ASIA-PACIFIC SOLAR RESEARCH CONFERENCE

Peer Reviewed
2017 Solar Research Conference

A method for classifying households to help forecasting their Photovoltaic electricity self-consumption patterns

Baran Yildiz[1], Jose Bilbao[1], Alistair Sproul[1] and Jonathon Dore[2]
[1]*School of Photovoltaic and Renewable Energy Engineering, University of New South Wales, Sydney, NSW, 2052, Australia*

[2]*Solar Analytics Pty Ltd., 9/245 Chalmers St, Redfern, NSW, 2016, Australia*

E-mail: baran.yildiz@unsw.edu.au

Abstract

Smart meter data can be used for various purposes within smart grids, including residential energy applications, such as Home Energy Management Systems (HEMS) and Battery Energy Management Systems (BEMS). Considering the low feed-in tariffs for rooftop photovoltaic (PV) and increasing customer electricity prices, maximizing PV self-consumption becomes a key objective for these energy management systems. This paper analyses the impacts of household electricity load consumption profile and PV size on PV self-consumption. A clustering model has been developed to classify households according to their daily load and generation profiles and PV size. The study is then extended to analyse the influence of different seasons on the self-consumption forecast. The results show that the clustering model can guide HEMS and BEMS in deciding more accurate strategies for forecasting day-ahead PV self-consumption.

1. Introduction

Year 2016 set a new record for PV installations with 70 GW new PV added globally (Climate Council 2017). Australia, being one of the world leaders for roof top PV systems, now has over 20% of houses owning PV systems, reaching up to 1.6 million rooftop systems (APVI & IEA-PVPS 2017). Previously, solar bonus schemes with high feed in tariffs have supported the uptake of PV and were common in many countries such as Germany, Australia and Italy. However most of these schemes have been abolished or reduced to lower rates and in particular for Australia, feed-in tariffs for PV systems are lower than electricity retail usage rates. As a result, household owners became financially more incentivized to directly consume the generated PV electricity (also known as PV self-consumption) rather than exporting it back to the grid. Besides the financial profits to the owners, increased PV self-consumption can also reduce stress on the electricity distribution grid, help with frequency and voltage regulation, while reducing the requirement for relatively expensive gas-fired electricity generators that only work during peak periods (Luthander et al. 2015). Therefore, there is significant interest in studying PV self-consumption, and in developing energy management tools that maximize self-consumption (Matallanas et al. 2012).

One common strategy to maximize PV self-consumption is to use energy storage, such that the excess PV generation can be stored in batteries or hot water tanks to be used at a later time by the household. Another common strategy is to use load shifting, which is a popular demand side management strategy (DSM). With load shifting, certain appliances can be shifted from early morning, afternoon, or night periods that have low or zero irradiation, to the periods where there is enough on-site generation. Both strategies can also be used simultaneously (Masa-Bote et al. 2014).

However, PV self-consumption is influenced by the household's daily consumption patterns and the size of the PV system. Hence, it is not an easy task to forecast PV self-consumption for the next day as both load and PV exhibits highly dynamic profiles at the individual household level. This study suggests a useful clustering method to have a better understanding of PV self-consumption profiles, according to consumption and PV generation patterns. This method can provide useful information to HEMS and BEMS for deciding an effective energy management and forecast strategy. Furthermore, climate and seasons have impact on self-consumption. For example, self-consumption characteristics can be different for houses with electric heating and cooling between two regions; one where heating requirement is the dominant load when the irradiation is at lower levels and the other where the cooling requirement is dominant when the irradiation is at higher levels. Here, in order to provide important insights on the impact of climate and seasons, household self-consumption profiles are investigated for both winter and summer seasons.

The paper is organized as follows: in section 2, the data set is described and a statistics summary is provided for the households' consumption and PV generation. In section 3, the methodology and application of the chosen clustering method, K-means is described. Obtained results are provided with further discussion in section 4. The paper is concluded in section 5.

2. Data-set

The data-set used consists of 300 households with PV systems randomly selected from the Ausgrid network, an electricity distribution network provider in the Greater Sydney area of New South Wales (NSW), Australia (Ausgrid 2014). Each household's electricity load and PV generation are measured by a gross meter between the period from 1 July 2010 to 30 June 2013. The measurements are in half hour resolution and none of the households have batteries. The average daily load profile of each customer can be seen in Figure 1.

