# Investigating Correlation of PV Generation Time Series

J. Taylor, A. Everard, A.R Buckley, L.M.H Hall, J.R.C Briggs

## Directive

This briefing paper explores the correlation/synchronicity between the generation time-series of solar photovoltaic (PV) systems on a half-hourly time period as a function of the distance between the systems. The resulting dependence will have ramifications concerning the level of PV penetration permissible on the electrical grid, since a high degree of correlation increases the risk of grid faults. This work aims to identify the critical length scales whereby the correlation of PV generation can be considered problematic. Key stakeholders in this study are those tasked with managing and administering the electrical grid in the UK, that is, National Grid and the Distribution Network Operators (DNOs), though the results are also extremely pertinent to anyone attempting energy flow simulations involving distributed PV generation.

## MgDB Ensemble

The Microgen Database (MgDB) [1] dataset used in this analysis comprises PV generation data of varying resolutions from across the UK, with historic data spanning up to seven years. The dataset is supplied by a combination of homeowners and commercial sources and includes both domestic and commercial scale installations with a wide range of azimuth and elevation angles.

## Method

Using half-hourly data from the MgDB dataset from across the UK, generation time series are compared between a sample of system pairs chosen at random from all possible pair permutations. The sample of systems is chosen for the month of July 2012 with the condition that the azimuth be between 160 and 200 degrees clockwise from North and the elevation between 20 and 40 degrees from horizontal, which yields 648 systems. The geographical spread and distribution of azimuth and elevation angles for the sample is shown in Figure 2. Of these systems, 11 are excluded due to suspected system failures during the period[1]. To avoid excessive computation time, it is necessary to only consider a proportion of the possible permutations (202566) of system pairs and as such, 20% of the possible permutations are chosen at random. To minimise the effect of shading, only data between 10am and 2pm is considered, since the vast majority of system shading occurs in the morning and evening (when the sun is lower in the sky). Correlation between time series is measured by the Pearson correlation coefficient [2]. See Figure 3 of the Appendix for more detail on the algorithm used to acquire the results.

[1] Systems are treated as failing if there are more than $x$ zero readings in the time series, where $x = 20\% \times total\ number\ of\ readings$.

# Results

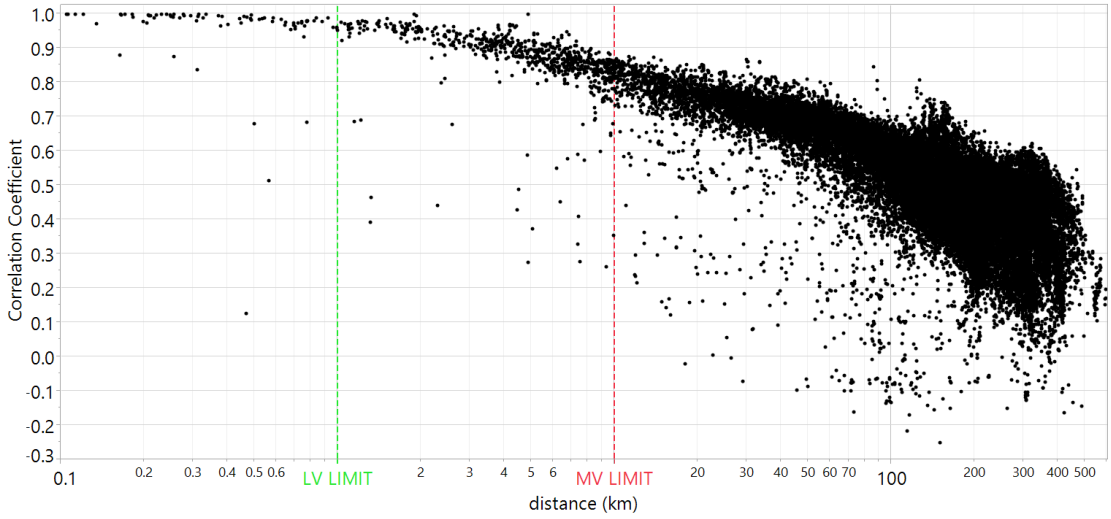


*Figure 1; Plot of correlation between generation time-series of PV systems at a range of length scales.*

It can be seen from Figure 1 that the correlation in the generation time-series decays exponentially with the system separation. Separations on a length scale less than or equal to the approximate length scale of low voltage (LV) networks show a coefficient close to 1, indicating a very high degree of correlation. The correlation varies considerably for length scales between those approximating LV and medium voltage (MV) network limits, which for the purposes of this report are taken as 1 and 10 km respectively. Since the Pearson correlation coefficient does not provide a well-defined measure of what level of correlation can be considered critical, it is not possible to determine from Figure 1 exactly what the critical length scales should be, however it is clear that the critical length scale lies somewhere below the MV (10 km) limit. Crucially, below 1 km separation generation for similarly oriented systems is totally correlated. LV networks lie in this spatial limit and should be modelled appropriately. Between 1 km and 10 km separation there is a drop in correlation, This drop is important for MV network performance since in some cases there will be high correlation and in others not. MV networks will need to be analysed on a case by case basis. At the national scale, at ~100 km separation, correlation is substantially reduced and the national PV generation can be considered to be well modelled by a representative ensemble from across the whole nation.

