

RESEARCH ARTICLE

OPEN  ACCESS

# Short-term forecasting of individual household electricity loads with investigating impact of data resolution and forecast horizon

Baran Yildiz\*, Jose I. Bilbao, Jonathon Dore, and Alistair B. Sproul

School of Photovoltaics and Renewable Energy Engineering, University of New South Wales, Sydney, NSW, Australia

Accepted: 24 April 2017

**Abstract.** Smart grid components such as smart home and battery energy management systems, high penetration of renewable energy systems, and demand response activities, require accurate electricity demand forecasts for the successful operation of the electricity distribution networks. For example, in order to optimize residential PV generation and electricity consumption and plan battery charge-discharge regimes by scheduling household appliances, forecasts need to target and be tailored to individual household electricity loads. The recent uptake of smart meters allows easier access to electricity readings at very fine resolutions; hence, it is possible to utilize this source of available data to create forecast models. In this paper, models which predominantly use smart meter data alongside with weather variables, or smart meter based models (SMBM), are implemented to forecast individual household loads. Well-known machine learning models such as artificial neural networks (ANN), support vector machines (SVM) and Least-Square SVM are implemented within the SMBM framework and their performance is compared. The analysed household stock consists of 14 households from the state of New South Wales, Australia, with at least a year worth of 5 min. resolution data. In order for the results to be comparable between different households, our study first investigates household load profiles according to their volatility and reveals the relationship between load standard deviation and forecast performance. The analysis extends previous research by evaluating forecasts over four different data resolution; 5, 15, 30 and 60 min, each resolution analysed for four different horizons; 1, 6, 12 and 24 h ahead. Both, data resolution and forecast horizon, proved to have significant impact on the forecast performance and the obtained results provide important insights for the operation of various smart grid applications. Finally, it is shown that the load profile of some households vary significantly across different days; as a result, providing a single model for the entire period may result in limited performance. By the use of a pre-clustering step, similar daily load profiles are grouped together according to their standard deviation, and instead of applying one SMBM for the entire data-set of a particular household, separate SMBMs are applied to each one of the clusters. This preliminary clustering step increases the complexity of the analysis however it results in significant improvements in forecast performance.

## 1 Introduction

Residential buildings make up significant proportion of end-use electricity demand. The U.S. Energy Information Administration projects that around 30% of global electricity end use will be attributed to residential sector by 2020 [1]. However, the residential energy sector has seen limited research interest compared to commercial and industrial sectors [2]. Amongst the forecast studies which targeted the residential sector, the main focus have been on aggregate regional or district level electricity loads over medium term (a week to a year ahead) to long term (more

than a year ahead) horizons. Due to the lack of financial interests, little emphasis has been given to individual household level consumption analysis and forecasts [2,3]. On the other hand, with the recent advancement of smart grid technologies, there is a growing interest for understanding residential electricity consumption at individual household level. In particular, the analysis and accurate prediction of electricity loads at individual household level can improve the effectiveness of smart grid applications; this include smart home energy management systems, battery energy management tools, and demand response operations.

At the same time, there has been a great expansion in the deployment of smart meters globally. Many governments mandate utilities to equip their customers with

\* e-mail: [baran.yildiz@unsw.edu.au](mailto:baran.yildiz@unsw.edu.au)

advanced metering infrastructures (AMI) [4], which is a necessary tool to accommodate smart grid technologies and assist networks in responding and shaping future's residential electricity demand. AMI can monitor the load at various data resolutions, and transmit and receive consumption information between the utility companies and households. AMI data enables household consumption analysis at high resolutions which can be effectively used for very short (less than an hour ahead) and short-term (an hour to a week ahead) consumption forecasts.

Individual household electricity load profiles exhibits high volatility. Especially compared to larger scale loads such as commercial buildings, substations or regional electricity loads, load variance and uncertainty is much greater at the household level. This is mainly caused by the unpredictability of occupancy behavior and relatively smaller base loads [5]. Therefore, it is a greater challenge to forecast electricity loads at individual household scale.

Previous load forecast studies on individual households which utilized smart meter and weather data within machine learning models, or smart meter based models (SMBM), mainly worked with a particular data resolution; with common forecast horizons chosen to be either one hour or 24 h ahead. Edwards et al. [6] carried hour ahead hourly forecast analysis for three research houses in Tennessee. The houses were unoccupied and fitted with automatically controlled appliances to create simulated occupancy. Authors used artificial neural networks (ANN) and least squares support vector regression (LS-SVR) and the latter method achieved the best overall mean absolute percentage error (MAPE) ranged between 16 and 21% for the three households. Ahmed et al. [7] used ANN to predict day ahead household loads with 15 min resolution. Although the household stock consisted of 28 households, authors presented results for a single household for a limited of 15 days test period where the model achieved MAPE of 13.90%. Rodrigues et al. [8] also used ANN, but they focused on daily maximum loads in addition to the hourly loads. The authors only presented results for two household for a small test period of three days where the MAPE were within 23.5% for the hourly loads. Ghofrani et al. [9] carried forecast analysis for various very short-term horizons; 15, 30 and 60 min-ahead however, the results were reported for only a single test day and the overall MAPE was 12.9%, 18.3% and 30.4% for the three horizons respectively. Gajowniczek et al. [10] forecasted day ahead hourly loads of sample households by using ANN and SVR models. The results were reported for a single household where the MSE was around 0.1 kWh. It is important to note that example load profiles shown in the study had small volatility with load values less than 0.5 kWh.