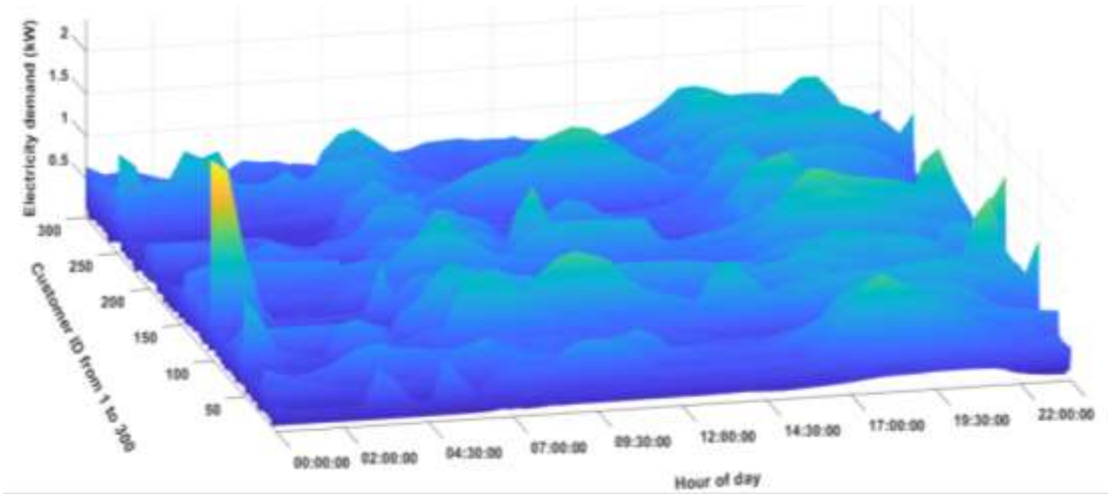


Figure 1 Average half hourly load profile of 300 customers over 3 years

Some other useful summery statistics of the household stock is presented below:

Table 1 Summary statistics of Ausgrid data set

Year	2010-2011	2011-2012	2012-2013
Average daily consumption in kWh (per household)	19.12	18.07	17.50
Average daily generation in kWh (per household)	5.81	5.71	5.98
Average daily generation in kWh/kWp (per household)	3.43	3.37	3.55

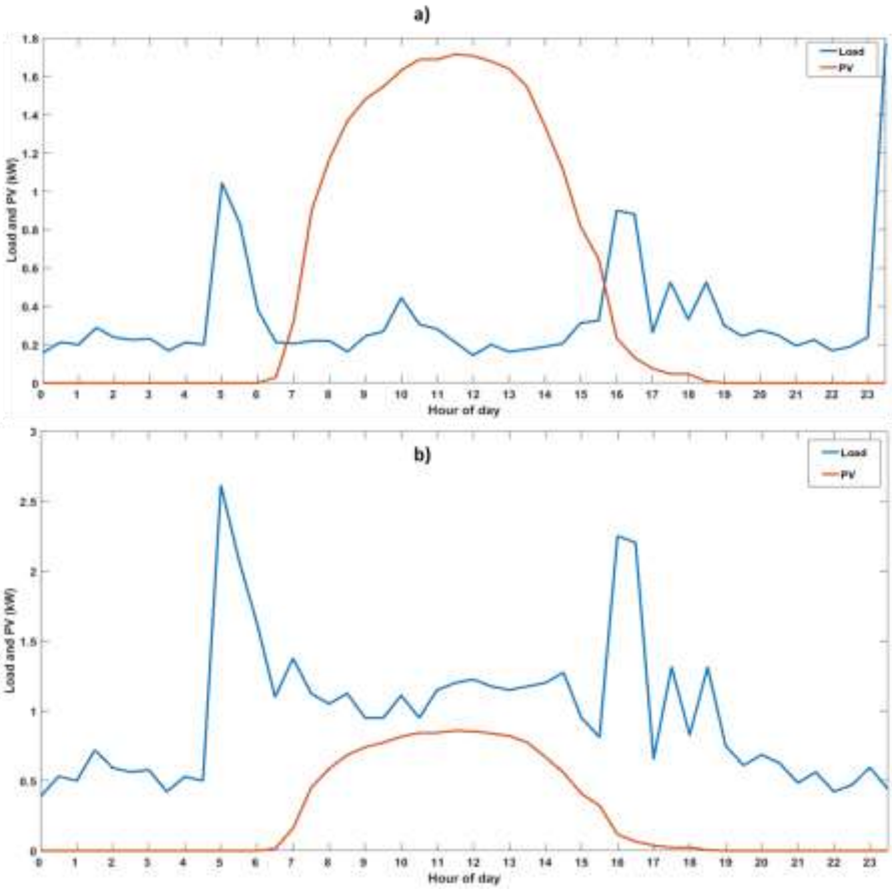
The average system size of the household stock is 1.68 kW. This number is consistent with the average household PV size in Australia back in 2010; however, it does not represent today's systems, since the average size of household rooftop PV installed in the last 5 years is around 4.1 kW (APVI 2017). Another limitation of the study is the half-hourly raw data resolution. It was shown by (Wright & Firth 2007) that averaging higher resolution data such as 1 or 5 minutes to hourly or sub-hourly may give misleading results when calculating PV self-consumption, import and exports. More specific data resolution recommendations were provided by (Beck et al. 2016); for example, 15 minute data could be sufficient resolution for calculating PV self-consumption profiles for households with moderate consumption.

