

Segmenting Residential Smart Meter Data for Short-Term Load Forecasting

Alexander Kell
School of Computing
Newcastle upon Tyne, UK
a.kell2@newcastle.ac.uk

A. Stephen McGough
School of Computing
Newcastle upon Tyne, UK
stephen.mcgough@ncl.ac.uk

Matthew Forshaw
School of Computing
Newcastle upon Tyne, UK
matthew.forshaw@ncl.ac.uk

ABSTRACT

In order to reliably generate electricity to meet the demands of the customer base, it is essential to match supply with demand. Short-term load forecasting is utilised in both real-time scheduling of electricity, and load-frequency control. This paper aims to improve the accuracy of load-forecasting by using machine learning techniques to predict 30 minutes ahead using smart meter data. We utilised the k -means clustering algorithm to cluster similar individual consumers and fit distinct models per cluster. Public holidays were taken into consideration for changing customer behaviour, as was periodicity of the day, week and year. We evaluated a number of approaches for predicting future energy demands including; Random Forests, Neural Networks, Long Short-Term Memory Neural Networks and Support Vector Regression models. We found that Random Forests performed best at each clustering level, and that clustering similar consumers and aggregating their predictions outperformed a single model in each case. These findings suggest that clustering smart meter data prior to forecasting is an important step in improving accuracy when using machine learning techniques.

KEYWORDS

Advanced metering infrastructure (AMI), k -means clustering, neural networks, random forest, support vector regression, smart meter, load profiles, electrical power load forecasting, short-term electricity forecasting

1 INTRODUCTION

The energy markets have undergone significant changes in recent years. The liberalisation of the energy industry, technological advancements and policy changes have had numerous effects [34]. These include a rise in both competition and data [6, 28].

Accurate load forecasting is essential for control and planning of electricity generation in electrical grids as supply must meet demand [20]. Accurate estimates of demand are required so that the correct amount of electricity is purchased on the wholesale market [8]. Failure to accurately forecast electricity demand can lead to financial loss or system-wide blackouts [13].

The introduction of smart meters in many countries (USA, Europe and South Korea) has led to an influx of high granularity electricity consumption data that can be used for load forecasting [7]. 800 million smart meters are projected to be installed worldwide by 2020 [30].

This paper explores short-term load-forecasting at an interval of 30 minutes ahead and clusters similar users based on their electricity consumption. A variety of different forecasting techniques were

evaluated such as Random Forests [32], Long Short-Term Memory Neural Networks (LSTM) [14], Artificial Neural Networks [33] (ANN) and Support Vector Regression (SVR) [9].

Random Forests are an ensemble-based learning method for classification and regression, and are made up of many decision trees. LSTMs are recurrent Neural Networks which remember values over arbitrary time intervals. Multilayer Perceptrons are a popular type of neural network which consist of a minimum of three layers and can be used to make non-linear predictions. SVRs are supervised learning models which analyse data for regression analysis.

To improve forecasting results, k -means clustering of smart meter data was evaluated. An average 24-hour electricity load profile per customer was calculated, and the result used for clustering. The clustered sub-system is then aggregated and separate models trained on these aggregates. The yearly, weekly and daily periodicity of electricity load is accounted for by feature vectors. Once forecasts for each cluster are made using the individual models, the results are aggregated for the final predictions. These predictions are compared to the actual results and the accuracy measured using mean absolute percentage error (MAPE) and mean absolute scaled error (MASE).

This paper provides researchers and utilities with methods to maximise forecasting accuracy through the selection of machine learning and clustering algorithms.

This paper is structured as follows. In Section 2 we explore related work of load forecasting. The experiments and their evaluation are discussed in Section 3. The results are discussed in Section 4. In Section 5 we conclude and consider future directions for this work.

2 RELATED WORK

The forecasting of aggregated and clustered electricity demand has been the focus of a considerable amount of research in recent years. The research can generally be classified into two classes, Artificial Intelligence (AI) techniques [19, 22, 25] and classical time series approaches [16, 23, 24]. For the purposes of our paper we have reviewed artificial intelligence techniques. Please refer to appendix A to explore the literature related to classical time series approaches.

Singh *et al.* produced a review of load forecasting techniques and methodologies and reported that hybrid methods, which combine two or more different techniques, are gaining traction, as well as soft computing approaches (AI) such as genetic algorithms [27].