# References

[1] Sheffield Solar, “Microgen Database,” Sheffield Solar - University of Sheffield, [Online]. Available: http://www.microgen-database.org.uk/.

[2] K. Pearson, “Notes on regression and inheritance in the case of two parents,” *Proceedings of the Royal Society of London,* vol. 58, pp. 240-242, 20 June 1895.

# Appendix

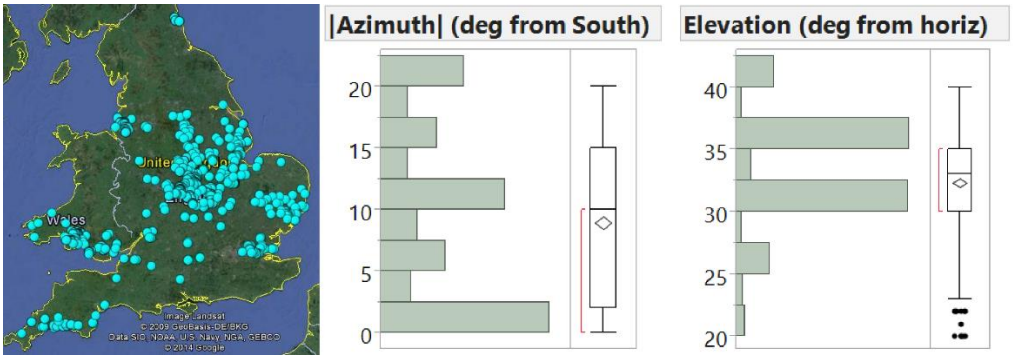


*Figure 2; Geospatial and geometric distribution for the 648 systems used in this analysis. The mean |Azimuth| is 9.0° from South and the mean elevation is 32.4° from horizontal.*

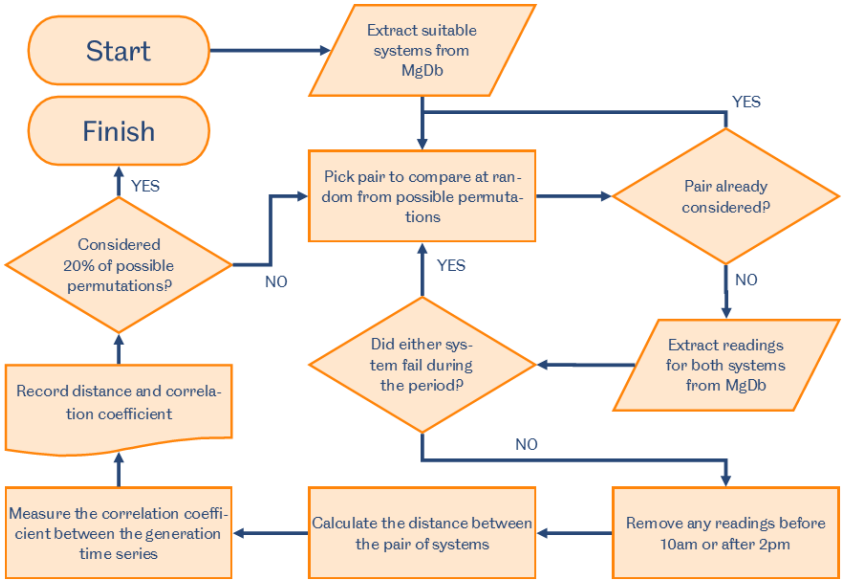


*Figure 3; Flowchart showing the simplified computational process to produce the results*

# Acknowledgements

The Microgen Database is a public-industry-academic collaboration providing solar photovoltaic performance data for use across the UK PV supply chain. The project is funded by the EPSRC and The University of Sheffield. It relies on data donation from individual domestic PV owners and commercial PV operators. If you have data and would like to contribute for the common good of UK PV please visit www.microgen-database.org.uk.

In particular, Sheffield Solar would like to acknowledge PassivSystems (www.passivsystems.com) for donating the PV generation data used in this analysis.

| The University Of Sheffield. | EPSRC Engineering and Physical Sciences Research Council | |
|---|---|---|
| Funded by The University of Sheffield | Solar Energy for Future Societies (Grant Code: EP/I032541/1) | WISE PV (Grant Code: EP/K022229/1) |