22nd International Conference on Renewable Energies and Power Quality (ICREPQ'24)
Bilbao (Spain), 26th to 28th June 2024
Renewable Energy and Power Quality Journal (RE&PQJ)
ISSN 2172-038 X, Volume No.22, September 2024

Modelling of solar power production at 65° 35′ N

Per Westerlund[1], Jesper Rydén[2] and Nils Persson[3]

[1] Electric Power Engineering
Luleå University of Technology
SE-931 77 Skellefteå, Sweden
Phone number: +46 910 58 57 38

[2] Department of Energy and Technology
Swedish University of Agricultural Sciences
SE-750 07 Uppsala, Sweden
Phone number: +46 18 67 35 57

[3] Department of Engineering Sciences and Mathematics
Luleå University of Technology
SE-971 87 Luleå, Sweden
Phone number: +46 920 49 10

Abstract. Distributed generation feeds into the distribution part of the electric grid. As it is constructed with power flows going from higher to lower voltages, there might be problems with overloading and high currents, caused by more production than consumption in the area fed by a substation.

To estimate the impact of generation connected to the distribution system, there is need for a statistical description of the production. Five years of data from a solar power plant in Luleå have been analysed statistically. Also, a physical model has been applied.

The results show that the physical model manages to predict reasonable the data. Of the five years, one has apparently a lower production, but it is due to the lack of data. There is a marked seasonality, where the power drops considerable in September and vanishes during winter. The variation at a certain day and a certain hour is large, as the interquartile distance is about one half of the maximum power in summer.

Key words. Solar power generation, power generation planning, time series analysis, extremes.

1. Introduction

The electric distribution grid has been constructed with power flows going from higher voltages to lower ones. Distributed generation is the generation connected to the distribution grid. Normally the generation is connected to the subtransmission grid or the transmission grid in case of nuclear power plants.

The distribution grid is not designed for receiving power. Small sources, i.e. of the same size as the load, do not cause problems. This is the case of solar panels on roofs. Then the power is consumed in the area fed by the same transformer.

Larger installations can cause problems since some power is feed to higher voltage levels, passing through the transformers in the secondary substation. The issues are overloading of transformers and cables and voltage rise.

The capacity of an electric grid for new production or consumption is called "hosting capacity", a denomination that was coined 20 years ago [1].

Even if the power grid can handle the new currents other problems, like increased losses in transformers and transmission lines, need to be taken into consideration.

The tariffs of the distribution company are determined by the power losses and the power flows to the company owning the subtransmission grid [2].

The purpose of this paper is to statistically compare and describe production data from a solar power plant farm in northern Sweden so that statistically sound data can be used for both grid simulations and economical models for future establishments.

The resulting model will be used for calculating the power flows after the installation. Then, it will be possible to estimate how the tariffs of the distribution-grid company will change and thus how much the owner of the installation should pay to the distribution network owner.

2. Data

The data set originates from a solar power plant in Luleå, a town in northern Sweden with latitude 65° 35′ north and longitude 22° 11′ east. The panels have an elevation of 30 degrees and are facing south.

There are five years of data, 2019-2023, with an hourly resolution. However, occasionally there have been interruptions in the storage of production data, most notably during 2020 when data is missing from 20th May to 25th June. Fig. 1 shows the yearly variation as well as the peak for each day for year 2021. The produced power shown has been scaled with respect to the maximum installed power. In the remainder of this paper, we will present power in that manner. Note in the beginning of the year the result of snow covering the panels.

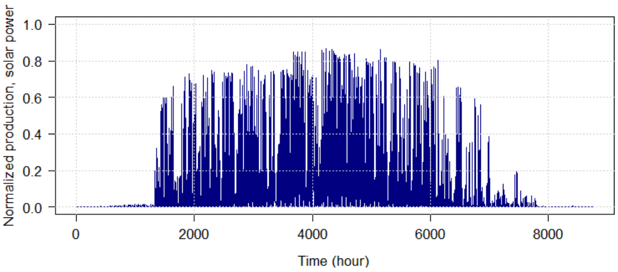

Fig. 1. Normalized production of the solar power plant in 2021 as a function of time.

Fig. 2 presents a comparison between the five years: production as a fraction of installed capacity. Judging from this, there is no apparent trend in the production.