2.1 Artificial Intelligence Techniques

Dillon *et al.* presented a Neural Network for short-term load forecasting. Their Neural Network consisted of three-layers and used adaptive learning for training [8]. They proposed the use of weather

information to augment their electricity load data. They found better results with the Adaptive Neural Network than with a linear model, or Non-Adaptive Neural Network. In contrast to Dillon our paper focuses on a Non-Adaptive Neural Network and does not take into account weather information.

Chen *et al.* used an Artificial Neural Network to predict electricity demand of three substations in Taiwan. They integrated temperature data into the model, and showed a higher degree of accuracy when forecasting demand in residential and commercial substations as opposed to industrial. This was due to the ability to model the high usage of air-conditioners in residential and commercial substations using temperature data [5]. In contrast to the work by Chen *et al.*, we focus on client-side prediction using smart meter data. We were, therefore, able to cluster the data based on load profile, as opposed to grouping based on geographical location.

2.2 Clustering

Multiple techniques have been proposed for the clustering of electricity load data prior to forecasting. Shu *et al.* and Nagi *et al.* propose a hybrid approach in which self-organizing maps are used to cluster the data, and Support Vector Regression is used for prediction [22, 26]. This technique proved robust for different data types. Shu showed that this hybrid approach out-performed a single SVR technique, whilst Nagi showed superior results to a traditional ANN system. In contrast to both Nagi *et al.* and Shu *et al.* our paper utilises *k*-means as the clustering algorithm.

Wijaya *et al.* demonstrated that implementing a certain number of clusters improved load-forecasting accuracy [35]. However, a study by Ilić *et al.*, showed that increasing the number of clusters did not improve accuracy [17].

3 METHODOLOGY

The work in this paper was run on a MacBook Pro with a quad-core 3.1GHz Intel Core i7 processor with 16 GB 1867 MHz DDR3 of RAM and a 500GB solid state drive (SSD).

3.1 Data Collection

Smart meter data obtained from the Irish Social Science Data Archive (ISSDA) was used in this study [11]. The Commission for Energy Regulation released a public dataset of anonymised smart meter data from the "Electricity Smart Metering Customer Behaviour Trials" [1]. This dataset is made up of over 5000 Irish homes and businesses and is sampled at 30-minute intervals.

The data was recorded between the 14th July 2009 and 31st December 2010. For the purposes of cross-validation this data was split into a training, validation, and testing set. The training set consisted of the first 9 months of data and used to train the models, the validation set consisted of the following 2 months of data and used to tune the hyperparameters, and the test set included the remaining 6 months and used for measuring error. These splits were chosen to balance the training data with the test data and give the models a chance to learn the periodicity inherent in a one year period. Due to the long training times for these algorithms, we worked with a sample of 709 individual Irish homes. However, due to the infrequent requirement to train these models, we believe our technique would be suited for the real life application.

Figure 1 displays four residential customer daily load profiles. It can be seen that whilst Customer 1 and Customer 2 have similar load profiles, Customer 3 and Customer 4 have significantly different load profiles. This demonstrates that, whilst electricity consumption changes per-person, it is possible to cluster similar customers by their load profiles.

3.2 Clustering

We propose that clustering similar customer load profiles and aggregating each cluster's electricity consumption improves the accuracy of the models. This is due to the fact that aggregated clustered loads decrease the stochasticity of the demand of the system, and therefore increase the predictive ability of the models.

The Euclidean distance and wavelet metrics were evaluated using hierarchical clustering [21]. However, *k*-means demonstrated to be the most robust and best-performing algorithm, and thus was chosen for use in this paper [12].

To cluster the data, each user's nine month electricity consumption in the training set was reduced to a single averaged 24 hour load profile (48 data points per customer). This was achieved by taking the average electricity consumed for each half hourly point of the day (eg. taking the mean of every 12-12:30pm point in the training set). We did not take into account the difference between weekend and weekdays for clustering. The data was then scaled so that different sized households, but with similar usage profiles were clustered together.

To select the optimum number of clusters (*k*) the test set was used for cross-validation. This allowed us to compare the results of each of the models and select *k* with the highest accuracy exhibited by mean absolute percentage error (MAPE) and mean absolute scaled error (MASE). In this paper *k* was varied between 1 and 7, which was chosen due to the fact that the error did not vary greatly past seven clusters. Multiple models were fit per cluster and used to predict the testing data.

To overcome local minima the *k*-means algorithm was run 1000 times and the most accurate partition chosen [18].