3. Methodology

There has been different ways of defining PV self-consumption in the literature; however, in this paper PV self-consumption will be defined as the amount of PV generation directly consumed by a household, also referred as absolute PV self-consumption (Widén 2014). Hence, PV self-consumption (PVSC) is equal to the minimum between the PV generation and the electricity load at any point, as shown in equation 1 below. Note that, this statement is valid for net-metering arrangements and may differ for other metering arrangements.

$$\square\square\square\square(\square) = \min[\square\square\square\square(\square), \square\square(\square)] \quad (1)$$

The following figure includes data from two households with different PV sizes on the 1st of Jan 2010, which illustrates two different PV self-consumption cases. In Figure 2.a, between 7 am to 15.30 pm, the PV generation is greater than the load, therefore the household consumption is completely off-set by the PV generation during this period. In this case, the value of the PV self-consumption is equal to the load. On the other hand, in Figure 2.b, the household load is always greater than the PV generation during the day. As a result, all PV generation is used to off-set a portion of the load and the value of the PV self-consumption is equal to the PV generation



.

Figure 2 Different PV self-consumption cases from two households on 1st Jan 2010: a) PVSC=Load between 7am-15.30pm, b) PVSC=PV at all times during sunshine hours

With the intention to identify households whose regular PV self-consumption profile is similar to one of the examples shown in figure 2, a well-known clustering model, K-means has been utilized. For the purpose of representing PV self-consumption characteristics of households, the following variables were for clustering:

i. Average daily consumption
ii. Average daily generation
iii. The ratio of average daily consumption to average PV generation
iv. The ratio of average daily consumption to PV size
v. The ratio of average daily generation to PV size
vi. PV size

After experimentation and with the purpose of preventing problems pertinent to the curse of dimensionality (James et al. 2006)[1], only two variables, **iii** and **vi** are found suitable to represent each household for this clustering problem. Working with a smaller number of variables also made the results easier to interpret. It was also observed that clustering results obtained by using other variables as household representatives such as **iv-v, iv-vi** or **i-ii-vi** gave similar results, with minor differences in cluster assignments of households.

[1] Curse of dimensionality refers to difficulties associated with analysing high dimensional data. For example, it is easier to plot and analyse data with two or three dimensions compared to data with higher dimensions.

K-means tries to minimise the intra-cluster variation within each cluster, such that points that are most similar to each other are grouped in the same cluster. For measuring the distances between two points in clusters, the well-known Euclidean distance was chosen. To improve the quality of the clustering outcome, chosen variables **iii** and **vi** were normalized with respect to their maximum values, such that both lied within [0-1] range. For more details on K-means clustering procedure, normalization, mathematical representation of the algorithm and Euclidean distance please refer to (James et al. 2006).