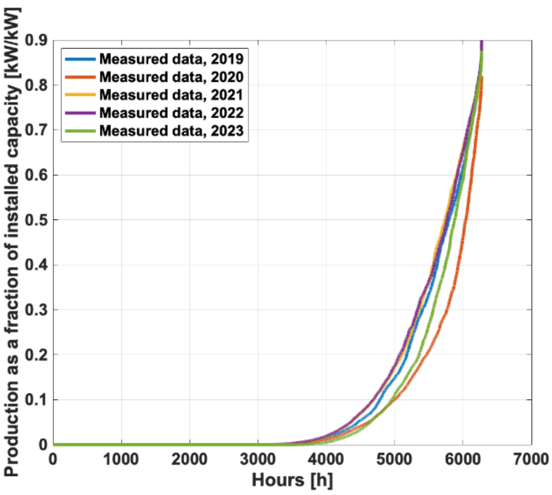

Fig. 2. Comparison between the data for the five years.

3. Method I: Comparison with a physical model

The available power can be calculated using a model of the physical aspects [3]. A prediction of the possible production was calculated by a web site and is shown in Fig. 3 for the case of year 2019 [4]. This figure is a cumulative distribution function in statistics, i.e., it shows how during how many times the production is below a certain level.

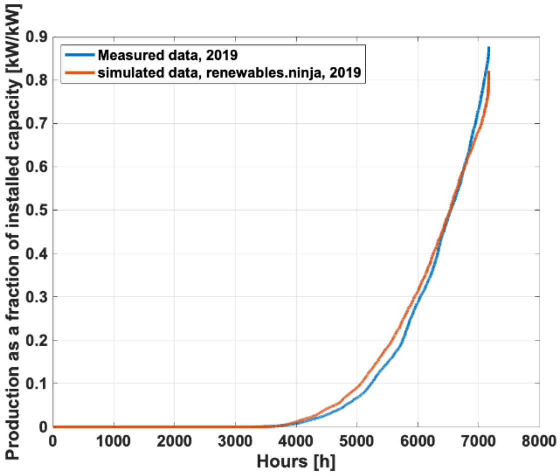

Fig. 3. Comparison between the data for 2019 and the calculation for the same year.

4. Method II: Variability over time

It is of interest to establish and quantify a "normal" year of produced power, for further investigations and e.g. simulation studies. Thus, a major concern is variability. More precisely, variability occurs in the form of seasonality as well as daily (diurnal) variation. There is a random effect caused by clouds passing in front of the sun.

A. Study of seasonal aspects

A simple algorithm could be as follows (recall that the original time series have hourly observations):

(1) Gather all observations day by day (1, …, 365).
(2) To handle diurnal variability, data are aggregated to a "typical" value (e.g. the mean). For each day (1, …, 365), five such values are then obtained (for each of the five years available).
(3) Investigate for each day (1, …, 365) the variability, e.g. through visualisations or computing measures like simple mean.

In step (3), one option might be to use a smoothing technique, e.g. generalized additive models (GAM), cf. [5]. However, one of the series (2020) had unfortunately a large portion of the month of June missing. Hence, in the proceeding, we present a less elaborated approach and simply plot the aggregated data monthly.

In Fig, 4, we present the results for summer months (June –September). In step (2), the sample mean was taken for each day (and each series). For the month of June, one notes the missing data (production 0), originating from 2020.

Now, consider extremes of produced power. In step (2), the daily maximum is then computed. The result is shown in Fig. 5. Again, for the month of June, the missing period is evident through zeros.

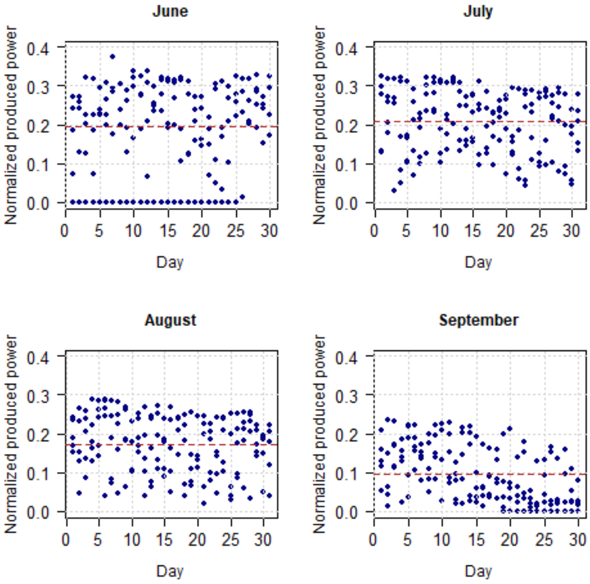

Fig. 4. Aggregated data for summer months, based on daily means. Dashed line: overall mean of these means.

