Topic: 5.1
Title: **Improved 'nowcasting' of residential PV generation using clustering**

Navid Haghdadi[1, a, i], Anna Bruce[2, a], and Iain MacGill[3, b]

[1] n.haghdadi@student.unsw.edu.au [2] a.bruce@unsw.edu.au [3] i.macgill@unsw.edu.au

[a] School of Photovoltaic and Renewable Energy Engineering, University of NSW Australia, Sydney, Australia

[b] School of Electrical Engineering and Telecommunications, University of NSW Australia, Sydney, Australia

[i] Corresponding author, address: University of NSW Australia, Sydney, NSW 2052, Australia, +61450816639

Summary: The number of small scale grid-connected photovoltaic (PV) systems installed in the low voltage distribution network is increasing, and predicting the output power of these systems is crucial for better understanding their impacts on the power systems. Since on-line monitoring of these PV systems is currently sparse, upscaling techniques can be used to estimate the real-time generation of of all PV generation in a geographical region, using on-line measured output data from a subset of the PV systems. In this paper a new clustering method is proposed for selecting the most representative PV systems for such upscaling to improve PV generation 'nowcasts'ug. The relationships between the output of different PV systems are established by, first, temporally clustering the historical half-hourly output power of all PV systems to establish partitions containing similar time periods; and second, clustering all PV systems to partition similarly performing PV systems. In this way, representative PV systems can be selected that, with on-line monitoring, best 'nowcast' the output of all systems. For each temporal cluster, the output power of any PV system, or a group of PV systems, can be expressed as a multivariate polynomial, and this model can be used for estimation of any new instance for which the output power of representative systems are known. The proposed clustering algorithm is evaluated using tenfold cross validation.

Purpose of the work: With a large and growing number of largely unmonitored PV systems connected to distribution networks, estimates of their overall generation can be a valuable input to operational power system processes including network management, intraday PV forecasting, and scheduling, balancing and the preparation of market offers in electricity systems [refs]. Since online monitoring, communicating and processing of data from very large numers of small PV systems is expensive and difficult, it is useful to be able to use a subset of on-line data to predict the output of some or all of the other PV systems in real time (nowcasting). A method is proposed to nowcast the output power of a group of PV systems using online monitoring of some of them based on previously identified relationships derived from the historical metered half hour data of generation from all systems.

Approach: In this paper a new method is developed for spatial and temporal clustering of PV system performance in an area, using a Fuzzy clustering approach. Spatial clustering is used to identify and partition similar PV systems by considering the historical metered data of each system as an observation with the time intervals as features. Temporal clustering uses the data from all PV systems at each time period as an observation, and each PV system as a feature. To reduce the dimension of this clustering problem, principal component analysis is used to reduce the number of features in both clustering steps. One year of half hour output power data for 290 grid-connected residential PV systems in Ausgrid's New South Wales network area are used as a case study. To test the model developed, 90% of the data is randomly selected and used for training the model while the other 10% is used for testing the method. This process is repeated for 10 times, in order to consider all 10% subsamples of data as test data. The final error of the nowcasting approach is obtained by averaging the error of each training and testing process.

Scientific innovation and relevance: The problem of estimating the aggregate output power of PV systems based on on-line monitoring from some of them, and up-scaling techniques has been explored in recent years [refs]. The two main approaches include upscaling of randomly sampled systems [refs], or using criteria to select representative systems [refs]. A stochastic method for finding representative PV systems and temporal and spatial clustering is introduced for the first time in this paper. This method can be used to nowcast one, some or all PV systems in a specific region, based on online monitoring some of them. While the method relies on historical data to find the representative systems, this technique does not require online monitoring, and could be used to develop improved criteria for selection of representative PV systems for forecasting.

Results and Conclusions: An individual PV system's output power can be estimated in real time using the proposed method with a mean absolute error of less than 0.05 W/Wp using 10 temporal and 10 spatial clusters, which was determined to be the optimal number of clusters for this dataset. The aggregated power of each spatial cluster can be nowcasted with an error of 0.02 W/Wp and the aggregated output power of all PV systems with error of less than 0.01 W/Wp. A relationship between the result of spatial clustering of PV systems and their geographical location is found, indicating that representative systems for nowcasting may be chosen based on criteria such as location. These patterns will be further examined in future work for larger areas with more different climate conditions.

Explanatory information

A. Steps of the method

In this section an overview of the model developed for nowcasting is presented. First the general scheme of the method with input and outputs are shown and then the procedure of each part of the method is introduced.

A.1 General scheme of the model

As depicted in Figure 1, the final aim of method is to develop a model for predicting output power of PV systems in a given area based on monitoring a portion of them. The input to the model is the output power of representative PV systems for a particular time instance, and the output of the model is the output power of one, some or all PV systems for that instance. Input values are first labeled by means of KNN method [refs] and then after identifying the temporal cluster to which these values belong, the output of other PV systems are predicted using the parameters of the model that have been previously determined using the training data.