This study incorporates data from 14 households sourced from Solar Analytics Solar Smart Monitor data base. Households are located in the state of New South Wales (NSW), Australia and have at least one year worth of 5 min resolution data. ANN, support vector regression (SVR), and LS-SVR models, are used as they have previously shown promising results [6,11].

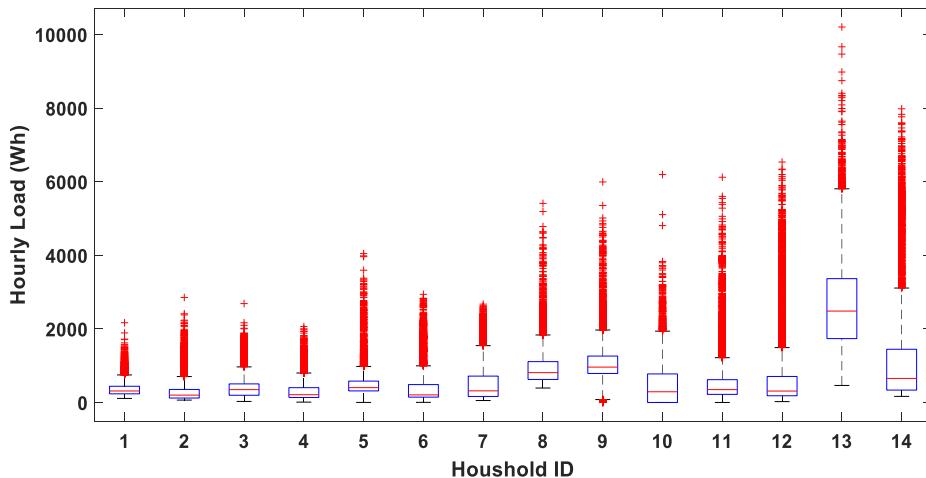
To address the limitations found so far in the literature, our analysis is focused on the following points:

- in order for SMBM to capture seasonal effects at different scales such as daily, weekly, monthly, and seasonally, households with at least a complete year worth of data were analysed;
- for results to be more comparable and meaningful, our study first analyses the households according to their load profile volatility. Performing and reporting separate results for different types of households gives important insights on the relationship between household load standard deviation and SMBM performance;
- instead of working with single data resolution and forecast horizon, our study extends the forecasts over four different data resolution; 5 min, 15 min, 30 min and 60 min, each analysed at four different horizons; 1 h, 6 h, 12 h and 24 h ahead. Results of the analysis carries great importance for the smart grid applications mentioned above;
- the application of one SMBM to the entire data period might not give sufficient amount of detail on the relationship between forecast performance and types of daily profiles. Therefore, for each one of the analysed households, similar daily load profiles are grouped together according to their standard deviation by the use of a pre-clustering step. Hence, instead of applying one SMBM for the entire period, separate SMBMs are applied to each one of the clusters.

The remainder of this paper is organized as follows: In [Section 2](#), the SMBM is introduced, followed by descriptions of the forecast and clustering methodologies. In the first part of [Section 3](#), results for the initial forecast and statistical analysis are presented. Next, results for the data resolution vs. forecast horizon analysis are shown with discussing the relevant reasons behind the impact of these two parameters on forecast performance. The section is concluded with presenting the results obtained from the combined clustering and forecast method. Finally, concluding remarks are given in [Section 4](#).

## 2 Methodology

SMBM use historical loads and weather variables, alongside with temporal variables (hour of day, day of week, holiday, etc.) as inputs for producing forecasts. The desired data resolution can be chosen depending on the available smart meter data, and the desired forecast horizon can be chosen depending on the application. For this study, four different data resolution were used; 5, 15, 30 and 60 min over four different horizon; 1, 6, 12 and 24 h ahead. For this study, historical load and weather variables are selected based on the autocorrelation selection method [12]. Depending on the household, chosen resolution, and forecast horizon, the predictor matrix,  $X$ , consists of 15–25 variables while target loads are organized in the form of an output vector  $Y$ , which are used to train, validate and test the chosen forecast models (ANN, SVR and LS-SVR). All the data analysis and forecasts models are implemented in MATLAB 2016b, using the Statistics and Machine Learning and Neural Network Toolboxes [13].



**Fig. 1.** Boxplot statistics for the analysed households.

## 2.1 Artificial neural networks (ANN)

ANN models are well known for their superior capabilities in modeling complex, non-linear relationships between input and output [14] and have previously been used in many load forecasting problems [15]. Models used in our study consisted of three layers: a single input, hidden and output layer. In order to choose the optimal number of neurons within the hidden layer, we tried a range of different neurons from 2 to 24 whilst observing the cross validation error. The results show that 12 neurons was the optimal number. For the training, Levenberg-Marquardt Backpropagation method was chosen due to its faster computation time and competitive accuracy over computationally more expensive methods [16]. In order to prevent converging to local minima which is identified as one of the inherent problems of ANN, an ensemble of 100 ANNs are used instead of a single ANN, as suggested by [16].

## 2.2 Support vector regression (SVR)

Support vector machines (SVM), found by Vapnik et al. [17] was originally designed to solve classification problems. SVR is a form of SVM to solve numerical regression problems. One of the advantages of SVR over ANN is that the SVR gives a unique solution that converges to a global minimum [18]. On the other hand, parameters such as  $C$  (cost parameter),  $\epsilon$  (acceptable error range) and Kernel function parameter; need to be optimized in order to obtain the optimum results which is also known as parameter tuning. Amongst various types, radial basis function kernel is used for our forecast problem due to its common use in the area [11,19] For more detailed information on SVR and tuning parameters please refer to [20].