K-means model requires the number of clusters to be specified. In order to decide the number of clusters, the Silhouette method is used (Rousseeuw 1987), which measures the similarity of inter-cluster points and dissimilarity between different clusters. A positive and higher number of Silhouette value is more desirable as it represents a more effective clustering outcome. The following figure shows Silhouette values obtained with number of clusters between 2 to 20.

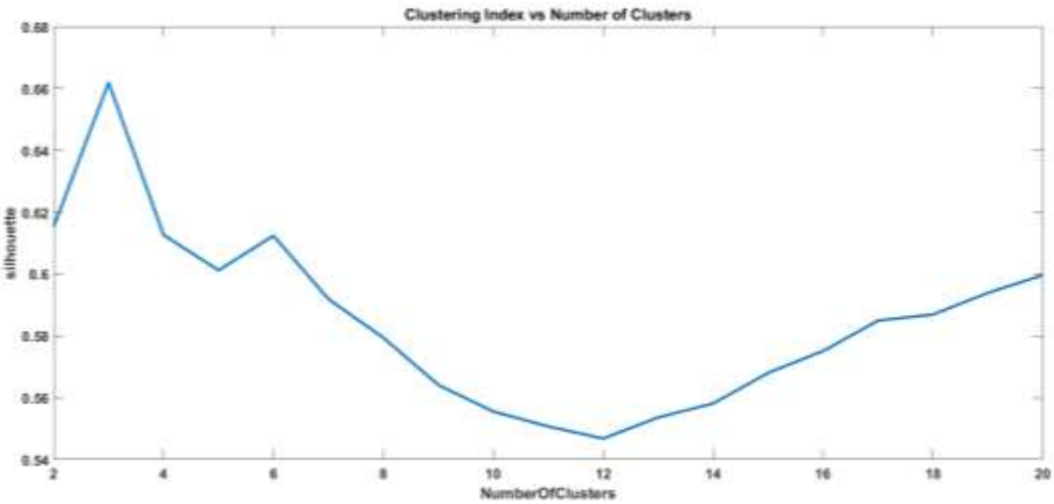


Figure 3 Silhouette values for number of clusters between 2 to 20

It can be seen that the highest Silhouette value was obtained with three clusters and hence the K-means was run to group households into three clusters.

4. Results and discussion

The obtained clustering assignment is shown in the following Figure 4. Each household corresponds to a single point represented by normalized PV size and normalized ratio of average daily load and average PV generation.

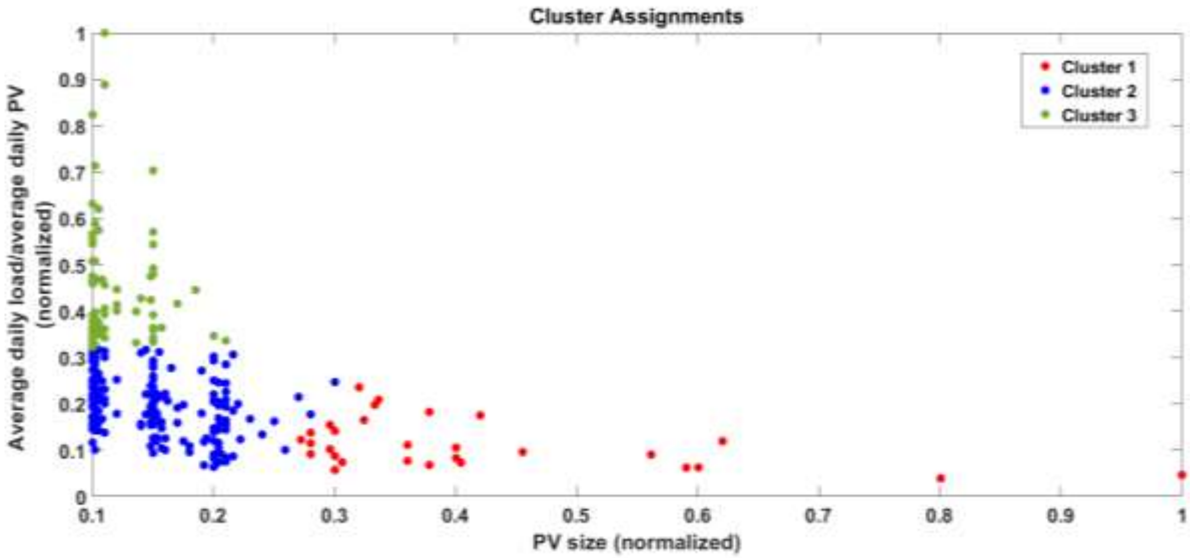


Figure 4 Clustering results obtained by K-means when number of clusters is equal to three