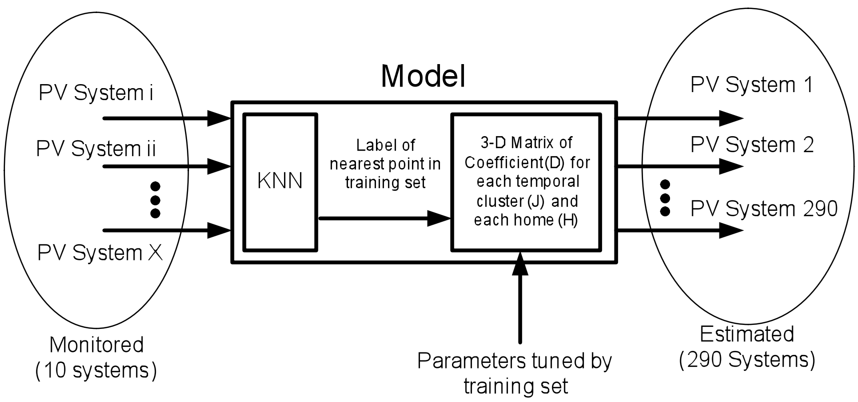

Figure 1.General scheme of the developed model

A.2 *Finding representative PV systems*

Nowcasting based on scaling-up relies on selecting a subset of PV systems that are representative of all systems. In this paper, a new clustering-based approach for finding these systems is introduced. First using historical data, all PV systems are portioned into different clusters with similar behaviour. For each cluster, a representative is selected. In this method each system is considered as an observation and each instance as a feature. Since the number of features is much higher than the number of observations, before the start of clustering, it is necessary to reduce the number of features by a feature (dimension) reduction method. The Principal Component Analysis (PCA) approach is used and the number of features is reduced as much as possible while retaining 95% data separability. Then all systems are spatially clustered using fuzzy clustering method and the representative of each cluster is found. Figure 2 shows the entire procedure.

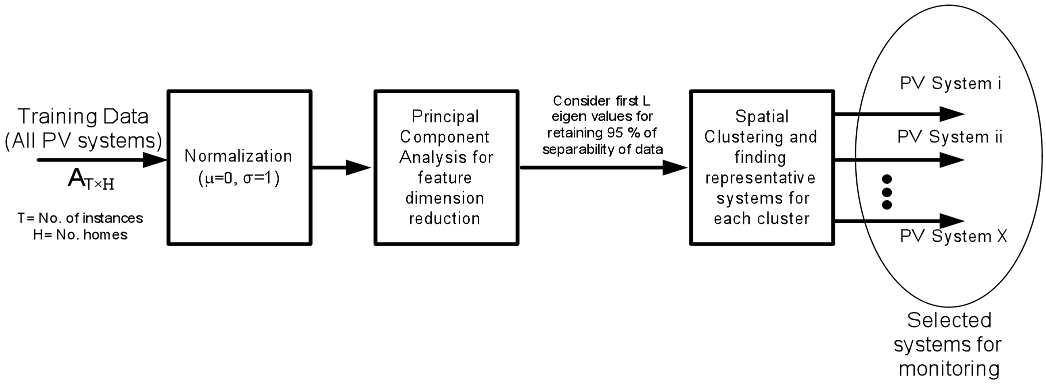

Figure 2. Procedure of finding representative systems

A.3 *Tuning the parameters of model*

The 3 dimensional coefficient matrix is tuned by multivariate polynomial fit. (1) shows the formula of this matrix for each system, *i,* in each temporal cluster and *J* is the total number of temporal clusters, R_K is the value of the representative system and *C* denotes the coefficients of the matrix. First, all PV systems are temporally clustered and then for each cluster the relationship between representative systems and other systems are modeled.

□□ □□□□□□□□=□□1=□□□,1,□×□1□1×□2□2×…×□10□10 …
∀ □1+□2+…+□10≤□ □□□=□□□,□,□×□1□1×□2□2×…×□10□10
(1)

B. Implementation of the method

290 roof mounted solar homes located in the New South Wales franchise region of Australia's largest distribution network service provider, Ausgrid , are considered as a case study for implementation of the method. The preliminary results from nowcasting, along with the error of estimation is shown in the following figures.