## 2.3 Least squares support vector regression (LS-SVR)

LS-SVR is an extension of SVR which is capable of linking important frameworks into SVR such as; Gaussian process, Bayesian inference (probabilistic interpretations and inference), and Fisher Discriminant Analysis [21]. Unlike

SVR, LS-SVR's main criteria function is based on the least squares method. For the implementation of LS-SVR, a library designed to operate in MATLAB environment, LS-SVM lab Toolbox [22] was used. A similar tuning procedure was used for LS-SVR in order to find the optimum parameters (regularization, Kernel function, and cost parameters). For more details on LS-SVR representation and method please refer to [23].

## 2.4 Clustering

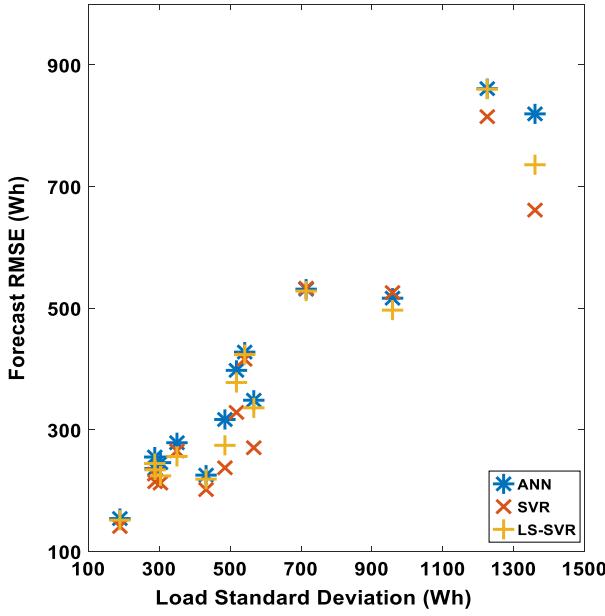
The initial forecast analysis showed that there is a significant relationship between the model performance and household load profile volatility, as shown in Figure 1. Since daily electricity profiles of a household significantly vary over days, using a single model to the entire analysed period may result in limited forecast performance. At this point with the use of an additional clustering step, similar daily load profiles can be grouped according to their volatility and separate models can be applied to each one of these groups. For the clustering of daily profiles, five households are chosen which are representative of distinctive variance, mean and box plot statistics (Fig. 2). For measuring the daily load profile volatility, daily load variance and daily peak index [24] measures were used shown below:

$$\text{Daily Load Variance} : \sum_{i=1}^{24} \frac{(P_i - \mu)^2}{24}, \quad (1)$$

$$\text{DailyPeak Index} : \frac{P_{\text{daily max}}}{\mu}, \quad (2)$$

where  $P_i$  represents household load at hour  $i$ ,  $P_{\text{daily max}}$  represents daily maximum load and  $\mu$  represents the daily average load. Consequently, for a household with  $n$  number of days, the cluster matrix,  $X_{\text{cluster}}$  consists of  $n \times 2$  variables.

For clustering daily load profiles, a well-known clustering technique, *K-means* is used. In our study we



**Fig. 2.** Forecast RMSE performance of the models vs. household load standard deviation.

used the well-known *squared Euclidean* distance [18]. In order to decide for the optimum number of clusters, we used the *silhouette method* which is a measure of similarity of a point to the other points in its own cluster when compared to the points in other clusters [25]. To overcome the potential problems of *K-means*, such as random assignment and finding local minimum instead of global minimum solution [18], *K-means* is ran for 100 iterations and the cluster assignment resulted in minimum within-cluster variation is used for following the forecast step. The methodologies of the described four sections can be found in Table 1.

## 2.5 Error metrics

In order to assess and compare the performances of the forecasting models, the following commonly used three error metrics are used:

$$\text{Root Mean Squared Error (RMSE)} : \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (3)$$

$$\text{Mean Absolute Error (MAE)} : \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (4)$$

$$\text{Mean Bias Error (MBE)} : \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i), \quad (5)$$

where  $y_i$  and  $\hat{y}_i$  correspond to real and predicted load values respectively. In an individual household forecast analysis, the models may perform relatively well on majority of the points, and fail on some other larger loads (spikes). This can significantly influence the error metrics, and hence may

give misleading information about the model performance. Therefore, in order to have a better idea of the model performance we have used additional accuracy metrics shown below [19]:

$$\begin{aligned} \text{For loads smaller than } 500 \text{ W : Accuracy}_{\text{small}} \\ = \begin{cases} 0 & \text{if } |y_i - \hat{y}_i| > 150 \text{ W} \\ 1 & \text{if } |y_i - \hat{y}_i| \leq 150 \text{ W,} \end{cases} \end{aligned} \quad (6)$$

$$\begin{aligned} \text{For loads greater than } 500 \text{ W : Accuracy}_{\text{large}} \\ = \begin{cases} 0 & \text{if } |y_i - \hat{y}_i| > 0.15 y_i \\ 1 & \text{if } |y_i - \hat{y}_i| \leq 0.15 y_i. \end{cases} \end{aligned} \quad (7)$$

Through these metrics, the model performance can be obtained on large and small-scale loads separately. As a result, the relationship between model performance and different scale loads can be investigated in more detail.

## 3 Results and discussion

A summary statistics of the hourly loads of the 14 households are presented in Figure 1 using the Boxplots. It is important to note that household electricity load profiles exhibit significant amount of outliers, shown as red marks. In practice, they correspond to the peak loads (spikes) and occur when electricity intensive appliances are turned-on.

