consumed}}{\sum_{1}^{n} P_{produced(DC)}}$$

These two indicators, that can be evaluated also in the case without battery considering null its contribution to self-consumption, give an immediate idea concerning the system analysed:

Higher is the SSR, higher is the grade of self-sufficiency/autonomy; if it reaches 1 (its maximum value) the owner becomes completely autonomous (it can avoid to buy electricity from the grid). At the same time, values of SSR really high can indicate also that probably the PV and/or the battery is oversized with respect to the consumption: in common cases is not advisable to reach value too high of SSR, because the higher cost of the investment is not justified by its advantages, and there will be always an optimal size of the system that varies from case to case.

Higher is the SCR, higher is the share of production self-consumed and higher should be the profitability of the system. In the other hand, this is not necessary true: it can mean also that the PV production is too small compared to the consumption and/or that the battery is oversized.

Done these considerations, it is always better to have values of SSR and SCR as high as possible at parity of cost and, to choose the system with the optimal size for each specific case, another opportunity to maximise the system's profitability is the demand side management (DSM).

It consists in shifting the demand (usually it is possible to do it only partially) considering two objectives (not necessary coincident):

The first is to make the profiles of consumption and production, that both vary during time, more similar in order to raise the SSR.

For instance, the owner could choose to activate his washing machine, manually or automatically, when the production is higher than the consumption.

The second, but not less important, is to make the interaction with the grid more profitable, trying to consume less (and if it is possible to sell instead) when the price of electricity on the local market is high and trying to buy energy when it is low.

A system equipped with artificial intelligence and sensors that can shift properly loads and storage devices, considering continuously information such as data of the system, data of the energy market and forecasts of the behaviour of both, is a solution that could justify its costs, especially in the future.

4.2 Plots

The creation of plots was helpful, starting from the first phases, to better understand the database's content: the only information available about each measure was contained in a short label and, even if the labels were quite clear (ex: "I_dc_bat_1_A" for the direct current exiting or entering the battery in Ampère), it was important to create some plots simply to verify to have well understood. They were also useful initially, together with computation, to check the coherency between interrelated measures.

In this section plots are used to show the behaviour of the systems object of studies and, when it is not specified, they refer to the household of reference (table t4054; while in Appendix C it is possible to see the plots for table t2098).

The first one lets to visualize the balance of power discussed in Section 4.1.5:

The four different powers (consumed by the house, produced by the PV, exchanged with the grid, exchanged with the battery) must give a null sum for each time step along the full period of study, else it is symptom of errors in the measurements.

Graphically, during a period of choice, the areas under these four curves, that represent the quantities of energy exchanged, must give a sum null if summed with the right convention of sign (ex: area positive when the power enters in the solar home battery system and negative when it leaves it due to consumption or losses or injection in the local grid).

Here are displayed the power flows during a winter day:

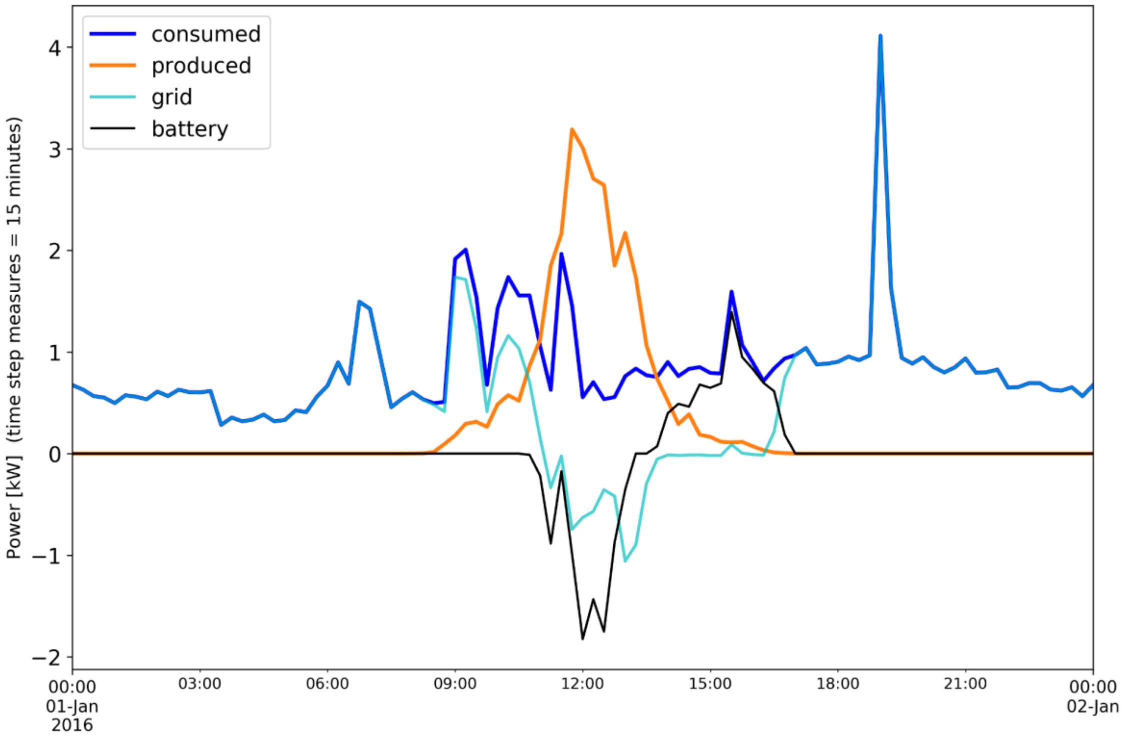


Figure 6. Power flows during the 1th January 2016.

It is possible to see that, during the night, contributions of PV and battery are null and therefore the house consumption is satisfied by the grid only.

Starting from 8:45 the PV starts to produce electricity, that is directly consumed, and consequently the grid will continue to supply power to satisfy the remaining share consumption. As the production increase, it exceeds the consumption (after 10:30) and the surplus is in part stored in the battery and in part sent to the grid. About three hours later the situation is reversed and the battery compensates the deficit of production, while the grid is almost inactive.

After 16:30 the production is null, the battery is not able anymore to satisfy the consumption alone and in one hour it is fully discharged so the grid comes back to play the main role for the rest of the day.

During a summer day the dynamics are the same:

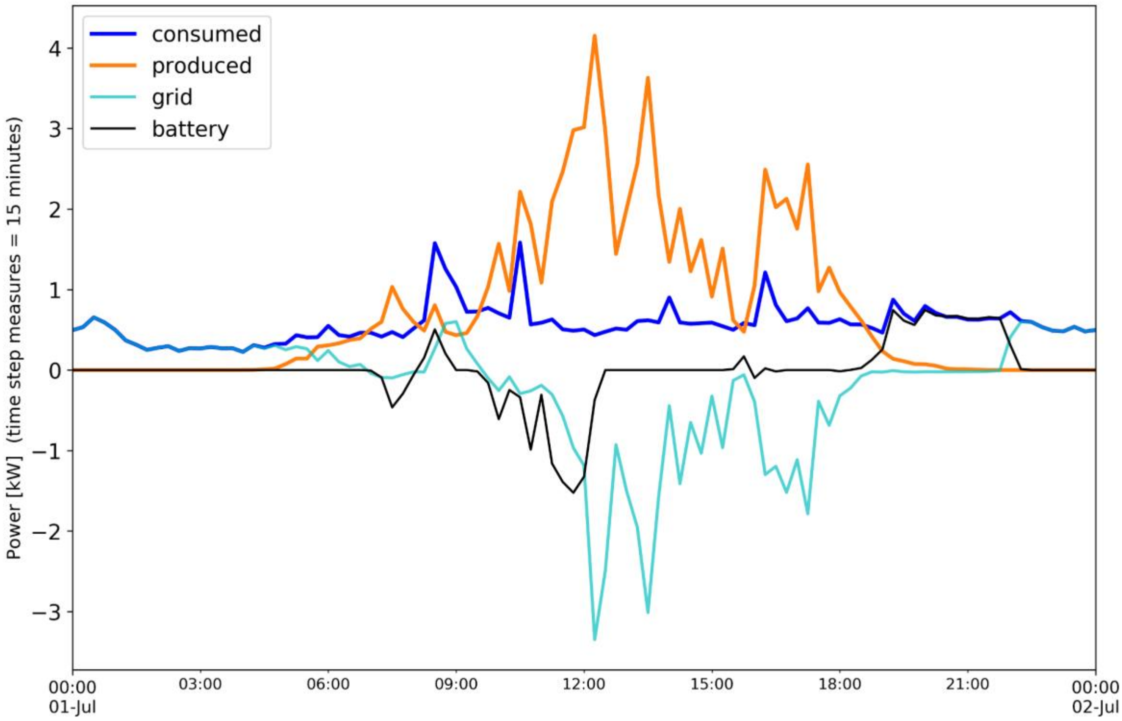


Figure 7. Power flows during the 1th July 2016.

Nevertheless, there are differences: the PV production is much higher and available during more hours, the battery does more switches between charge and discharge and there is not the peak of consumption that was registered in winter during evening, since it was probably due to heating devices.