Cluster 1 represents households with larger PV and smaller ratio of average daily load/PV, whereas cluster 3 represents households with smaller PV size and larger ratio of average daily load/PV. The majority of households fall under cluster 2 with a moderate PV size and ratio of average daily load/PV. Figure 5 below shows examples of daily load and PV profiles for 16 days from summer and winter seasons taken from three example households belonging to clusters 1, 2 and 3 respectively. The example household (ID: 74) from cluster 1, exhibits load profiles smaller than its PV generation profiles during sunshine hours except in early mornings and late afternoons, where solar irradiation is very low. Other example household (ID: 228) from cluster 3, exhibits load profiles greater than its solar generation almost all times regardless of the season. The remaining seasons (autumn and spring) showed a similar behavior for these clusters. There were only a limited number of days where due to unexpected load consumption or whether conditions, these typical profiles were altered.

On the other hand, this regular type of PV and load profile relationship was not clearly observed for the households in cluster 2, as shown by the example household (ID: 253) in the same figure below. There is no clear distinction between days where load or PV is more dominant during sunshine hours. However, some of the households from cluster 2 behaved similar to cluster 3 during summer, where PV was mostly greater than the load, but behaved similar to cluster 2 during winter where load was mostly greater than PV.

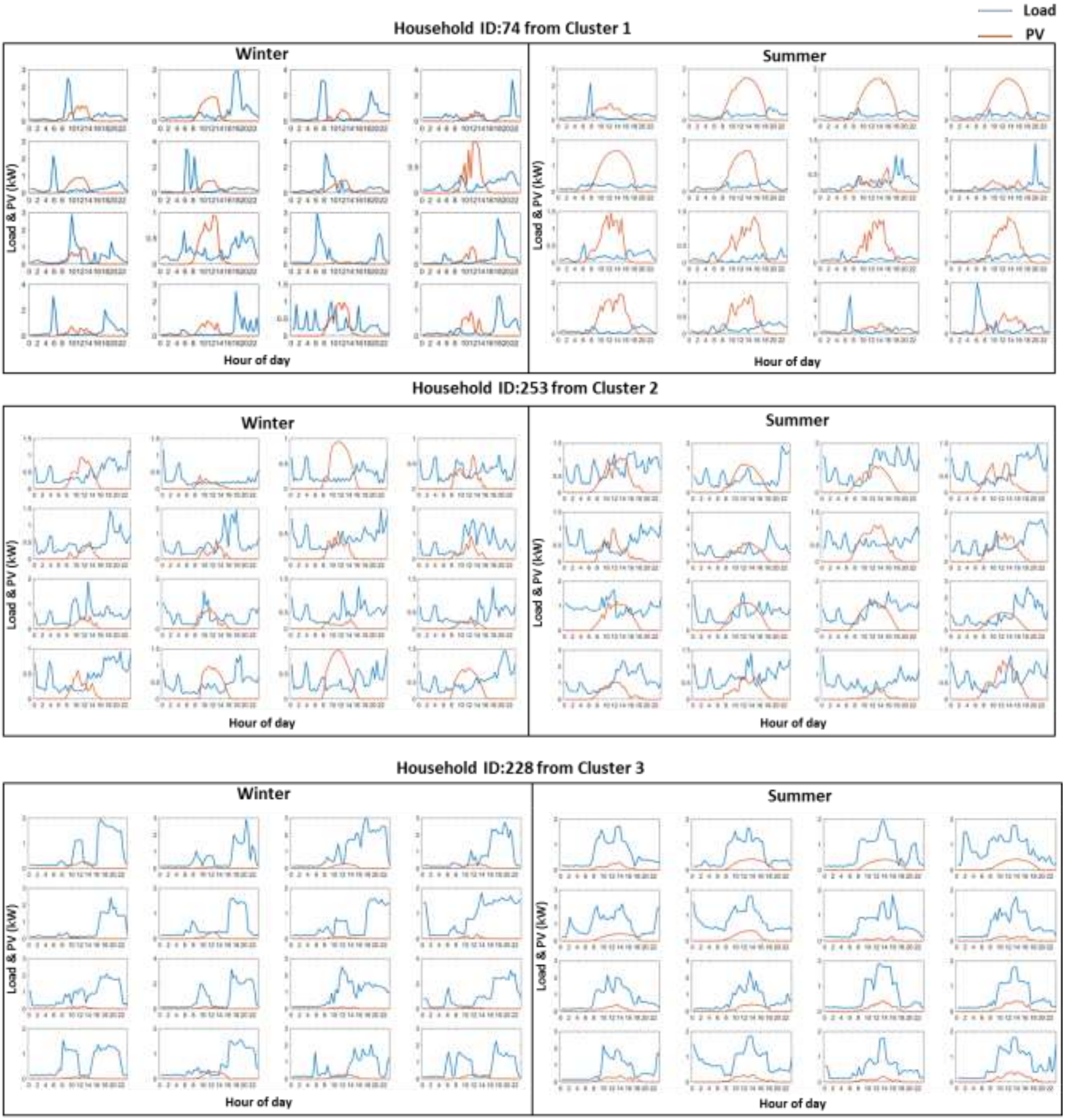


Figure 5 Example household load and PV profiles for two weeks from summer and winter seasons

As mentioned previously, in order to predict PV-self consumption for the next day, both PV and load forecasts are required. For each time point, the smaller value of these forecasts constitutes the PV self-consumption forecast. Previous research has shown the high variability and uncertainty around household level load and PV forecasts (Yildiz et al. 2017). Therefore, requiring only one of these forecasts instead of both can be advantageous in terms of forecast accuracy. In particular, for households which fall under cluster 1, load forecasts will be more relevant and may