For the initial forecast analysis, a forecast horizon of 24 h is used with hourly resolution data. Figure 2 below shows the RMSE performance of the three models against the household load standard deviation. The plot reveals that increased standard deviation causes an increase in RMSE. This result shows that as the load profiles get more volatile, forecasting becomes more difficult for the models. SVR outperforms the two other models in most households. More detailed results obtained with SVR can be found in Figure 3a and b.

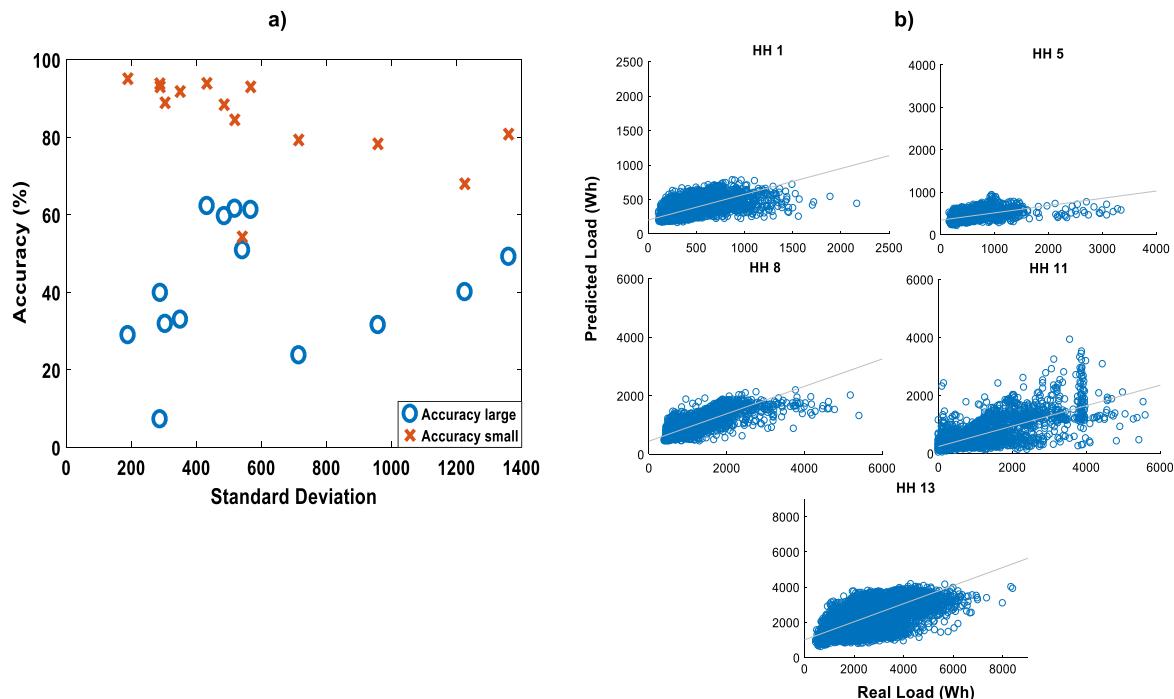
We can see that SVR model shows significantly better accuracy predicting smaller loads which is also observed for the other two models. It is important to note that accuracy metrics depend on the user choice (Sect. 2.5) and hence may vary with different threshold values.

Figure 3b shows forecast results obtained from five representative households (ID: 1, 5, 8, 11 and 13) at hourly resolution and 24 h-ahead horizon. The model is generally under predicting, in particular for larger loads the difference between real and predicted loads are more significant. These results once again show the challenges associated with predicting larger consumption due to use of electricity intensive appliances at unprecedented hours.

The impact of resolution and horizon on forecast performance is investigated on the same five households. SVR model outperformed the others for almost all of the resolution and horizons. The RMSE results are shown in Figure 4. It was observed that for the same resolution, errors stay relatively constant amongst three horizons; 6, 12, and 24 h ahead and smallest errors are obtained at hour-ahead horizon. This can be explained by further investigating the predictor variable selection done by the autocorrelation method. This method aims to select the

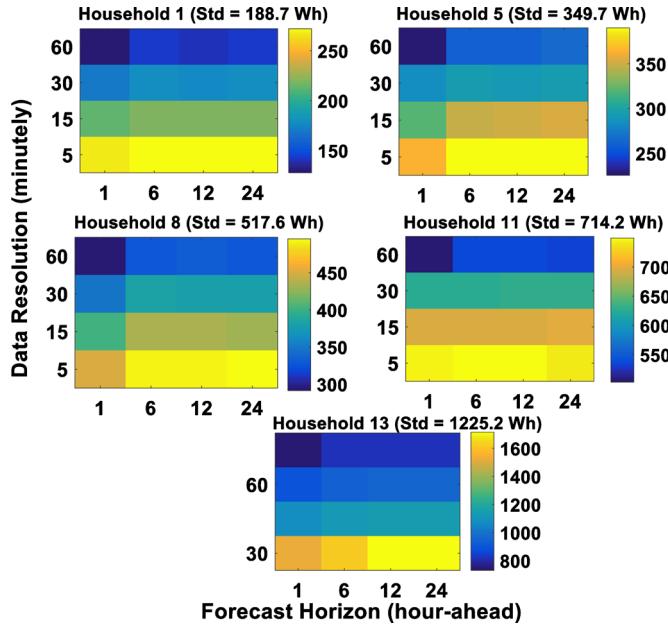
**Table 1.** Summary of the methodology used for the forecasting and clustering analysis.