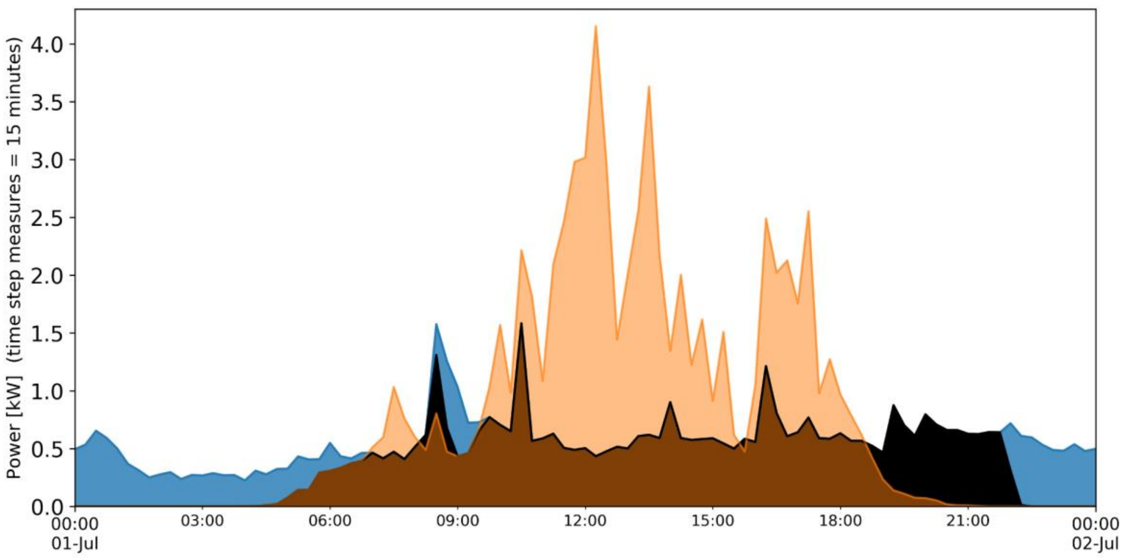


Figure 8. Self-consumption during the 1th July 2016.

Last figure lets to appreciate graphically the concept of self-consumption:

Without battery, the system at each instant would simply try, as priority, to fulfil its consumption with the production available in that moment and the self-consumption is the area given by the intersection of production's area (in orange) and consumption's area (in blue), coloured in brown.

When a battery is added, the self-consumption raises because the battery charges when the consumption is lower than the production, exploiting the surplus, and it can release the energy when the situation is reversed.

Looking at figure 7 we can see that it happens at least two time during that day (areas in black represent the amount of energy provided by the battery) and therefore the self-consumption in this case is equal to the brown area plus the black areas.

Next two figures show, on daily basis, the amount of energy self-consumed along the year compared with the total consumption (figure 8) and with the PV production (figure 9):

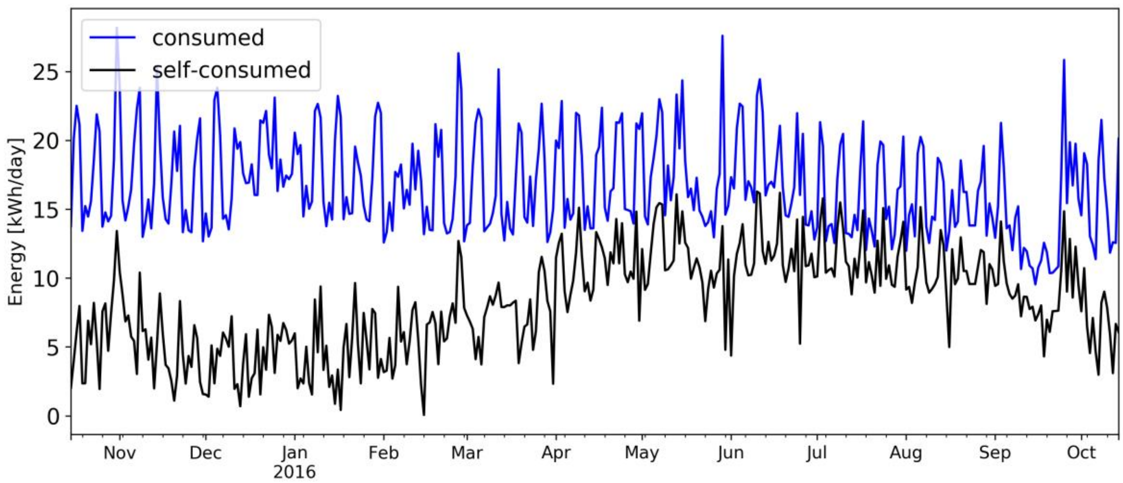


Figure 9. Daily energy consumed and self-consumed along the year.

We can see that the consumption tends to oscillate between 15 and 20 kWh/day, with some peaks of about 25 and a negative peak of 10 kWh/day during the middle of September.

The self-consumption oscillates in a similar way but it is much smaller in winter and closer to the consumption in summer, so we will expect to have the highest values of SSR (that is given by the ratio of the two curves) during the summer months.

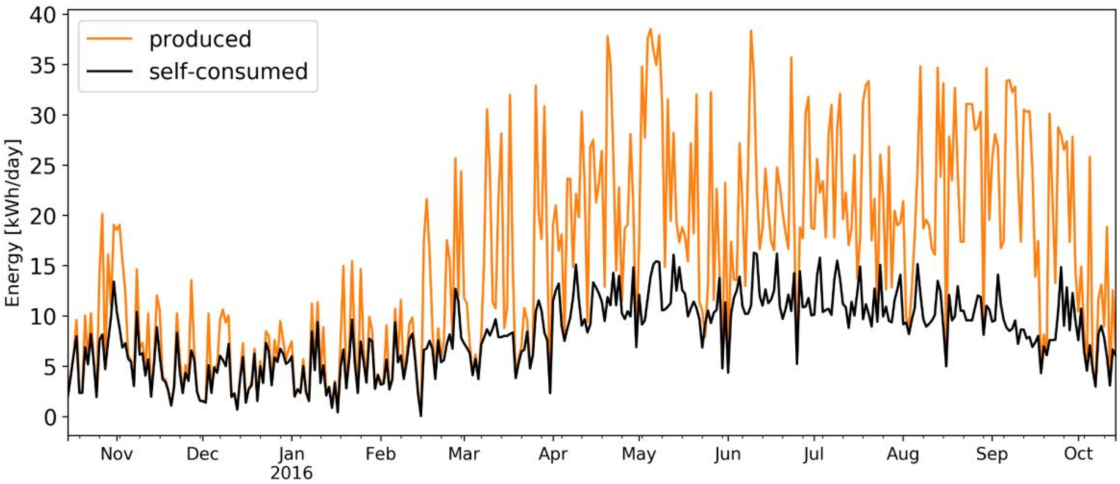


Figure 10. Daily energy produced and self-consumed along the year.

Here we can remark that the profile of production is the most variable:

During summer the production is higher and there are peaks over 35 kWh/day, while during several winter days is under 5 kWh/day and, as we expect since the battery is not of big capacity and it usually fully discharges before the end of each day, it is almost coincident with the self-consumption when it is lower than the daily consumption.

We will expect to have the highest values of SCR (that is given by the ratio of these two curves) during winter months when the surplus of production is lower (mostly due to the production curve since the consumption curve has a more constant trend along the year).

The next plot, that shows the values of SSR and SCR calculated during each week, confirms the previous considerations about these two indicators.

It is possible to deduce that yearly SCR must be higher because the area comprised between the two curves is bigger when SCR is higher (if compared with the area between the two curves when instead SSR is higher).

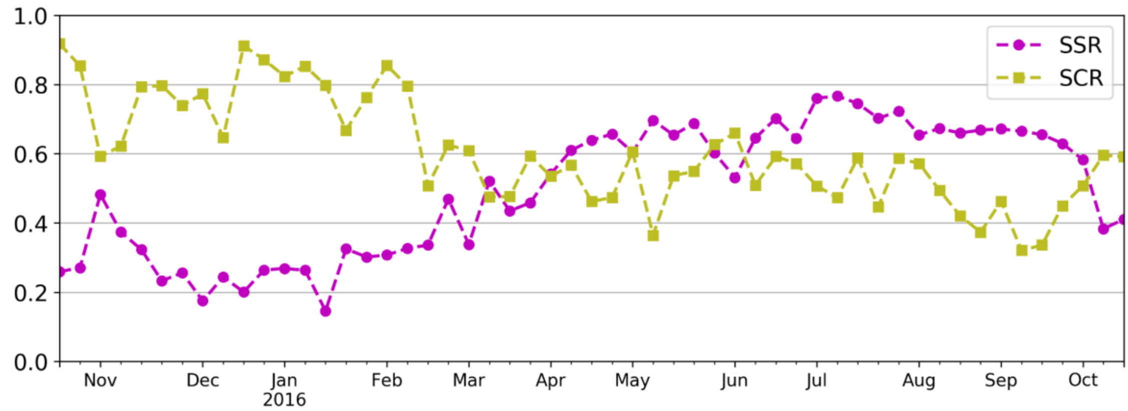


Figure 11. Weekly average of SSR and SCR along the year.

Furthermore, it is possible to note the difference between values collected in middle October 2015 and the ones collected exactly one year later.

It is also interesting to see how the SSR and SCR change if we consider the same system deprived of its battery; we will indicate them as SSR* and SCR* when they are evaluated in this situation.

This change is here displayed considering the ratio of the two indicators in the case without and with battery:

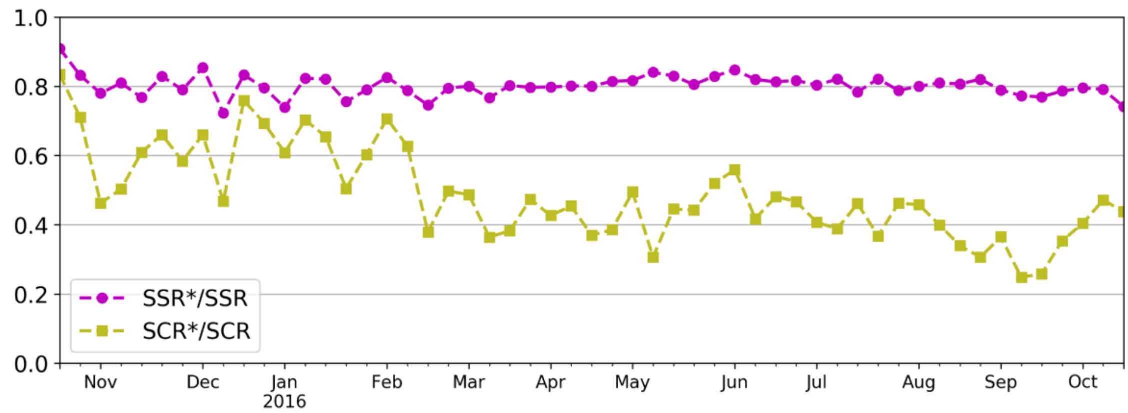


Figure 12. Weekly constant of proportionality to have SSR* and SCR*, respect the normal case.