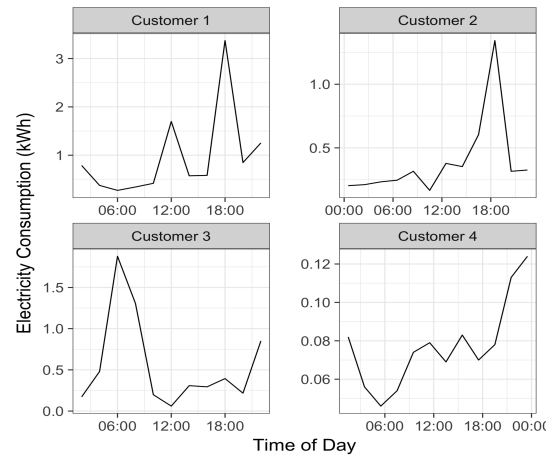


Figure 1: Figure showing daily load profiles for four different customers on the 22nd July 2009.

3.3 Aggregating Demand

Once each customer is assigned to their respective cluster, the total electricity consumed by each cluster is calculated. This is achieved by summing the electricity consumed at every time interval per cluster. This creates a partial system load. A model is trained on each of the different partial system loads, and the resultant forecasts are aggregated to generate the total load forecast. This forecast is then evaluated by calculating the MAPE and MASE for each model.

3.4 Feature Selection

3.4.1 Calendar Attributes. Due to the daily, weekly and annual periodicity of the electricity consumption, daily calendar attributes were deemed important to accurately model the problem. The calendar attributes included are: hour of the day, day of the month, day of the week, month, and public holidays.

Public holidays were used as features for the model due to the change in electricity consumption exhibited on these days.

We evaluate the increase in performance due to the modelling of calendar attributes in the results section.

3.4.2 Time Series Data. The time-series element was modelled using lagged data inputs. This was achieved using the previous 3 hours, the equivalent 3 hours from the previous day, and the equivalent 3 hours from the previous week.

Long Short-Term Memory neural networks remember values over arbitrary time intervals. And thus can remember short-term memory over a long period of time, for this reason, 5 lagged inputs of the previous two and a half hours were used as features to the Long Short-Term Memory network.

3.4.3 Data Representation. Numerical representations were used for the lagged data input. A single binary was used for public holidays. One hot encoding is a method which allows categorical variables to be distinguished from ordinal data. One hot encoding was used to encode day of the week and month of the year. Table 1 displays the input data for SVR, Neural Network and Random Forest.

3.5 Experiments

This section explores the experiments made to design the models. Cross-validation was used on the validation set of each of the models to tune the hyperparameters. Each of the models were then created five times per cluster to explore the variance of the results.

3.5.1 Support Vector Regression. The validation set was used to tune the hyperparameters and select the kernel of the SVR model.

The kernels compared were polynomial, radial basis function (RBF) and the linear kernel [4, 31]. They were chosen due to their popularity, support and speed of computation.

The parameter values are shown in Table 2. The linear kernel produced the best results, and therefore chosen as the final model.

3.5.2 Random Forest. The number of variables sampled as candidates at each split was tuned using the validation set.

The optimum number for this was 23. It is proposed that the value 23 was found to be optimum due to the 21 lagged inputs, which are crucial to learn the underlying nature of the time series.

Table 1: List of Input Data for Models

Input	Variable	Detail description
1	Hour	Single numeric input representing hour of the day
2	Day of month	Single numeric input representing day of the month
3-9	Day of week	Seven binary digits representing calendar information regarding day of the week
10-21	Month	Twelve binary digits representing calendar information regarding month
22-42	Lagged inputs	Twenty one numeric inputs representing lagged inputs of previous 3 hours, previous 3 hours of previous day including hour to be predicted, and previous 3 hours of previous week including hour to be predicted
43	Holiday	Single binary digit representing whether the day was a public holiday

Table 2: Prediction Accuracy Based on Type of Kernel

Kernel Type	Kernel Parameters	RMSE
Linear	No values	0.02102779
RBF	$C=2, \gamma = 0.016$	0.02444950
Polynomial	$C=2, d = 2, r = 2$	0.03145719

3.5.3 Artificial Neural Network. The first step when creating an Artificial Neural Network is to design the architecture. In our case, the number of input neurons is set to 43 (see Table 1). Only one output neuron is required, due to the fact that we are only forecasting one step (30 minutes) ahead.