Step	ANN	SVR and LS-SVR	Clustering and forecast
1	Import and clean raw data from missing points and outliers		
2	Choose a resolution and horizon		Resolution: 60 min, horizon:24 h
3		Organize input matrix $X$ and output vector $Y$	
4	Normalize $X$ and $Y$		Organize $Y$ as 24 hourly daily and re-arrange $X$ accordingly
5	Randomly partition data into 10 folds		Use daily variance and peak index values to be used for clustering
6	Allocate 9 folds for training and validation of the models		Run $K$ -means with optimum number of clusters
7	Tune the ANN ensemble	Tune the SVR and LS-SVR parameters using the training set	Assign daily profiles to the relevant cluster
8	Choose optimum parameters according to the smallest 10 fold cross-validation error		Repeat steps 4–8 of ANN, SVR and LS-SVR methods on the clusters
9	Test the models on the independent test set and store the results in the “ <i>Predictions</i> ” vector		Store results separately for each cluster and also for the entire clustered data in <i>Predictions_wCluster</i>
10	Repeat the first 9 steps for each of the 10 folds.		Calculate error metrics and compare the results obtained by <i>Predictions_wCluster</i> and <i>Predictions</i>
11	Calculate error metrics by using the “ <i>Predictions</i> ” vector and the corresponding real load values		

**Fig. 3.** (a) Accuracy metrics versus standard deviation. (b) Real vs. predicted loads at hourly resolution and 24 h-ahead horizon by SVR for 5 representative households.

historical loads which shows the highest correlation with the target forecast load. For example for hourly resolution, the highest correlation is obtained at the 1st lag (previous

hour) followed by 24th (previous day) and multiples of 24 up to 168th lag (previous week). The remaining lags between the aforementioned shows less significant



**Fig. 4.** RMSE results of resolution versus horizon analysis on chosen households.

correlation values hence the predictor matrices for horizons; 6, 12 and 24 h includes very similar variables and consequently, obtained results are very similar. On the other hand, 1 h-ahead horizon predictor matrix includes the 1st lag in addition to the other highly correlated lags and hence the obtained results are better than other three horizons. It is important to mention that the predictor selection by autocorrelation is very much dependent on the household consumption profiles and hence highly correlated historical lags may significantly vary between different households.

Another important point to mention is the impact of resolution on forecast performance. The results show that errors increase at higher data resolutions. This outcome is not surprising as all the lower resolutions are obtained by averaging the five minute raw data. Hence, both the variance and magnitude of peak loads (spikes) are much greater for higher resolution data. This is shown in [Figure 5](#) below, where load profiles of the Household 11 are plotted for the four different resolutions between Jan 01 and Jan 15 2015.

As previously discussed, different households may exhibit different load profiles. Moreover, a particular household's daily electricity load profiles may significantly vary between different days. An example is given in [Figure 6](#) below, where eight daily load profiles are drawn for the Household 8 between January 01 and January 08 2015. It can be clearly seen that daily profiles have distinctive characteristics; load profile shape, mean, peak, time of peak, standard deviation are notably different for each profile.

Machine learning algorithms are expected to capture these daily profile changes with the help of predictor variables such as historical loads, calendar and time information. However, it is not as straight forward to find

out how well these changes can be captured with the given inputs, especially when working with machine learning models which are also known as “black box” models. Therefore, this gives the motivation to assist the models by pre grouping similar daily load profiles in the same clusters and further investigating forecast performance by applying a separate model for each cluster. Results of an example clustering assignment is shown in [Figure 7](#) for the Household 1. In this particular example, the optimum number of clusters were found to be four:

- cluster 1 represents days with minimal morning consumption but more significant after work hours usage;
- cluster 2 represents working days with morning and evening peaks where the prior is more emphasized;
- cluster 3 represents minimal household activities when household members are mostly outside the dwelling;
- cluster 4 represents days with a mid-day peak, which may correspond to weekends with a later start.

All cluster profiles show minimal activity during late night and early morning hours where household members are sleeping.

The improvement of performance obtained by the combining the clustering approach with respect to sole SMBM approach can be found in [Figure 8](#). Considering the superior performance of SVR in the previous analysis, it was selected as the preferred technique for this combined method.

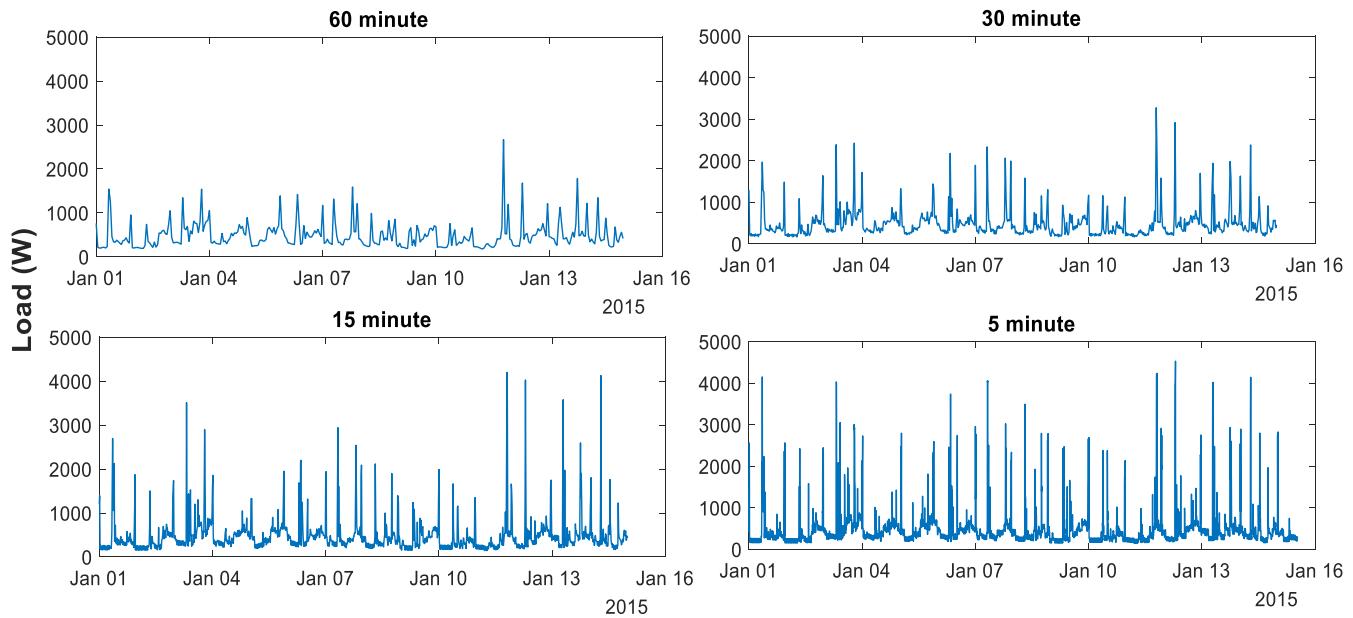
It can be seen that the pre-clustering step improves the forecast performance which is given as RMSE percentage of the average household load. The improvement is as high as 14% for HH11, and 9% for the average household stock.