These ratios can reach 1 as maximum and weekly they are always lower (if evaluated daily there are instead days in which they reach the maximum instead). This means that in the case without battery the weekly performances are always worse (as it could be expected):

The SSR* is about 20% lower than the SSR along the year and the SCR* in average is about 50% lower than the SCR; this last one is small especially in the period in which the surplus of production is higher since the system without battery lacks of the flexibility necessary to exploit better this surplus.

To conclude this section, monthly quantities of energy exchanged are reported in figure 13:

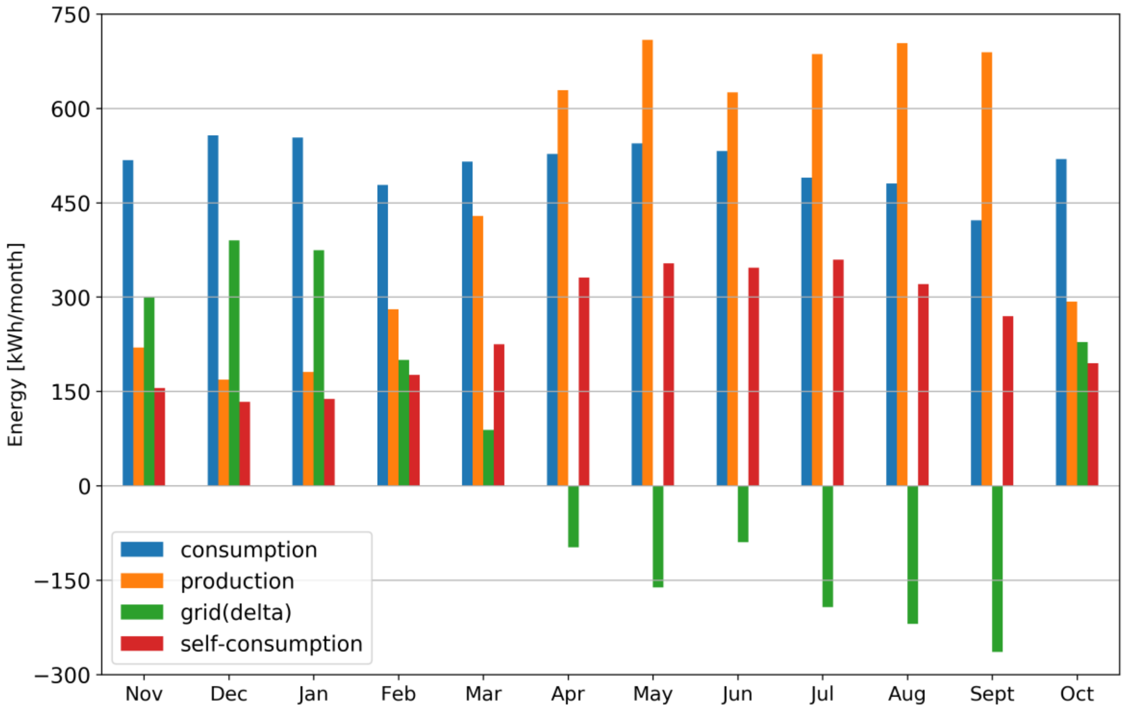


Figure 13. Monthly energy exchanged by the solar home battery system.

Here we see again trends observed before for the consumption (that varies from 420 to 560 kWh/month), the production (from 160 to 710 kWh/month) and the self-consumption (from 130 to 360 kWh/month).

In addition, it is displayed the difference between energy received from the grid and energy sent to the grid (from -270 to 370 kWh/month):

We see that during the six months more productive (from April until September included) the difference is negative, so the share of energy sold is higher, while during other months it is higher the share of energy bought from the grid.

5 Results of the analysis

Following the methodologies described in the previous chapter, the physical quantities of interest are evaluated for each table selected and results are collected in the following tables. They all refer to the whole period of 1 year if it is not specified the contrary.

5.1 PVs

Tables	Energy Produced [MWh]	Relative PV Size [kWh/kWh]	Peak Power installed [kW]	Production for kWp [kWh/kWp]	Inverter Efficiency [%]
t1584	8.550	1.0	11.6	765.6	96.3
t2009	11.110	1.6	12.1	941.8	97.4
t2098	5.985	0.9	6.7	916.1	97.6
t2207	10.503	1.6	12.5	865.9	97.1
t4054	5.613	0.9	6.8	855.6	97.1

Table 2. Results about each photovoltaic system.

From this table we can see that there are two PVs (of t2009 and t2207) characterized by much higher production (almost double than the minimum) and relative PV size with respect to the others.

5.2 Yearly energy exchanged

Tables	Energy Consumed [MWh]	Self-Cons. Energy [MWh]	Energy from Grid [MWh]	Energy to Grid [MWh]	Energy to Battery (DC) [MWh]	Battery Efficiency [%]
t1584	9.165	4.592	4.573	3.958	2.090	76.1
t2009	7.187	5.670	1.517	5.441	2.895	94.4
t2098	6.990	3.747	3.244	2.239	1.864	80.8
t2207	6.613	4.495	2.118	6.007	2.045	97.5
t4054	6.135	3.003	3.133	2.611	0.672	94.9

Table 3. Other relevant energy flows, in addition to PV production, and efficiency of the batteries.

From Table 3 we can observe that consumption of household t1584 is about 50% higher than t4054, while the consumption of the two equipped with the biggest PVs is average; these last households exchange with the grid mainly to sell electricity.

Differences about the energy received by the battery are coherent considering the differences of production and of battery capacity (value that will be displayed in Table 5).

The distribution of each group of 5 results (one for each household) can be better appreciated if displayed through a box plot[3], in which each box shows their statistical distribution:

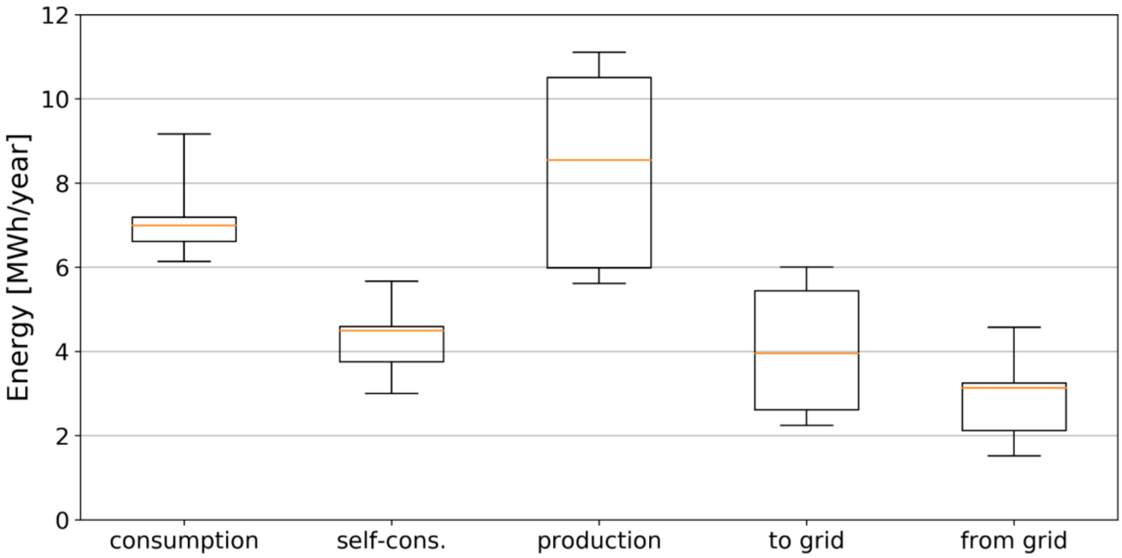


Figure 14. Box plot about energy exchanged (case with battery).

5.2.1 Case without battery

Removing the battery from each system (that it corresponds to consider null its capacity) we obtain different results:

The consumption and the production are not affected but the self-consumption will be lower and the flows entering or exiting the grid will be higher.

[3] With the convention chosen, each box is built to show the minimum, the maximum and the quantiles 25%, 50% and 75%.

Results obtained in this condition are marked by an asterisk and they are collected in next table and figure:

Tables	Energy Consumed* [MWh]	Self-Cons. Energy* [MWh]	Energy from Grid* [MWh]	Energy to Grid* [MWh]
t1584	9.165	3.231	5.934	5.319
t2009	7.187	3.166	4.021	7.945
t2098	6.990	2.460	4.531	3.525
t2207	6.613	2.885	3.728	7.618
t4054	6.135	2.412	3.724	3.202

Table 4. Energy flows in the case without battery.

Comparing these results with the previous table, it is possible to realise that:

- In average self-consumption decreases by 34%.
- In average the energy received from the grid increases by 50%.
- In average the energy sent to the grid increases by 36%.