To design the number of hidden layers the Levenberg-Marquardt technique was used. An optimal architecture with three hidden layers was obtained. The first layer contained two neurons, the second contained five, and the third contained four.

3.5.4 LSTM. The Levenberg-Marquardt techniques was once again used to select number of layers and number of memory units. Using this technique, the optimum number of layers was found to be 2 with 50 memory units each.

4 RESULTS

To determine the optimal number of clusters a range of values for k were explored, thus, k was varied between one and seven. 28 forecasting models were therefore constructed per type of model. The models were fit five times to explore the variation in the output. The model accuracy was evaluated using both MAPE and MASE.

The results are shown in figure 2. These show that clustering similar users improves accuracy. The optimum value for k for every model was shown to be four. After this, the accuracy diminishes slightly. The error bars shown in Figure 2 display a slight variance

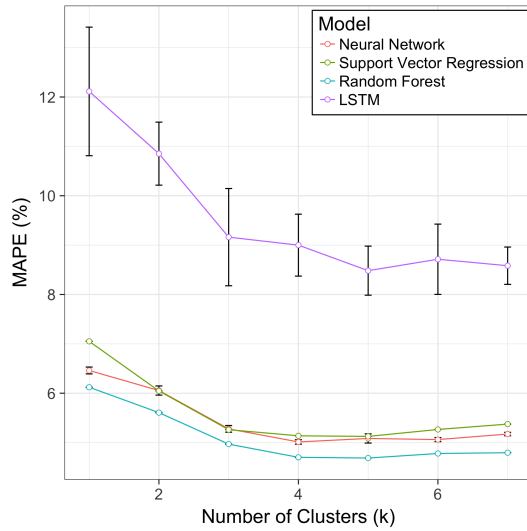


Figure 2: Comparison of accuracy of models forecasting electricity with varying number of clusters.

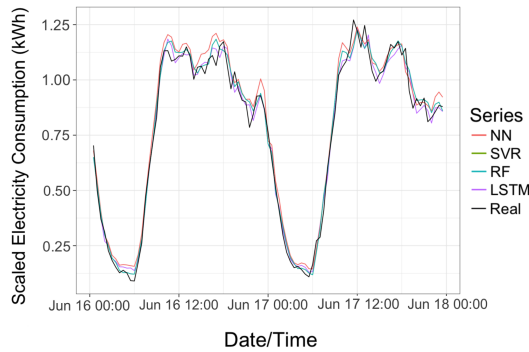


Figure 3: Real electricity consumption versus predicted electricity consumption between June 16th and June 18th 2010.

in MAPE for SVRs, ANNs and Random Forests. However, the MAPE of the LSTMs seem to vary by up to 11%.

The MASE metric also demonstrated that four clusters were optimal for SVR, Neural Networks and Random Forests.

The impact of using calendar attributes improves prediction accuracy by 6% for neural networks, 4% for Random Forests and 1% for SVR. For these results please see figure 6 in appendix D.

It is proposed that the optimum value for k cluster centres was four due to the distinct patterns observed in each of the clusters. At $k = 5$ one of the distinct clusters is split, and leads to an increase in stochasticity. At $k = 3$ the stochasticity is also increased by the aggregation of load profiles which are dissimilar.

However, the optimum number of clusters will vary for different datasets. Differing geographies may display varying usage characteristics due to culture, weather or social norms.

The results demonstrate that SVR, Random Forests and the ANN have similar accuracy, and adequately predict electricity consumption. The LSTM shows a similar pattern in increasing accuracy with number of clusters, but performs worse than the other models. The Random Forest seems to outperform each of the other models. This may be due to the internal operation of the Random Forest which undertakes its own cross-validation using out-of-bag samples.

Figure 3 displays actual electricity consumption versus predicted results. It shows that the LSTM model predicts a similar value in the next time step as the previous time step, which would explain its inferior results to the other models.

The training times were tested by timing how long the models took to fit for four clusters. The Support Vector Regression took less time than all of the other methods, whereas the LSTM took the longest. The Support Vector Regression model required 3 minutes and 18 seconds to run. The Random Forest required 14 minutes and 58 seconds. The Artificial Neural Network required 17 minutes and 48 seconds, whilst the LSTM took 21 minutes 11 seconds to run.

The time to make a single prediction was recorded at sub microseconds and therefore deemed negligible for our use-case.