We need to emphasize that the results obtained in the above figure assumes that each daily profile is modeled by the correct SMBM from the corresponding cluster. In particular, forecasts are carried for daily profiles which are already assigned to the correct cluster. In practice, a daily load profile which is to be forecasted, first needs to be assigned to a cluster and then to be modeled with the corresponding SMBM of the cluster. So far, we have used various calendar information in order to predict the clustering assignment and classify the daily profiles into the respective clusters; however, this approach has not given sufficient classification accuracy. This limitation is aimed to be improved in a future study.

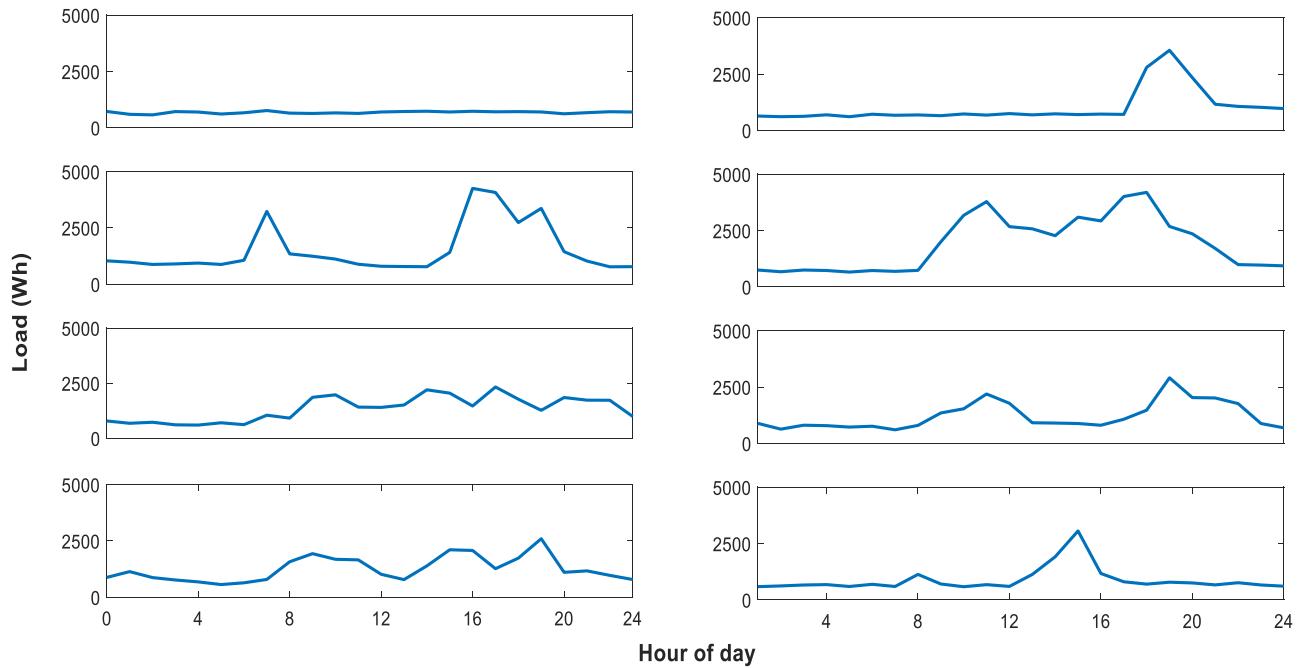
## 4 Conclusion

Individual household electricity loads pose great challenges for forecast models. So far, results in the literature lag behind of the results obtained for larger scale loads, which exhibit more predictable and stable load profiles. This is a sign that there can be improvement for predicting these highly volatile profiles.