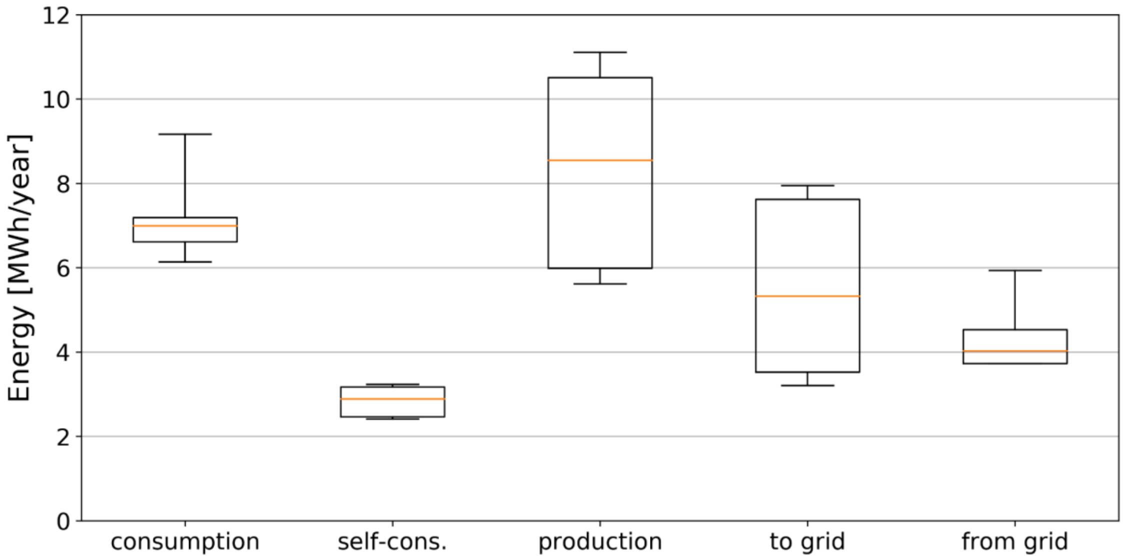


Figure 15. Box plot about energy exchanged (case without battery).

In term of statistical distribution, it is interesting to note that variability of results about self-consumption and energy from the grid decreases (the correspondent boxes are smaller than before), while it happens the opposite for the results relative to the energy sent to the grid.

5.3 SSR and SCR

Tables	Relative Battery Size [kWh/MWh]	Battery Capacity [kWh]	SSR [%]	SCR [%]	SSR* [%]	SCR* [%]
t1584	1.4	12.5	50.1	51.7	35.3	36.4
t2009	1.8	12.7	78.9	49.7	44.0	27.8
t2098	1.5	10.8	53.6	61.1	35.2	40.1
t2207	1.4	9.2	68.0	41.6	43.6	26.7
t4054	0.4	2.2	48.9	51.9	39.3	41.7

Table 5. Battery parameters; Self-Sufficiency Rate and Self-Consumption Rate with and without battery.

These results let to quantify the performance improvement due to the battery:

It is maximum in correspondence of the maximum value of relative battery size and battery capacity (It is the case of t2009 that has a record improvement of SSR and SCR close to +80%), and minimum in the opposite case (about +25% for t4054).

About absolute values of SSR and SCR, they look in line with results of similar analysis [7], according to which they can differ much even between comparable households, even if a little high in some cases (for example for 4054 the SSR*=39.3% is a value high considering that its relative PV size is 0.9). Eventually this can be due to the use of some DSM strategy that, as seen before, consists in increasing the self-consumption shifting the loads in an advantageous way with respect to demand.

It follows the box plot corresponding to Table 5:

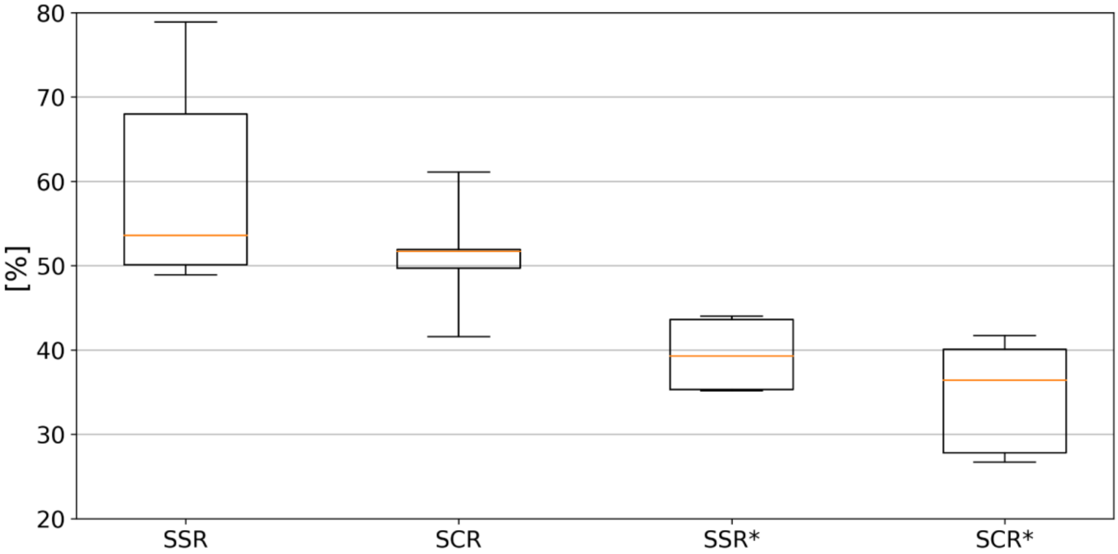


Figure 16. Self-Sufficiency Rate and Self-Consumption Rate, with and without battery.

To study the influence of data's time step on results, the 4 indicators were calculated also using a time step of 1 hour, instead of 15 minutes. Here it is shown the difference between the new results and the ones of Table 5:

Tables	Δ(SSR) [%]	Δ(SCR) [%]	Δ(SSR*) [%]	Δ(SCR*) [%]
t1584	0.7	0.7	1.7	1.8
t2009	0.6	0.4	3.2	2.0
t2098	0.6	0.7	1.5	1.7
t2207	0.7	0.4	2.4	1.4
t4054	0.8	0.8	2.0	2.1

Table 6. Difference between results obtained with a time step=1 hour and with time step=15 min.

It is possible to note that differences are all positive. It means that the time step at lower resolution (of 1 h), in which the fast variations are smoothed out, leads to overestimate all these indicators.

The error is lower than 0.9 % and so almost negligible in the case with the battery, while in the other case it reaches a maximum of 3.2%.

5.4 Economic analysis

To estimate the profitability of a PV battery system grid-connected, its levelized cost of electricity (LCOE) is calculated, according to [7].

This economic indicator can be seen as the minimum average price at which the system should sell the energy produced, during its life, to recover its total cost (that takes into account the investment and costs of operation and maintenance, while the grid is considered as a costless generator producing at its retail price). Its formula is the following:

$$LCOE = \frac{A + E_{fromGrid} * P_{Retail} - E_{toGrid} * P_{toGrid}}{E_{load}}$$

In which $E_{fromGrid}$ and E_{toGrid} are the amounts of energy yearly bought (at price P_{Retail}) and sold to the grid (at price P_{toGrid}) and E_{load} is the yearly consumption of the house.

A is the annuity (constant in this case) that takes into account the investment of the PV (I_{PV}) and of the battery (I_{bat}). Because of their different lifetime (N_{PV} and N_{bat}), we assume the I_{bat} is payed another time after N_{bat} years:

$$A = \left(I_{PV} + I_{bat} * \left[1 + \frac{1}{(1+i)^{N_{bat}}}\right]\right) * (CRF * OM)$$

The other terms are the discount rate (i), the fraction of yearly operation and maintenance, respect the total investment, (OM) and the capital recovery factor (CRF):

$$CRF = \frac{i * (1+i)^{N_{PV}}}{(1+i)^{N_{PV}} - 1}$$

Considering references [7] and [17], the assumptions done to perform the economic analysis are:

- N_{PV} = 20 years;
- N_{bat} = 10 years;
- I_{PV} = 1300 €/kW_p;
- I_{bat} = 300 € + 200 €/kWh of capacity;
- OM = 1.5 %;
- i = 5 %;
- P_{Retail} = 0.293 €/kWh;
- P_{toGrid} = 0.122 €/kWh.

In the same way it can be analysed the case in which the battery is absent:

For installations of PV only, the investment of the battery will be null and this is the only positive change that could make the LCOE lower.

In this case however the yearly self-consumption will be lower (while E_{load} does not vary) and so for each kWh self-consumed in less there will be an equivalent raise of the energy bought by the grid and of the energy sold. This entails a negative change:

Even if the difference of $E_{fromGrid}$ and E_{toGrid} does not change, the difference of their cost will augment because the price to buy electricity is higher than the feed-in tariff for this systems, especially in Germany that has a price of retail above the European average.

The values of yearly consumption and of yearly energy exchanged with the grid are already been calculated (see Table 3 for the case with battery and Table 4 for the other case).

It is important to highlight that the cost chosen for the battery is much lower than the current price, nevertheless it could become a realistic value in the next year due to improvements in this technology.

The results are the following:

	t1584	t2009	t2098	t2207	t4054
LCOE with Battery [€/kWh]	0.297	0.239	0.270	0.267	0.253
LCOE PV only [€/kWh]	0.276	0.238	0.247	0.259	0.251

Table 7. Levelized cost of electricity for each household, with and without battery.

It is possible to see that these investments look all profitable except the first case, since for t1584, in the case with battery, the LCOE is higher than the price of retail considered (0.293 €/kWh).

Each PV only investment has a lower LCOE than the same with battery included. They look to be more profitable then but the difference is almost negligible in two cases (t2009 and t4054).

We can conclude that investments in PVs can already be profitable, in a similar contest, while the PVs coupled with batteries have potential but are not competitive yet.

Incentives here are not considered since they can differ considerably from year to year and from country to country, nevertheless they play an important role because they can lower the LCOE of any investment.