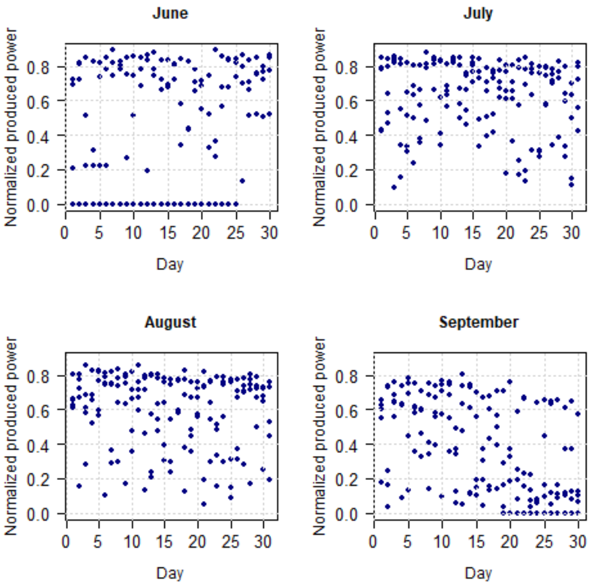

Fig. 5. Aggregated data for summer months, based on daily maxima.

B. Study of diurnal patterns

To get an idea of the diurnal variability, data were aggregated into roughly two-week periods (half months). For such an aggregated data set, for a particular hour, 0, …, 23, there are hence about 15 observations. The results are illustrated in Fig. 6 for the months of July and September, all years considered. As expected, maximum produced power occurs around noon. Moreover, in comparison to July, the average levels in September are considerably lower and the tails over the day are less fat as the season grows darker. Finally, around noon, the inter-quartile ranges (widths of boxes in boxplots) seem to be of the same order for July and early September.

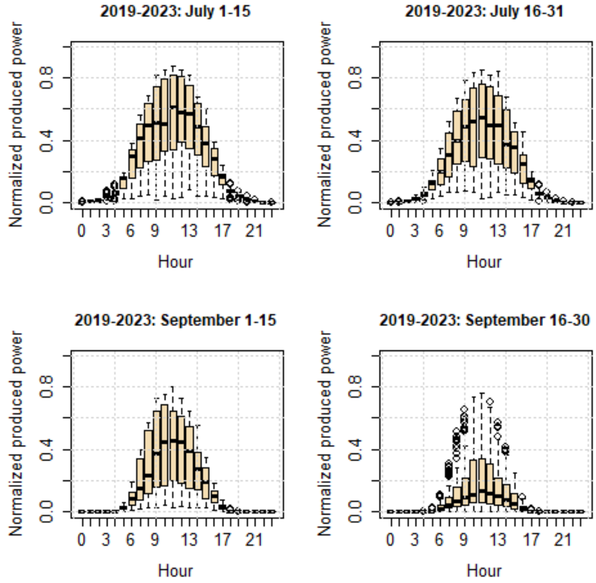

Fig. 6. Diurnal patterns of produced power, all years considered.

5. Method III: Extreme-value analysis

The right-hand tail is interesting, as there are the values that will affect the grid the most. For instance, the 95-quantile is a straight-forward measure of how far the tail goes. It is interesting to see when this large production occurs.

A. Descriptive statistics: quantiles and extremes

In Table 1, some descriptive statistics is presented of produced effect, normalized with respect to maximum installed effect. For the annual maximum production, the level is quite stable over the years. The influence of the missing data in 2020 is seen more clearly for the 95% quantile, which is considerably lower in 2020 compared to other years.

Table 1. Descriptive statistics

	2019	2020	2021	2022	2023
95% quantile	0.579	0.348	0.629	0.631	0.578
Maximum production	0.877	0.820	0.866	0.900	0.878

In addition to Table 1 are illustrated in Fig. 7 the empirical upper quantiles from 0.95 to 0.999 for the five years under study. Again, note the different features resulting from year 2020.

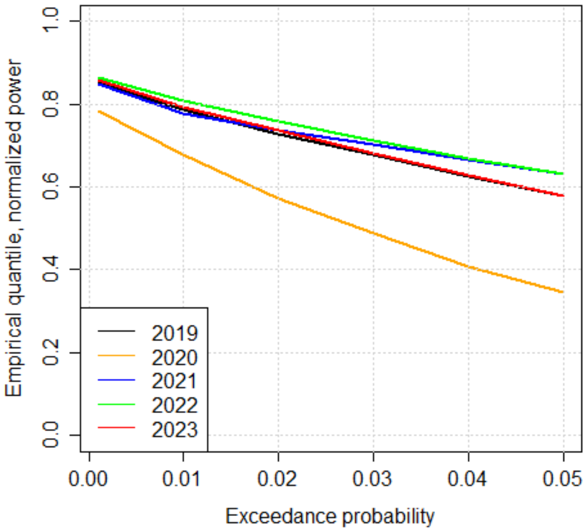

Fig. 7. Upper quantiles of normalized produced power, based on data 2019-2023.