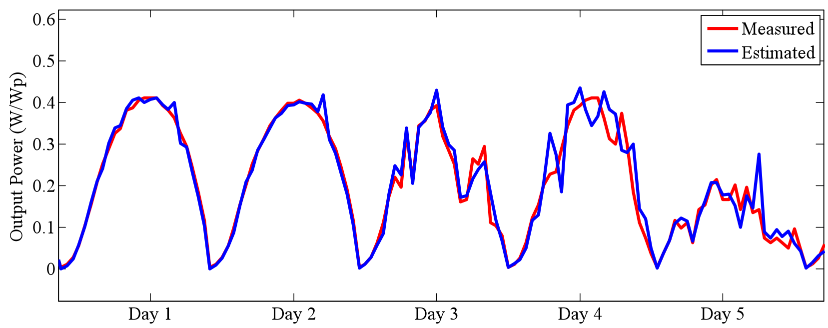

Figure 3. Comparison of estimated and measured values for one home in 5 consecutive days (night times are removed) *might want to place a small dashed line between the days to make clear that night period has been removed … just might look a bit strange to some readers*

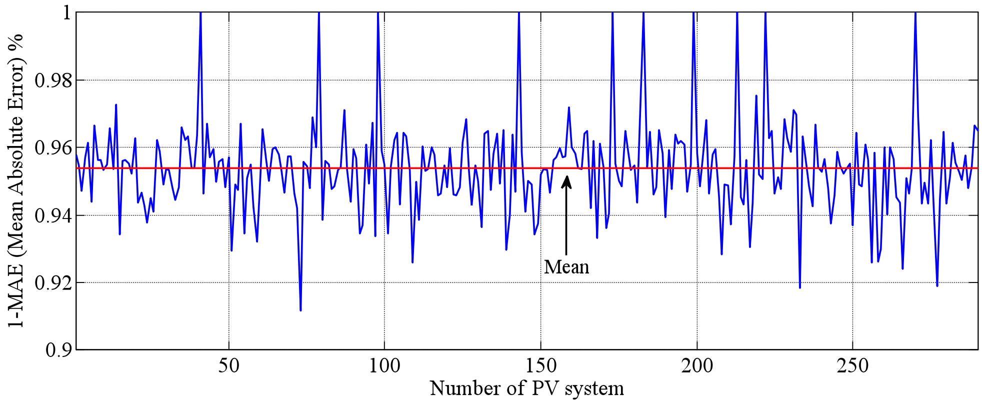

Figure 4. Accuracy of Estimation for all homes

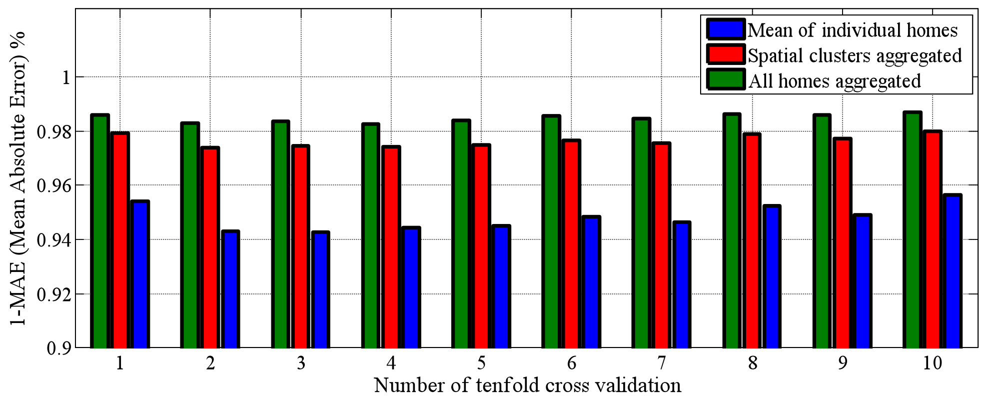

Figure 5. Error of estimation for each home and for aggregating homes together

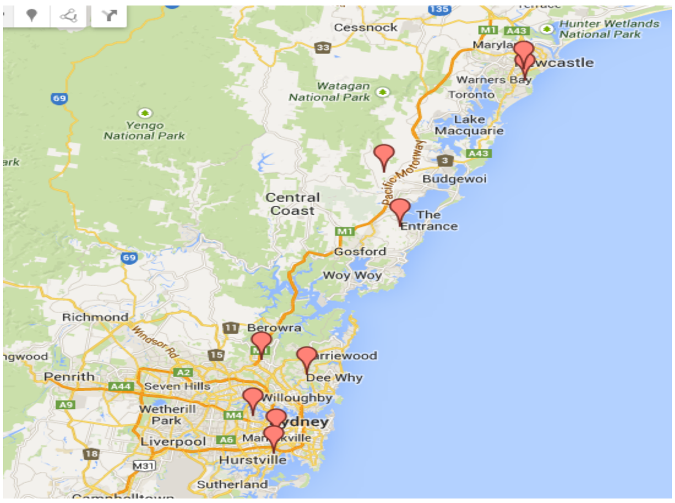

Figure 6. Geographical distribution of representative PV systems