6 Case Optimized

In the previous chapters, the behaviour of five solar home battery systems has been described, but still it is not clear how the controller present in each system operates to manage the different flows of power.

In fact, this can be done trying to reach one or more objectives, such as for instance to maximize the self-consumption, to improve the interaction with the local grid with the aim of increasing its stability or to increase the life time of the battery.

In this chapter each real case analysed before will be optimized changing its controller with one that lets to maximize the self-consumption, to see how much the two cases differ and which is the maximum values of self-consumption rate that is possible to achieve.

According to this dispatch strategy, the battery, if it has not already reached its limit of capacity, will charge every time that there is a surplus of production (with respect to the consumption) and, if it has not already reached the maximum depth of discharge of project, it will discharge every time that there is a deficit of production.

The modelling process is done using a Python toolkit called "*prosumpy*". It has already been used in reference [7] and currently it can be freely downloaded by the platform GitHub [9].

Inputs given to the algorithm of prosumpy are:

The house consumption and the PV generation (evaluated during a year with a time step of 15 minutes).

The efficiency of the inverter and 3 other parameters, already calculated before, important to characterise the battery: its capacity, its efficiency and its maximum power of charge and discharge[4].

[4] With these same inputs normally it is also possible modelling according to an alternative strategy of dispatch (that lets to maximize the self-consumption in a more grid-friendly way, assuming perfect forecast of the demand) but this will not be considered because the corresponding results were affected by an error too high.

The two following figures, realised with prosumpy, show the behaviour of the system optimized during a winter and a summer week:

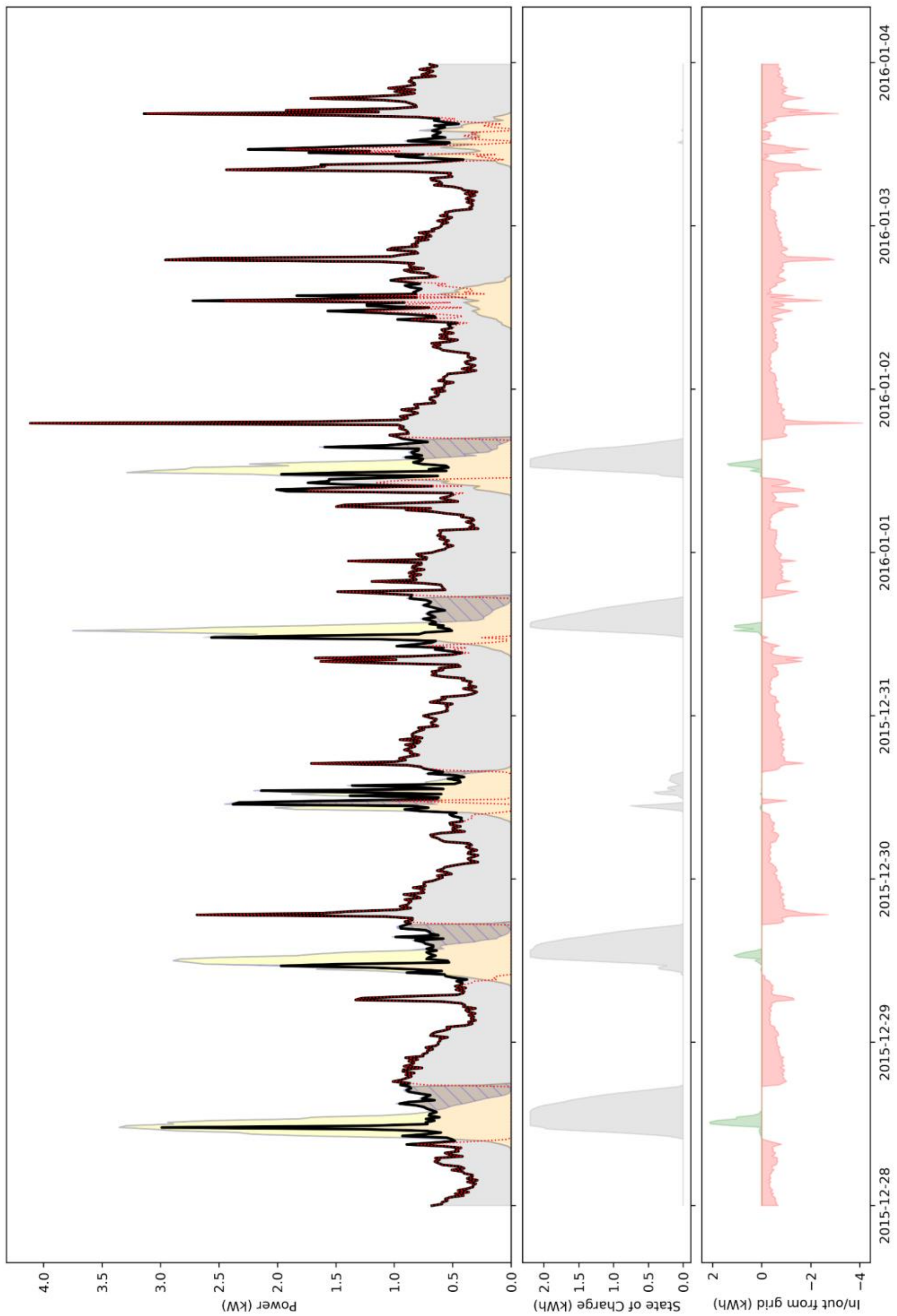


Figure 17. Energy flows for t4054, during a winter week, with strategy of dispatch that maximizes the self-consumption.

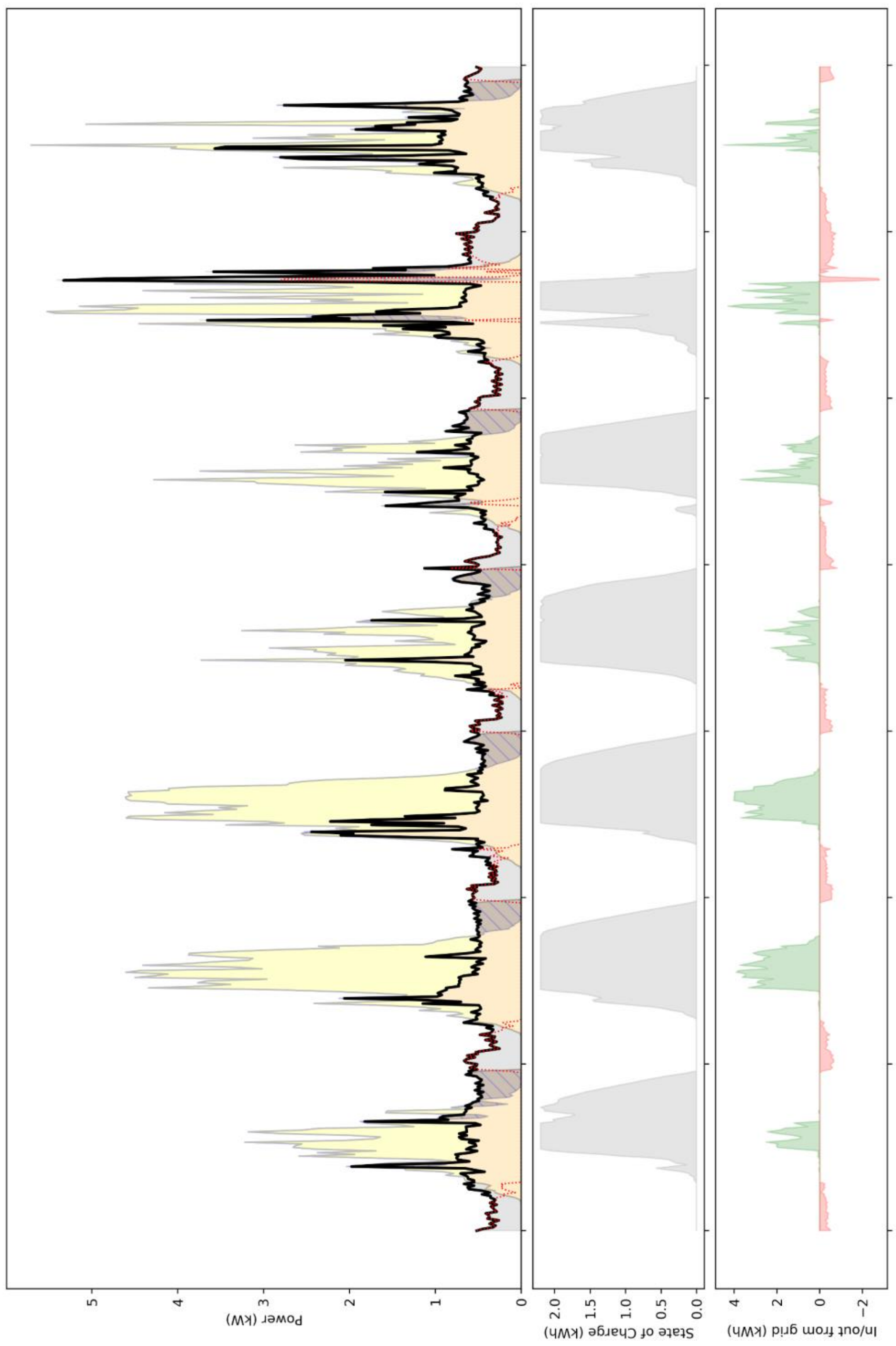


Figure 18. Energy flows for t4054, during a summer week, with strategy of dispatch that maximizes the self-consumption.

In the first subplot, yellow areas correspond to energy produced, while the curve in black is the demand and the area below is the energy consumed (in dark yellow if directly self-consumed and in dark grey if self-consumed thanks to the battery).

In the second subplot it is represented the state of charge of the battery, that can vary from 0 to the maximum capacity available.