Our analysis has shown that different households show different load characteristics, which affect the performance of the models. In particular, load volatility and standard deviation are two important parameters. This could indicate that different types of models can be used for



**Fig. 5.** Load profile of the Household 11 at four different data resolutions between Jan 01 and Jan 15 2015.



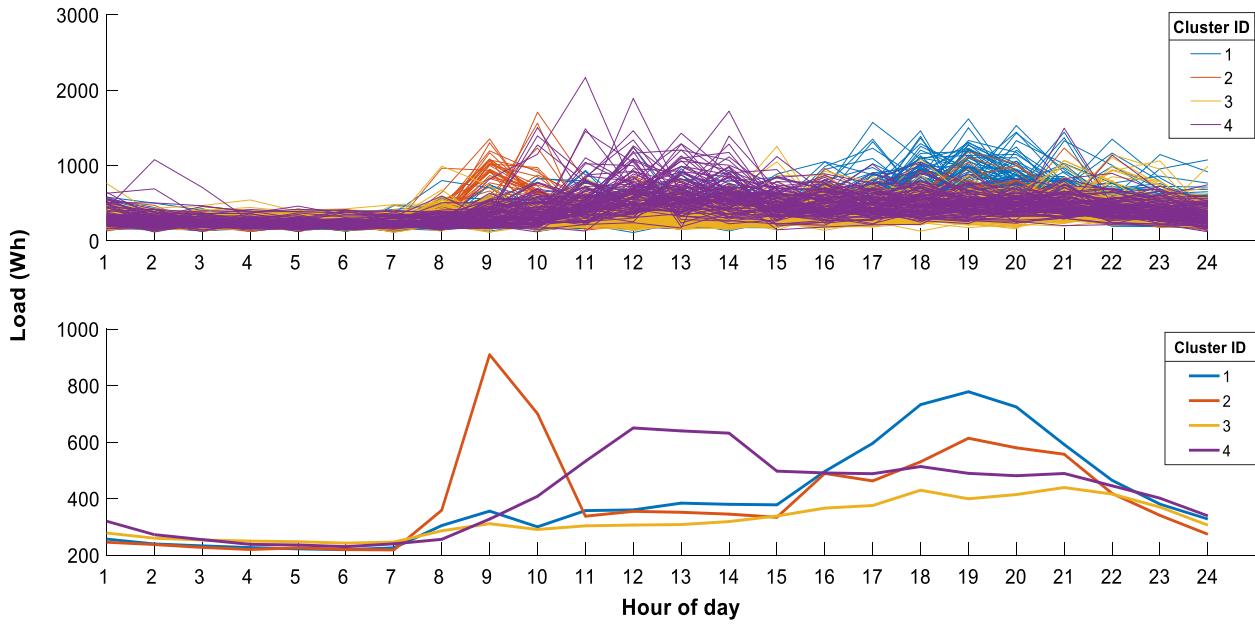
**Fig. 6.** Daily load profiles of the Household 8 between Jan 01 and Jan 08 2015.

different households according to their load profile characteristics. For example, a simple model could be used for a household with a more stable profile, whereas another household with higher volatility may require a more dynamic and responsive model.

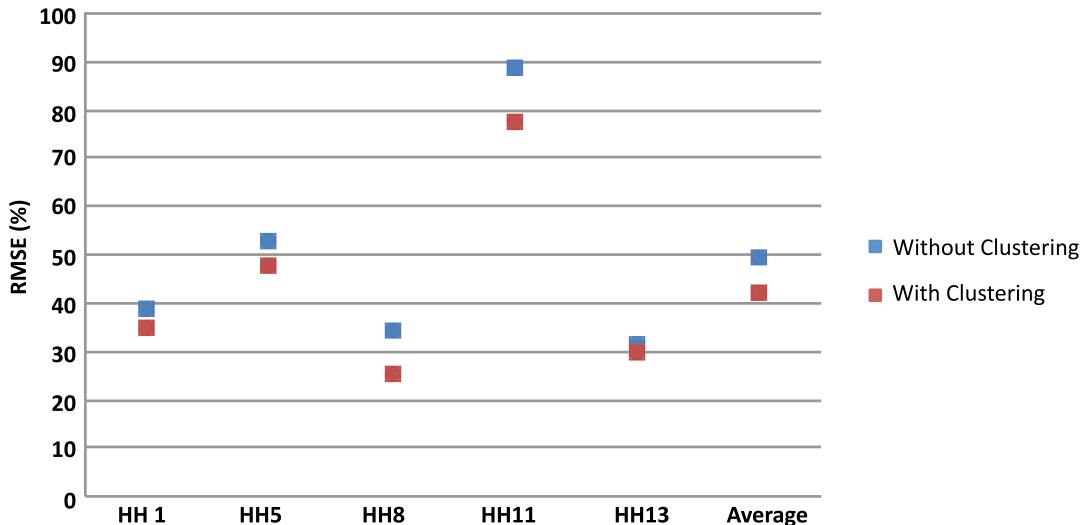
Results of data resolution versus forecast horizon has shown that the model performance stays relatively the same for horizons other than hour-ahead. This indicates that smart meter technologies at individual household scale may not gain much value from using a horizon shorter than

24 h unless it is hour ahead. These results are representative only for the studied household stock and must be confirmed for a broader range of households. So far, there has been limited amount of studies investigating the importance of load forecast resolution on smart grid technologies. Nonetheless, our analysis confirms that using higher data resolution results in higher errors.

We have also shown that assisting machine-learning algorithms through a pre-clustering step significantly improves the forecast performance. Although the use of



**Fig. 7.** Clustering assignment and cluster centroids for the Household 1.



**Fig. 8.** Improvement of RMSE with pre-clustering step.

a pre-clustering step increases the complexity of the process, the obtained improvements seem to justify the burden of this step. As previously mentioned, we will be looking for an affective way to assign forecasted daily profiles into the appropriate cluster in a future study.