5 CONCLUSION

The availability of data produced by smart meters enables network operators to gain greater insights into their customer behaviour and electricity usage. We demonstrated that implementing the k -means clustering algorithm to group similar customers improved the accuracy of the models tested. Distinct models were trained for each cluster and the individual forecasts aggregated for the total aggregated forecast. It was found that Random Forests outperformed all other models at each level of clustering. The optimum number of clusters was found to be four. Whilst the dataset used focused on residential data it is expected that applying a similar clustering technique on commercial properties would have a similar effect.

It is considered that monthly retraining of the models would be required to ensure continued accuracy. However, this is not expected to be a problem due to the short time time required for model training. Once trained, the prediction times are negligible.

In future work, we will look into the features that best aid in the forecasting of electricity consumption, try a wider variety of models in an ensemble manner and try different clustering techniques such as self-organizing maps (SOM) to obtain better accuracy results. We will also compare a greater variety of forecasting metrics.

ACKNOWLEDGMENTS

This work was supported by the Engineering and Physical Sciences Research Council, Centre for Doctoral Training in Cloud Computing for Big Data [grant number EP/L015358/1].

REFERENCES

- [1] 2011. The Irish Electricity Smart Metering Customer Behaviour Trials. (2011). <https://setis.ec.europa.eu/energy-research/project/irish-electricity-smart-metering-customer-behaviour-trials> [Online; accessed 21-January-2018].
- [2] Hirotugu Akaike. 1974. A New Look at the Statistical Model Identification. *IEEE Trans. Automat. Control* 19, 6 (1974), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- [3] M.S. Al-Musaylh, R.C. Deo, J.F. Adamowski, and Y. Li. 2018. Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated

- demand data in Queensland, Australia. *Advanced Engineering Informatics* 35, November 2017 (2018), 1–16. <https://doi.org/10.1016/j.aei.2017.11.002>
- [4] Yin-Wen Chang, Cho-Jui Hsieh, Kai-Wei Chang, Michael Ringgaard, and Chih-Jen Lin. 2010. Training and Testing Low-degree Polynomial Data Mappings via Linear SVM. *Journal of Machine Learning Research* 11 (2010), 1471–1490. http://www.csie.ntu.edu.tw/~cjlin/papers/lowpoly_journal.pdf
 - [5] C.S. Chen, Y.M. Tzeng, and J.C. Hwang. 1996. The application of artificial neural networks to substation load forecasting. *Electric Power Systems Research* 38, 1996 (1996), 153–160. [https://doi.org/10.1016/S0378-7796\(96\)01077-2](https://doi.org/10.1016/S0378-7796(96)01077-2)
 - [6] Cédric Clastres. 2011. Smart grids: Another step towards competition, energy security and climate change objectives. *Energy Policy* 39, 9 (2011), 5399–5408. <https://doi.org/10.1016/j.enpol.2011.05.024>
 - [7] Soma Shekara Sreenadh Reddy Depuru, Lingfeng Wang, and Vijay Devabhaktuni. 2011. Smart meters for power grid: Challenges, issues, advantages and status. *Renewable and Sustainable Energy Reviews* 15, 6 (2011), 2736–2742. <https://doi.org/10.1016/j.rser.2011.02.039>
 - [8] T S Dillon, S Sestito, and S Leung. 1991. Short term load forecasting using an adaptive neural network. *Electrical Power & Energy Systems* (1991), 186–192.
 - [9] Harris Drucker, Chris J C Burges, Linda Kaufman, Alex Smola, and Vladimir Vapnik. 1997. Support vector regression machines. *Advances in Neural Information Processing Systems* 1 (1997), 155–161. <https://doi.org/10.1.1.10.4845>
 - [10] Abdollah Kavousi Fard and Mohammad-Reza Akbari-Zadeh. 2014. A hybrid method based on wavelet, ANN and ARIMA model for short-term load forecasting. *Journal of Experimental & Theoretical Artificial Intelligence* 26, 2 (2014), 167–182. <https://doi.org/10.1080/0952813X.2013.813976>
 - [11] Commission for Energy Regulation (CER). 2012. *CER Smart Metering Project - Electricity Customer Behaviour Trial, 2009-2010* (2012). www.ucd.ie/issda/CER-electricity
 - [12] E. Forgy. 1965. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics* 21 (1965), 768–780.
 - [13] Paul Hines, Jay Apt, and Sarosh Talukdar. 2008. Trends in the history of large blackouts in the United States. *IEEE Power and Energy Society 2008 General Meeting: Conversion and Delivery of Electrical Energy in the 21st Century, PES* 15213 (2008), 1–8. <https://doi.org/10.1109/PES.2008.4596715>
 - [14] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (1997), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735> arXiv:https://doi.org/10.1162/neco.1997.9.8.1735
 - [15] Charles C. Holt. 2004. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting* 20, 1 (2004), 5–10. <https://doi.org/10.1016/j.ijforecast.2003.09.015>
 - [16] Shyh-jier Huang, Senior Member, and Kuang-rong Shih. 2003. Short-Term Load Forecasting Via ARMA Model Identification Including Non-Gaussian. 18, 2 (2003), 673–679.
 - [17] Dejan Ilić, Per Goncalves da Silva, Stamatis Karnouskos, and Malte Jacobi. 2013. Impact assessment of smart meter grouping on the accuracy of forecasting algorithms. *Proceedings of the 28th Annual ACM Symposium on Applied Computing - SAC '13* (2013), 673. <https://doi.org/10.1145/2480362.2480491>
 - [18] Anil K. Jain. 2010. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* 31, 8 (2010), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011> arXiv:cond-mat/0402594v3
 - [19] Kwang-ho Kim, Hyoung-sun Youn, Student Member, and Yong-cheol Kang. 2000. Short-term load forecasting for special days in anomalous load conditions using neural networks. *IEEE Transactions on Power Systems* 15, 2 (2000), 559–565. <https://doi.org/10.1109/59.867141>
 - [20] C.N. Lu, H. T. Wu, and S. Vemuri. 1993. Neural network based short term load forecasting. *IEEE Transactions on Power Systems* 8, 1 (1993), 336–342. <https://doi.org/10.1080/02533839.1995.9677697>
 - [21] K. V. Mardia, J. T. Kent, and J. M. Bibby. 1979. *Multivariate analysis*. Academic Press London ; New York. xv, 521 p. : pages. <http://www.loc.gov/catdir/toc/els031/79040922.html>
 - [22] J Nagi, S. K Yap, S K Tiong, and S K Ahmed. 2008. Electrical Power Load Forecasting using Hybrid Self-Organizing Maps and Support Vector Machines. *The 2nd International Power Engineering optimization Conference (PEOCO)* June (2008), 51 – 56.
 - [23] J Nazarko, A Jurczuk, and W Zalewski. 2005. ARIMA models in load modelling with clustering approach. In *2005 IEEE Russia Power Tech, PowerTech*. 1–6. <https://doi.org/10.1109/PTC.2005.4524719>
 - [24] Hung Nguyen and Christian K. Hansen. 2017. Short-term electricity load forecasting with Time Series Analysis. *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)* (2017), 214–221. <https://doi.org/10.1109/ICPHM.2017.7998331>
 - [25] Franklin L Quilumba, Wei-jen Lee, Heng Huang, David Y Wang, Senior Member, and Robert L Szabados. 2014. Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities. (2014), 1–8.
 - [26] Fan Shu and Chen Luonan. 2006. Short-term load forecasting based on an adaptive hybrid method. *Power Systems, IEEE Transactions on* 21, 1 (2006), 392–401. <https://doi.org/10.1109/TPWRS.2005.860944>
 - [27] Arunesh Kumar Singh, Ibraheem, S. Khatoon, Md Muazzam, and D. K. Chaturvedi. 2012. Load forecasting techniques and methodologies: A review. *ICPCES 2012 - 2012 2nd International Conference on Power, Control and Embedded Systems* (2012). <https://doi.org/10.1109/ICPCES.2012.6508132>
 - [28] Fereidoon P. Sioshansi. 2009. *Competitive electricity markets: design, implementation, performance*. Elsevier.
 - [29] James W. Taylor. 2008. An evaluation of methods for very short-term load forecasting using minute-by-minute British data. *International Journal of Forecasting* 24, 4 (2008), 645–658. <https://doi.org/10.1016/j.ijforecast.2008.07.007>
 - [30] Telefónica. 2014. The Smart Meter Revolution. Towards a smarter future. (2014), 11.
 - [31] Sergios Theodoridis. 2009. *Pattern recognition*. Academic Press, Burlington, MA London.
 - [32] Tin Kam Ho. [n. d.]. Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition 1* ([n. d.]), 278–282. <https://doi.org/10.1109/ICDAR.1995.598994>
 - [33] Dr. Günther Palm Dr. Ad Aertsen V. Braitenberg, G. Palm. 1986. *Brain Theory: Proceedings of the First Trieste Meeting on Brain Theory, October 1âŠ4, 1984* (1 ed.). Springer-Verlag Berlin Heidelberg. <http://gen.lib.rus.ec/book/index.php?md5=0c346f052f1aae1ea130393eaa3e43b>
 - [34] Joaquim L. Viegas, Susana M. Vieira, R. Melicio, V. M.F. Mendes, and João M.C. Sousa. 2016. Classification of new electricity customers based on surveys and smart metering data. *Energy* 107 (2016), 804–817. <https://doi.org/10.1016/j.energy.2016.04.065>
 - [35] Tri Kurniawan Wijaya, Matteo Vasirani, Samuel Humeau, and Karl Aberer. 2010. Residential Electricity Load Forecasting : Evaluation of Individual and Aggregate Forecasts. (2010), 1–8.