In the third, areas in green correspond to the energy sent to the grid, while the areas in red to the energy received.

Comparing graphically with respect to the real case, the difference for this two weeks displayed is not evident but, looking to the corresponding numerical values, it resulted that the battery tends to charge and to discharge at higher power (at its maximum power whenever it is possible). In reality this behaviour could be not recommended because fast variations in power can lower the lifetime of the battery.

New values of self-sufficiency rate and self-consumption rate, evaluated with prosumpy for the optimized case ("SSR opt" and "SCR opt"), are all higher as it was expected:

Tables	Relative Battery Size [kWh/MWh]	Battery Capacity [kWh]	SSR [%]	SCR [%]	SSR opt [%]	SCR opt [%]
t1584	1.4	12.5	50.1	51.7	59.2	61.1
t2009	1.8	12.7	78.9	49.7	84.2	53
t2098	1.5	10.8	53.6	61.1	58.1	66.3
t2207	1.4	9.2	68.0	41.6	75.4	46.1
t4054	0.4	2.2	48.9	51.9	50.7	53.9

Table 8. Battery parameters; Self-Sufficiency Rate and Self-Consumption Rate in the real and in the optimized case.

The difference between indicators of profitability, in the two cases, is in average of about 5%.

Nevertheless, for the data of table t4054 this difference is much lower (around 2%) and this is justified by the small dimension of its battery.

At the contrary, for table t1584 there is an increase record of over 9%.

These results could indicate that, for all the five systems considered, the strategy of dispatch applied differs from the one that lets to maximize the self-consumption (and that would let consequently to maximize the economic gain if evaluated with the same hypotheses of before).

This could be due to the choice of do not fully exploit the performances of the battery in order to preserve its durability and/or avoiding rapid fluctuation in power that is positive also for the grid.

Anyway another explanation of the difference in profitability indicators is that the technical parameters of the battery (especially its capacity) could have been overestimated, since the values were not the ones provided by the constructor, but derived by measures of the database.

LCOE, after the optimization, has been calculated like seen in Section 5.4 and, in accordance with higher values of SSR and SCR, it is always lower with respect to the real case with battery and, for three households, it is lower even if compared with the case without battery.

	t1584	t2009	t2098	t2207	t4054
LCOE with Battery [€/kWh]	0.297	0.239	0.270	0.267	0.253
LCOE PV only [€/kWh]	0.276	0.238	0.247	0.259	0.251
LCOE with Bat. optimized [€/kWh]	0.282	0.230	0.262	0.254	0.250

Table 9. Levelized cost of electricity in the three cases, for each household.

7 Conclusions

The main goal of this thesis was to evaluate residential PV battery systems, analysing the results obtained processing historical data.

The starting database contained information gathered monitoring 19 households in Germany, between the year 2015 and 2016.

Final results seem to be in accordance with mostly of the studies present in literature. In particular, it emerged that:

- It is important the quality of data at disposal to perform analysis of this kind and to have a clear understanding of them.
- The starting high time-resolution of 1 second was excessive: replacing it by a time-resolution of 15 minutes lets to simplify the analysis without affecting the results evaluated on the whole period of 1 year.
- Mostly of the data were not enough complete: only 5 households passed the selection criteria. This especially because the others contained information relative to periods shorter than 1 year.
- It was possible to obtain several results using a low variety of measure, anyway the lack of detailed information about some specific technical parameter could have led to errors.
- SSR was a little higher compared with similar case in literature, therefore it is suspected that some DMS strategy could have been applied in the cases studied.
- SSR and SCR depend strongly by the system design. Nevertheless, there will be always a more or less evident difference since they depend by several parameters and, for instance, each household has a different profile of consumption.
- For the five systems studied the use of the battery lets to reach values considerably higher of SSR and SCR, even if no battery is bigger than 13kWh. One of them had a record increment of the self-sufficiency rate from 44% (case without battery) to about 79%.

- Simulating an ideal dispatch strategy to maximize the SSR, the profitability indicators raised more and the highest value of SSR became 84.2%. Anyway, according to the literature, generally is not possible to reach value really high of SSR without oversizing the PV battery system.
- The price of the battery is still high so it does not look economic viable to equip the small PV systems with this kind of storage. Despite this, even if today its employ can be convenient only if incentivised, in the future this can change as its cost continues to follow a decreasing trend.

8 References

[1] H. S. Adolf Goetzberger, “Photovoltaic materials, history, status and outlook,” *Materials Science and Engineering R 40,* pp. 1-46, 2003.

[2] International Energy Agency, “World Energy Outlook,” 2017.

[3] International Renewable Energy Agency, “Renewable capacity highlights,” 31/03/2018.

[4] Frankfurt School-UNEP Centre/BNEF, “Global Trends in Renewable Energy Investment,” 2018.

[5] World Energy Council, “World Energy Resources - Solar,” 2016.

[6] International Energy Agency, “Global shifts in the energy system,” 14 11 2017. [Online]. Available: https://www.iea.org/weo2017/.

[7] K. M. P. Z. Sylvain Quoilin, “Quantifying self-consumption linked to solar home battery systems: Statistical analysis and economic assessment,” *Applied Energy,* vol. 182, p. 58–67, 2016.

[8] V. D. A. De Boeck L., “Comparison of support policies for residential photovoltaic systems in the major EU markets through investment profitability,” *Renewable Energy,* pp. 42-53 (Volume 87), Mars 2016.

[9] “prosumpy – Energy prosumer analysis with Python,” [Online]. Available: https://github.com/energy-modelling-toolkit/prosumpy.

[10] V. S. H. Joern Hoppmann, “The economic viability of battery storage for residential solar,” *Renewable and Sustainable Energy Reviews,* pp. 1101 - 1118, 2014.

[11] K. T. J. H. Maik Naumann, “Lithium-ion battery cost analysis in PV-household application,” *Energy Procedia,* pp. 37-47, 2015.

[12] N. K. M. J. Cong Nam Truong, “Economics of Residential Photovoltaic Battery,” *Batteries,* 2016.

[13] R. Velik, “The Influence of Battery Storage Size on Photovoltaics: Energy Self-Consumption for Grid-Connected Residential Buildings,” *International Journal of Advanced Renewable Energy Research,* 2013.

[14] Python Softare Foundation, “What is Python? Executive Summary,” [Online]. Available: https://www.python.org/doc/essays/blurb/.

[15] W. McKinney, Python for Data Analysis, O'Reilly, 2012.

[16] European Commission Joint Research Centre, “ Country and regional maps (PVGIS),” 2017. [Online]. Available: http://re.jrc.ec.europa.eu/pvg_download/map_index.html.

[17] Fraunhofer ISE, “Recent Facts about Photovoltaics in Germany,” 2018.

9 Appendices

Appendix A

Country	Policy / Regulatory Target	Supply Side Drivers	Demand Side Drivers	Fiscal Incentives	Remarks
Germany	Yes	Feed-in tariff; Competitive bidding	Mandatory interconnection	Capital subsidy	Grid parity achieved, capital subsidy now provided for energy storage.
China	Yes	Feed-in tariff; Competitive bidding		Capital subsidy	
Japan	Yes	Feed-in tariff	Net metering	Capital subsidy	Shifted from net to gross metering in 2009.
Italy	Yes	Feed-in tariff			
United States	Yes	Investment tax credit (ITC)	Renewable Portfolio Standards (RPS); Net metering	Capital subsidy; Tax credits	A few states have gross metering in place
France	Yes	Feed-in tariff			
Spain	Yes	Feed-in tariff		Capital subsidy	New projects not eligible for FiT from 2012, additional 6% on participating projects.
United Kingdom	Yes	Feed-in tariff	Net metering; Renewable Obligation (RO)	Capital subsidy	
Australia	Yes	Feed-in tariff	Net metering	Capital subsidy	
India	Yes	Feed-in tariff; Competitive bidding	Renewable Portfolio Obligation (RPO); Renewable Energy Credits (REC)	Capital subsidy; Viability gap funding; Accelerated depreciation; Tax holidays; Priority Sector Lending; Concessional Duties	Competitive bidding on tariff is preferred instrument rather than FiT.

Fig. App. 1. Key drivers for growth in major solar PV markets [5]