This research was funded by the CRC for Low Carbon Living Ltd supported by the Cooperative Research Centers program, an Australian Government initiative. One of the authors (BY) particularly acknowledges scholarship funding provided by the Cooperative Research Centre for Low Carbon Living Ltd and the ongoing support of Solar Analytics Pty Ltd. Solar Analytics, an automated fault monitoring service for rooftop solar energy systems.

## References

1. U.S. Energy Information Administration, *International Energy Outlook 2016* (2016), [www.eia.gov/forecasts/ieo/pdf/0484\(2016\).pdf](http://www.eia.gov/forecasts/ieo/pdf/0484(2016).pdf).
2. L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: a review of modeling techniques, *Renew. Sustain. Energy Rev.* **13**, 1819–1835 (2009)
3. M. Kavcic, et al., A review of bottom-up building stock models for energy consumption in the residential sector, *Build. Environ.* **45**, 1683–1697 (2010), <http://dx.doi.org/10.1016/j.buildenv.2010.01.021>.

4. M. Lawless, M. Lawless, K. Tsuji, *Global Market for Smart Electricity Meters: Government Policies Driving Strong Growth* (2014), [https://www.usite.gov/publications/332/id-037smart\\_meters\\_final.pdf](https://www.usite.gov/publications/332/id-037smart_meters_final.pdf).
5. M. Rossi, D. Brunelli, Electricity demand forecasting of single residential units, in *2013 IEEE Workshop on Environmental, Energy and Structural Monitoring Systems, EESMS 2013 – Proceedings, (Dii)* (2013)
6. R.E. Edwards, J. New, L.E. Parker, Predicting future hourly residential electrical consumption: a machine learning case study, *Energy Build.* **49**, 591–603 (2012)
7. K.M. Ahmed, M.A. Al Amin, M.T. Rahman, Application of short term energy consumption forecasting for household energy management system, in *IEEE, Green Energy and Technology (ICGET)* (2015)
8. F. Rodrigues, C. Cardeira, J.M.F. Calado, The daily and hourly energy consumption and load forecasting using artificial neural network method: a case study using a set of 93 households in Portugal, *Energy Procedia* **62**, 220–229 (2014), <http://www.sciencedirect.com/science/article/pii/S1876610214034146>
9. M. Ghofrani, et al., Smart meter based short-term load forecasting for residential customers, in *NAPS 2011-43rd North American Power Symposium* (2011) pp. 13–17
10. K. Gajowniczek, T. Ząbkowski, Short term electricity forecasting using individual smart meter data. *Procedia Comput. Sci.* **35**, 589–597 (2014), <http://www.sciencedirect.com/science/article/pii/S1877050914011053>
11. R.P. Singh, P.X. Gao, D.J. Lizotte, On hourly home peak load prediction, in *2012 IEEE 3rd International Conference on Smart Grid Communications, SmartGridComm 2012* (2012) pp. 163–166.
12. M.A. Hall, *Correlation-based Feature Selection for Machine Learning* (The University of Waikato, 1999), <http://www.cs.waikato.ac.nz/~mhall/thesis.pdf>
13. MathWorks, MATLAB 2016b, <https://au.mathworks.com/products/matlab.html>
14. A. NG, *Neural Networks Learning in Machine Learning* (Stanford University, 2015), <https://www.coursera.org/learn/machine-learning/lecture/Wh6s3/putting-it-together>
15. A.S. Ahmad, et al., A review on applications of ANN and SVM for building electrical energy consumption forecasting, *Renew. Sustain. Energy Rev.* **33**, 102–109 (2014), <http://dx.doi.org/10.1016/j.rser.2014.01.069>
16. M.H. Beale, M.T. Hagan, H.B. Demuth, *Neural Network Toolbox User's Guide* (MathWorks, 2015), pp. 2–3, [http://medcontent.metapress.com/index/A65RM03P4874243N.pdf%5Cnhttp://onlinelibrary.wiley.com/doi/10.1002/0471221546.fmatter\\_ndsub/summary](http://medcontent.metapress.com/index/A65RM03P4874243N.pdf%5Cnhttp://onlinelibrary.wiley.com/doi/10.1002/0471221546.fmatter_ndsub/summary)
17. C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* **20**, 273–297 (1995)
18. G. James, et al., *An Introduction to Statistical Learning* (2006), <http://books.google.com/books?id=9tv0taI8l6YC>.
19. F. Javed, et al., Forecasting for demand response in smart grids: an analysis on use of anthropologic and structural data and short term multiple loads forecasting, *Appl. Energy* **96**, 150–160 (2012), <http://dx.doi.org/10.1016/j.apenergy.2012.02.027>.
20. A.J. Smola, B. Schölkopf, A tutorial on support vector regression, *Stat. Comput.* **14**, 199–222 (2004)
21. J.A.K. Suykens, et al., *Least squares support vector machines* (2002), Vol. **xiv**, p. 294
22. K. De Brabanter, et al., 2011. *LS-SVMlab Toolbox User's Guide*, <http://www.esat.kuleuven.be/sista/lssvmlab/>
23. K. De Brabanter, *Least Squares Support Vector Regression with Applications to Large-Scale Data: a Statistical Approach* (2011)
24. F. McLoughlin, *Characterising Domestic Electricity Demand for Customer Load Profile Segmentation* (2013), p. 274
25. L. Kaufman, P.J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis* (John Wiley & Sons, Inc., 1990)

**Cite this article as:** Baran Yildiz, Jose I. Bilbao, Jonathon Dore, Alistair B. Sproul, Short-term forecasting of individual household electricity loads with investigating impact of data resolution and forecast horizon, *Renew. Energy Environ. Sustain.* **3**, 3 (2018)