provide adequate information for most days, especially sunny ones. On the other hand, in order to predict next day's PV self-consumption for a household which falls under cluster 3, PV forecasts will be more relevant than load forecasts. For a household whose regular PV and load relationship depends on the season such as some households from cluster 2, PV self-consumption forecast strategy can be altered between summer and winter months using cluster 1 and cluster 3 strategy respectively. Future research aims to investigate the forecast accuracy of PV-self consumption using these strategies for different clusters in order to validate the implications of the clustering.

Besides PV-self consumption forecast, estimation of imports and exports can be highly valuable for HEMS, BEMS and utilities in planning for next day's electricity operations. Especially for households which fall under cluster 1 and 2, next day's predicted exports can be planned to be utilized by energy storage, or demand side management strategies, such as load shifting. Once again, both load and PV forecasts are required in order to predict the next day's imports and exports where subtracting PV generation and load forecasts from each other gives the import and export forecasts. At this point, future research is also aimed to further investigate different forecast strategies for these clusters in predicting next day's imports and exports. Specifically, for certain households, using historical import/export time series by itself within a Smart Meter Based Model (SMBM) framework (Yildiz et al. 2017) may increase predictive accuracy compared to forecasting two independent time-series (load and PV) and subtracting them from one another.

Another implication of this clustering model is when deciding on purchasing storage options. In particular, if a household is thinking of purchasing a battery, it can first be assigned to one of these clusters according to its PV size and historical average daily load/PV ratio. If the household falls under cluster 3, then it may be advised not to buy batteries since all its PV generation is used to off-set load for most of the days. However, if it falls under cluster 1, there may be more confidence in advocating for purchasing a battery as there is excess PV generation in most of the days.