Appendix B

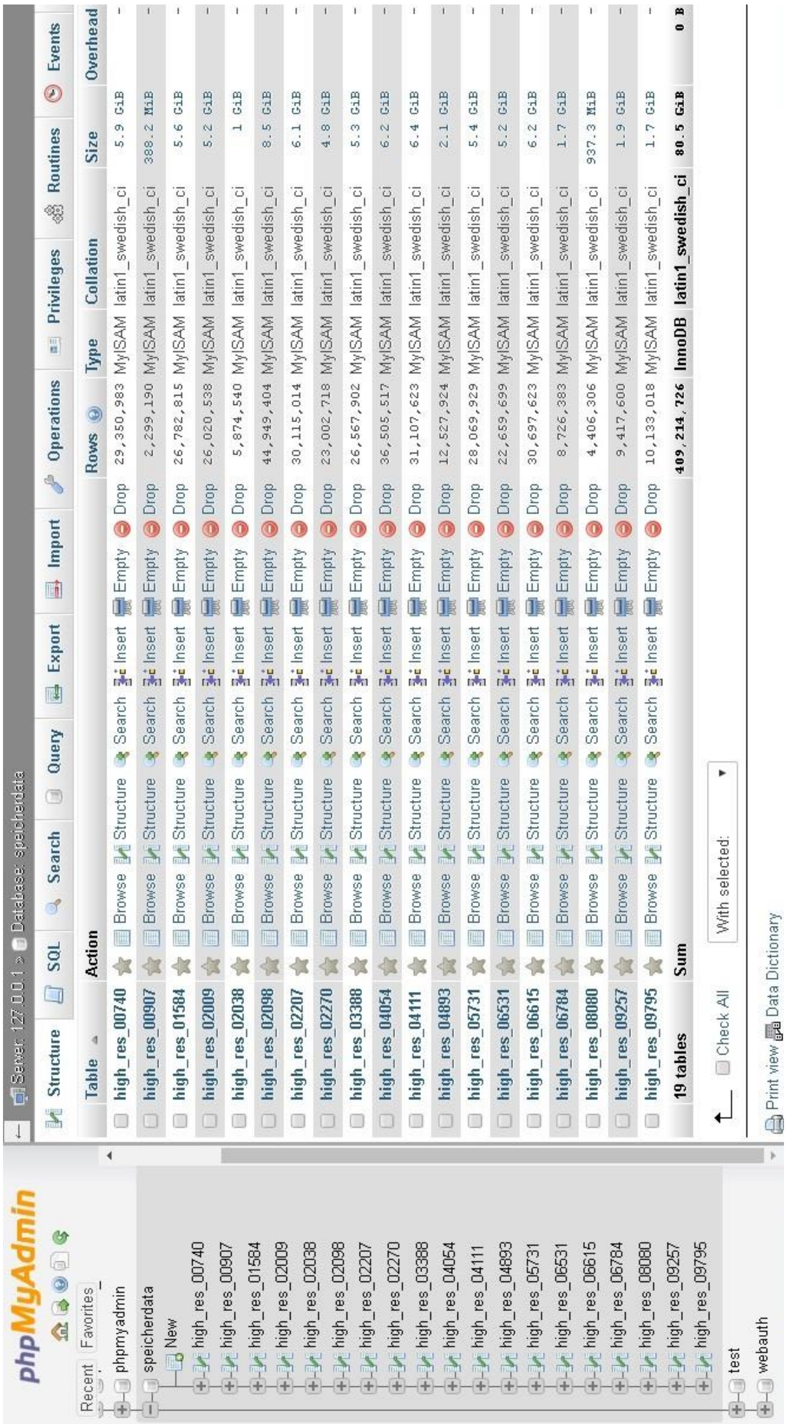


Table	Action	Rows	Type	Collation	Size	Overhead
high_res_00740	Browse Structure Search Insert Empty Drop	29,350,983	MyISAM	latin1_swedish_ci	5.9 GiB	-
high_res_00907	Browse Structure Search Insert Empty Drop	2,299,190	MyISAM	latin1_swedish_ci	388.2 MiB	-
high_res_01584	Browse Structure Search Insert Empty Drop	26,782,815	MyISAM	latin1_swedish_ci	5.6 GiB	-
high_res_02009	Browse Structure Search Insert Empty Drop	26,020,538	MyISAM	latin1_swedish_ci	5.2 GiB	-
high_res_02038	Browse Structure Search Insert Empty Drop	5,874,540	MyISAM	latin1_swedish_ci	1 GiB	-
high_res_02098	Browse Structure Search Insert Empty Drop	44,949,404	MyISAM	latin1_swedish_ci	8.5 GiB	-
high_res_02207	Browse Structure Search Insert Empty Drop	30,115,014	MyISAM	latin1_swedish_ci	6.1 GiB	-
high_res_02270	Browse Structure Search Insert Empty Drop	23,002,718	MyISAM	latin1_swedish_ci	4.8 GiB	-
high_res_03388	Browse Structure Search Insert Empty Drop	26,567,902	MyISAM	latin1_swedish_ci	5.3 GiB	-
high_res_04054	Browse Structure Search Insert Empty Drop	36,505,517	MyISAM	latin1_swedish_ci	6.2 GiB	-
high_res_04111	Browse Structure Search Insert Empty Drop	31,107,623	MyISAM	latin1_swedish_ci	6.4 GiB	-
high_res_04893	Browse Structure Search Insert Empty Drop	12,527,924	MyISAM	latin1_swedish_ci	2.1 GiB	-
high_res_05731	Browse Structure Search Insert Empty Drop	28,069,929	MyISAM	latin1_swedish_ci	5.4 GiB	-
high_res_06531	Browse Structure Search Insert Empty Drop	22,659,699	MyISAM	latin1_swedish_ci	5.2 GiB	-
high_res_06615	Browse Structure Search Insert Empty Drop	30,697,623	MyISAM	latin1_swedish_ci	6.2 GiB	-
high_res_06784	Browse Structure Search Insert Empty Drop	8,726,383	MyISAM	latin1_swedish_ci	1.7 GiB	-
high_res_08080	Browse Structure Search Insert Empty Drop	4,406,306	MyISAM	latin1_swedish_ci	937.3 MiB	-
high_res_09257	Browse Structure Search Insert Empty Drop	9,417,600	MyISAM	latin1_swedish_ci	1.9 GiB	-
high_res_09795	Browse Structure Search Insert Empty Drop	10,133,018	MyISAM	latin1_swedish_ci	1.7 GiB	-
19 tables	Sum	409,214,726	InnoDB	latin1_swedish_ci	80.5 GiB	0 B

Fig. App. 2. Database Speicherdata in MySQL

Appendix C

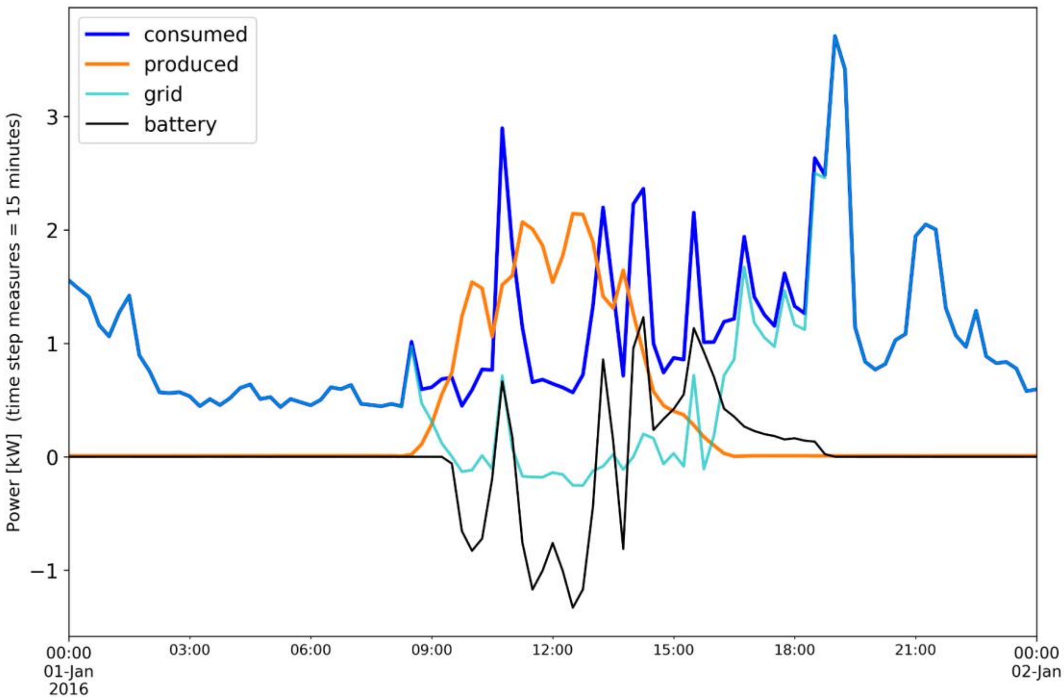


Fig. App. 3. Power flows during the 1th January 2016 for table t2098.

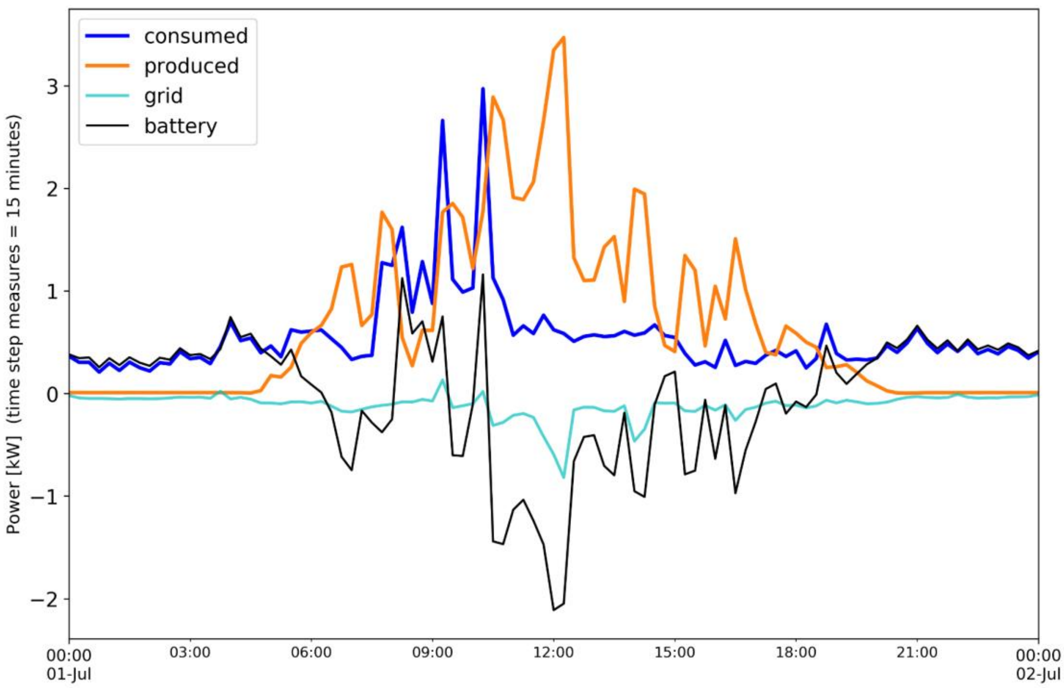


Fig. App. 4. Power flows during the 1th July 2016 for table t2098.