B. Extreme-value analysis

There are several ways to analyse statistically the extremes [6-8]. The classical approach is to investigate block maxima, where each block is one year, and to fit a Generalized Extreme Value (GEV) distribution. We have merely five years of data, but alternative statistical frameworks exist for a decent analysis; cf. [5], Chapter 1, for a review, or an application of the so-called ACER technique to coastal engineering in [9]. For small samples, the estimation technique of so-called L-moments, presented in [10], is often considered more robust than e.g. estimation by maximum likelihood. Algorithms are implemented in an R package, lmomco [11].

Annual maxima were extracted from the series and a GEV distribution was fitted by L.moments. Of particular concern in the GEV distribution is the shape parameter, its estimate is here found to be -1.1. A negative shape parameter implies from theory an upper bound (the case of reversed Weibull), which suits our situation, since there is an upper limit of installed power.

6. Results and discussion

Recall from Fig. 2 that there is no apparent trend in the production and that year 2020 appears to have a lower production, due to the lack of data.

There is a clear drop in September. Table 7.9 in [3] displays a minimum insolation of more than a half of the maximum for a latitude of 40 degrees. Comparing Fig. 4 and Fig. 5 the extremes do not fall so much then as the means do. The diurnal variation in Fig. 6 had the expected pattern. However, the variation for a specific hour is quite large with an interquartile distance of more than a half in July and in the beginning of September. In the last part of that month. there are some extremes in the box plots that correspond to days when there is a clear sky.

The variation between the days and the years is not that important when it comes to estimate the effect on the tariffs that distribution network owner pays to the subtransmission network owner, as the electric circuit model is linear.

Table 1 and Fig. 7 indicate again that 2020 was a particular bad year, which is due to the lack of data from about a month. The estimation of a generalised extreme-value distribution gives the expected result, a reversed Weibull.

An aspect of solar power production in northern Scandinavia is that panels might get covered by snow in winter season. The effect of this was seen in Fig. 1, the low levels at the beginning last longer that those at the end.

Further work includes a more detailed physical model, which takes into account the position of the sun in the sky and meteorological data. In addition, fitting different distribution to the data, can give more insight into the variation.

7. Conclusion

This physical modelling of the solar power plant in Luleå gives reasonable results when compared to the actual production. One year out of five has clearly lower production than the other ones due to lack of data. There is a marked seasonality and a large variation. The inter-quartile distance is about half of the maximal production.

References

[1] M. Bollen, T. Tavares de Oliveira, N. Etherden, S. Bhattacharyya, and S. Bahramirad, 20 years of hosting capacity studies, 2004 - 2024, CIRED Workshop, Vienna, 2024.

[2] Swedish Energy Market Inspectorate, Ellagens bestämmelser om ersättning vid inmatning av el: Metod vid prövning av tvister om nätnyttoersättning, Ei PM2020:04, http://www.ei.se, (2020).

[3] G.M. Masters, Renewable and efficient electric power systems, John Wiley & Sons, Hoboken, NJ (2004), ch. 7.

[4] Renewables.ninja, https://www.renewables.ninja, accessed in February 2024.

[5] S.N. Wood, Generalized additive models. An introduction with R, Chapman & Hall/CRC Press, Boca Raton (2006).

[6] D.K. Dey and J. Yan, Extreme value modelling and risk analysis. Methods and applications, Chapman & Hall/CRC Press, Boca Raton (2016).

[7] I. Rychlik and J. Rydén, Probability and risk analysis: an introduction for engineers, Springer-Verlag, Berlin (2006).

[8] P. Westerlund and W. Naim, “Extreme value analysis of power system data”, International Conference on Time Series and Forecasting, Granada (2019), pp. 322-327.

[9] J, Rydén, Estimation of return levels with long return periods for extreme sea levels by the Average Conditional Exceedance Rate method, GeoHazards 2024 5(1), 165-175 (2024)

[10] J.R.M. Hosking and J.R. Wallis, Regional frequency analysis. An approach based on L-moments, Cambridge University Press, (1997).

[11] W.H. Asquith, lmomco -- L-moments, censored L-moments, trimmed L-moments, L-comoments, and many distributions. R package version 2.4.7, Texas Tech University, Lubbock, Texas, (2024).