2014 ASIA-PACIFIC
SOLAR RESEARCH CONFERENCE

Correcting Daily Output Data from Distributed PV Systems for Performance Analysis

Navid Haghdadi[1,3], Anna G Bruce[1], Jessie K Copper[1], Emily J Mitchell[1], and Iain MacGill[2,3]

[1]*School of Photovoltaic and Renewable Energy Engineering University of New South Wales, Sydney 2052, Australia*
[2]*School of Electrical Engineering and Telecommunications, University of New South Wales, Australia, Sydney 2052, Australia*
[3]*Centre for Energy and Environmental Markets - CEEM, University of New South Wales, Australia, Sydney 2052, Australia*

E-mail: n.haghdadi@student.unsw.edu.au

Performance analyses of installed photovoltaic (PV) systems are essential for understanding the yield of PV systems, and hence the cost, quality and reliability of the electricity they provide. Such analyses can be used to identify loss mechanisms in PV systems and to improve performance predictions.

There are more than 1 million rooftop PV systems already installed in Australia, but only a few good quality publicly available PV output data sets that cover a limited geographical area are available. Data from significant numbers of small PV inverters are being contributed to publicly available online PV monitoring web sites such as PVOutput.org. However, despite the great potential value of these data for engineering studies, they have limited usefulness due to poor quality of the data and missing information. Large numbers of systems in this database contribute only daily output data, so any statistically robust comparison of PV system performance across different climate regions must make use of daily data. This paper presents a method for evaluating the existing recorded daily PV output data and cleaning and completing these datasets for further PV performance evaluation studies.

First, initial cleaning is performed based on statistical analysis of the output. Second, we identify any missing data or underperforming PV systems based on investigating historical daily output of PV systems. The temporal and spatial correlations between different PV systems are calculated and used to identify reporting or monitoring problems, or partial or complete system failure. Erroneous data is removed from the dataset, which can then be used for performance analysis, particularly analyses relating to a group or groups of systems where the missing data will not bias the results. Additionally, using the relative expected performance of PV systems, missing data can be filled with the most likely value. In this way, the existing dataset can be made complete for the purpose of assessing individual system performance. Further, the validity of each existing recorded data can be measured and a 'validity index' for each PV system can be reported to assist selection of data for further studies.

The method has been applied to the very large PVOutput.org dataset, consisting of more than 2 million records of daily energy generation for more than 7,000 PV systems located in different climate regions of Australia. The method has the potential to make much more data useful for performance analysis with improved confidence, as it can reduce errors in user-entered system parameters, reporting and communication, and does not rely on measured or modelled weather data, which is not available for all regions of Australia.