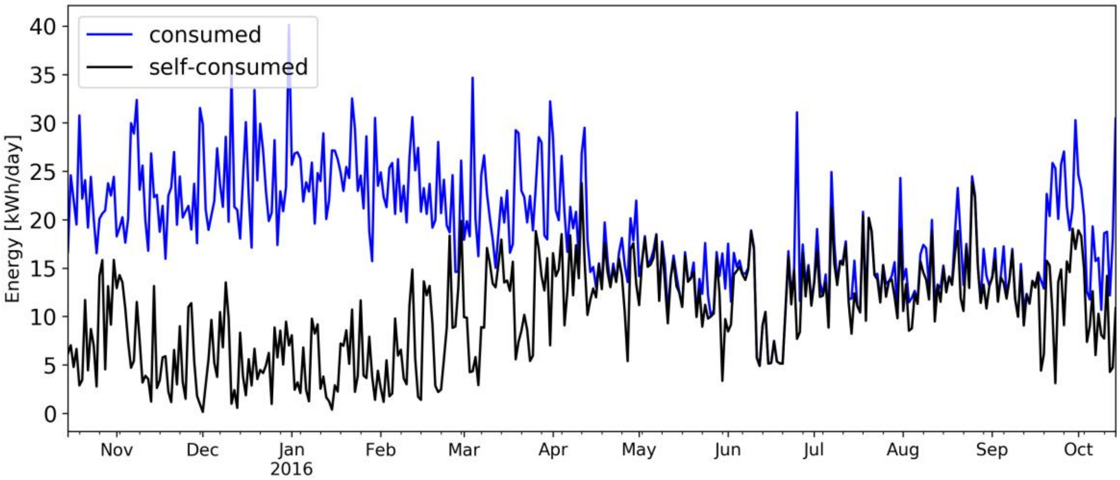


Fig. App. 5. Daily energy consumed and self-consumed along the year for table t2098.

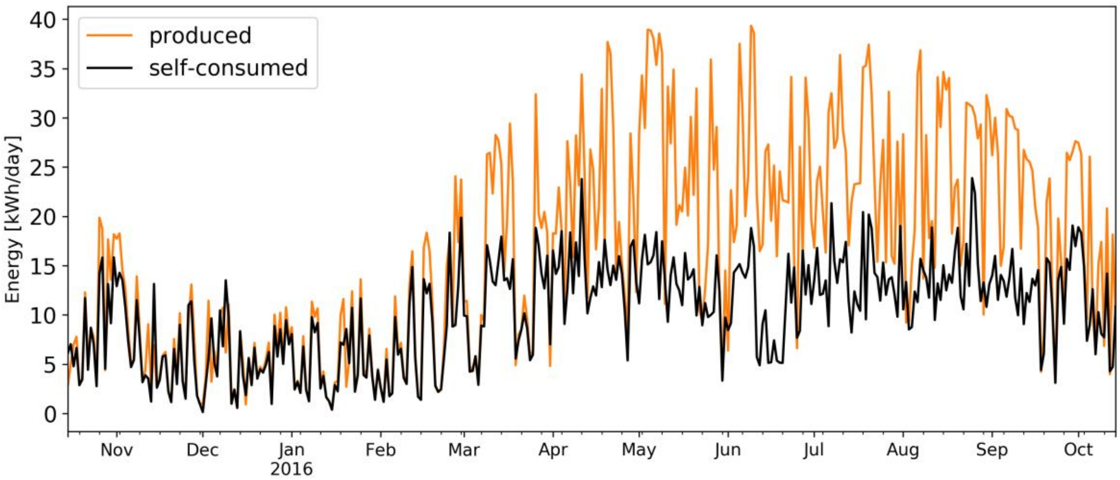


Fig. App. 6. Daily energy produced and self-consumed along the year for table t2098.

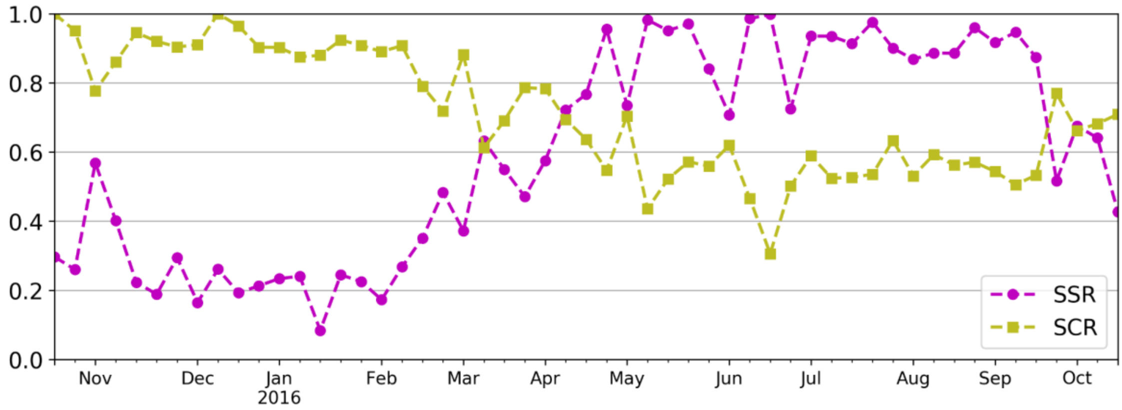


Fig. App. 7. Weekly average of SSR and SCR along the year for table t2098.

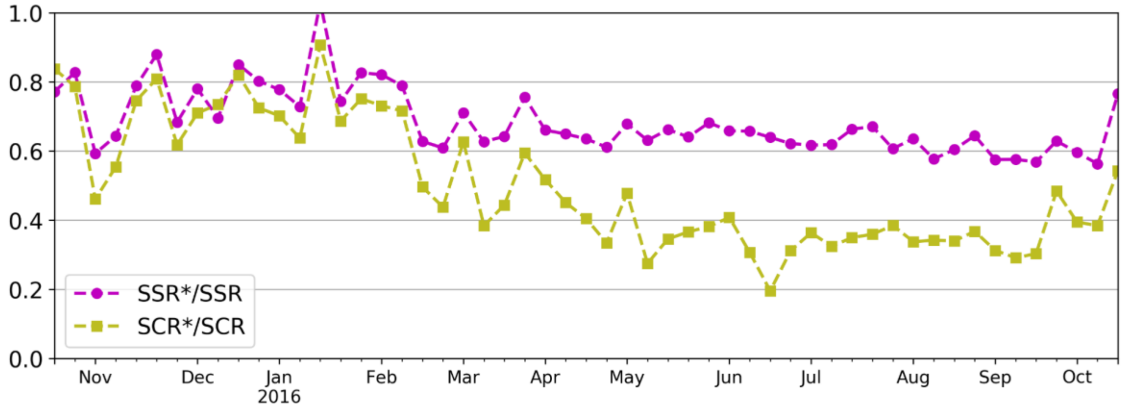


Fig. App.8. Weekly constant of prop. to have SSR* and SCR*, respect the normal case, for t2098.

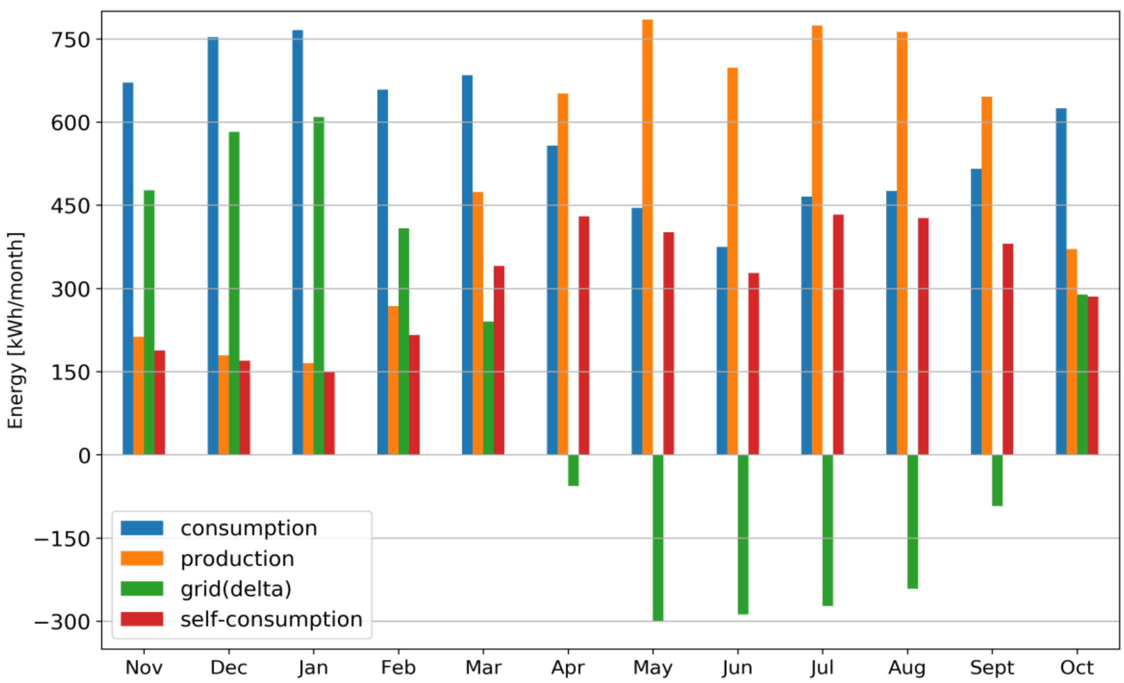


Fig. App. 9. Monthly energy exchanged by the solar home battery system of table t2098.