In order to test the validity of this method on a data set which is more representative of today's average PV size, 4.1 kW, some preliminary analysis has been carried. Whilst assuming the consumption profiles remained constant, each household's PV size was multiplied with a constant to give the average household stock PV size as 4.1kW. With the new simulated PV size and profiles, the K-means clustering was applied and it was observed that there was insignificant change in the household clustering assignments such that most households were assigned to the same clusters and the number of households in each cluster did not change. This is expected since the new clustering variables of the simulated data set **iii** and **vi** were multiples of previous data-set's variables caused by the multiplication of PV size. However this time, clusters had different implications. In particular for the new cluster 1 households, the excess PV amount significantly increased compared to previous cluster 1. For the new cluster 2, more households exhibited profiles similar to the previous cluster 1 households with most days of excess PV. Finally, for the new cluster 3, it was observed that the amount of imports significantly reduced and some households showed profiles similar to previous cluster 2 households where there PV self-consumption characteristic varied across seasons. It is important to note it is possible to find new clustering variables for the simulated data set, instead of using **iii** and **vi**. This may result in more effective clustering outcome such that the clusters preserve their characteristic load and PV profiles but the number of households assigned to each cluster changes. For example, due to the increase in PV size, more households could be assigned to cluster 1 and less households to cluster 3 while both clusters preserving their characteristic profiles.

5. Conclusions

This study proposes a clustering method to group households according to their PV size and average daily load/PV ratio. Results show that households which fall under different clusters exhibit distinctive load and PV profile relationship. When a new household is assigned to one of these clusters, important preliminary information can be provided to HEMS and BEMS, which can then use the information to devise appropriate forecast strategies for PV self-consumption, import and exports. Future research is aimed to find more affective clustering variables which can be more effective for recent data-sets with higher average PV size. Furthermore, future research is going to further investigate different strategies for forecasting PV self-consumption, import and exports.

References

APVI, 2017. *Mapping Australian Photovoltaic installations*,

APVI & IEA-PVPS, 2017. National Survey Report of PV Power Applications in 2016.

Ausgrid, 2014. Solar home electricity data. , (August), pp.2013–2014. Available at: http://www.ausgrid.com.au/Common/About.

Beck, T. et al., 2016. Assessing the influence of the temporal resolution of electrical load and PV generation profiles on self-consumption and sizing of PV-battery systems. *Applied Energy*, 173, pp.331–342. Available at: http://dx.doi.org/10.1016/j.apenergy.2016.04.050.

Climate Council, 2017. State of Solar 2016: Globally and in Australia. *Climate Council of Australia Limited*.

James, G. et al., 2006. *An Introduction to Statistical Learning*, Available at: http://books.google.com/books?id=9tv0taI8l6YC.

Luthander, R. et al., 2015. Photovoltaic self-consumption in buildings: A review. *Applied Energy*, 142, pp.80–94. Available at: http://dx.doi.org/10.1016/j.apenergy.2014.12.028.

Masa-Bote, D. et al., 2014. Improving photovoltaics grid integration through short time forecasting and self-consumption. *Applied Energy*, 125, pp.103–113. Available at: http://dx.doi.org/10.1016/j.apenergy.2014.03.045.

Matallanas, E. et al., 2012. Neural network controller for Active Demand-Side Management with PV energy in the residential sector. *Applied Energy*, 91(1), pp.90–97. Available at:

ASIA-PACIFIC SOLAR RESEARCH CONFERENCE

http://dx.doi.org/10.1016/j.apenergy.2011.09.004.

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C), pp.53–65.

Widén, J., 2014. Improved photovoltaic self-consumption with appliance scheduling in 200 single-family buildings. *Applied Energy*, 126, pp.199–212.

Wright, A. & Firth, S., 2007. The nature of domestic electricity-loads and effects of time averaging on statistics and on-site generation calculations. *Applied Energy*, 84(4), pp.389–403.

Yildiz, B. et al., 2017. Recent advances in the analysis of residential electricity consumption and applications of smart meter data. *Applied Energy*, (January), pp.1–25. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0306261917314265.

Yildiz, B. et al., 2017. Short-term forecasting of individual household electricity loads with investigating impact of data resolution and forecast horizon. *Proceedings of the World Renewable Energy Conference*.

Acknowledgements

This research was funded by the CRC for Low Carbon Living Ltd. supported by the Cooperative Research Centre program, an Australian Government initiative. One of the authors (B. Yildiz) acknowledges scholarship funding provided by the Cooperative Research Centre for Low Carbon Living Ltd., as well as a scholarship from the Research Training Program of the Australian Government and the ongoing support of Solar Analytics.