Appendices

A TIME SERIES RELATED WORK

This section explores the work done in the literature for a classical time series approach.

Al-Musaylh *et al.* proposed the use of Support Vector Regression (SVR), an autoregressive integrated moving average (ARIMA) model and a multivariate adaptive regression spline (MARS) in their short term electricity demand forecasting system [3]. They found that for a half, and one-hour forecasting horizons, that the MARS model outperformed both the ARIMA and SVR.

Taylor evaluates different statistical methods including ARIMA, an adaptation of Holt-Winters’ exponential smoothing [15], and an exponential smoothing method which focuses on the evolution of the intra-day cycle [29]. He found that the double seasonal adaptation of the Holt-Winters’ exponential smoothing method was the most accurate method for short lead times between 10 and 30 minutes.

In contrast to Taylor, Fard *et al.* proposed a novel hybrid forecasting method based on both artificial intelligence and classical time series approaches. They utilised the wavelet transform, ARIMA and ANNs for short term load forecasting [10]. The ARIMA model is created by finding the appropriate order using the Akaike information criterion [2]. The ARIMA model models the linear component of the load time series, and the residuals contain the non-linear components. These residuals are then decomposed by the discrete wavelet transform into its sub-frequencies. ANNs are then applied to these sub-frequencies and the outputs of both the ANN and ARIMA models are summed to make the final prediction. They found that this hybrid technique outperformed traditional methods. Our paper, however, does not integrate artificial intelligence and classical time series techniques.

B METHEDOLOGY

B.1 Levenberg-Marquardt Technique

The Levenberg-Marquardt is a technique suitable for training medium-sized Artificial Neural Networks with a low mean-squared error.

The fundamental rule is to select the minimum number of neurons in the hidden layer so as to capture the complexity of the model, but not too many as to introduce over-fitting, which results in a loss in generalization of the algorithm.

The method begins by choosing a small number of neurons and gradually increasing the number each time the model is trained and the forecast error obtained. The forecast error is monitored until an optimum value is found, to which no further improvement is noted. Once the optimum number of neurons in the layer is obtained an additional layer is added, and the same technique is used.

C EXPERIMENTS

C.1 Parameter Tuning of Random Forest

Figure 4 displays the tuning of the random forest hyperparameter "number of variables randomly sampled as candidates at each split."

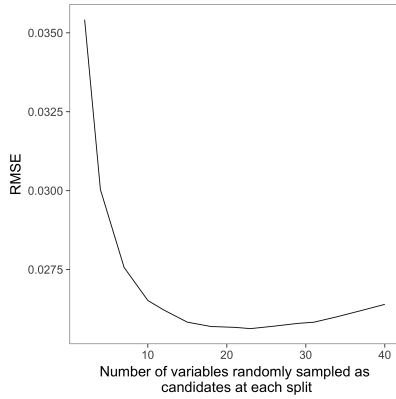


Figure 4: RMSE vs Number of variables randomly sampled as candidates at each split in the Random Forest model.

D RESULTS

Figure 5 displays the average load profile for each of the cluster centres. Here, distinct load profiles can be seen in each cluster. High night time use is exhibited in Cluster 1, a typical residential load profile in Cluster 2, a spread of usage in Cluster 3, and high daytime usage in Cluster 4.

Figure 6 displays the improvement in accuracy attained by including calendar attributes into the training of the model. It shows a clear improvement in training accuracy for all three models when calendar attributes are taken into account.

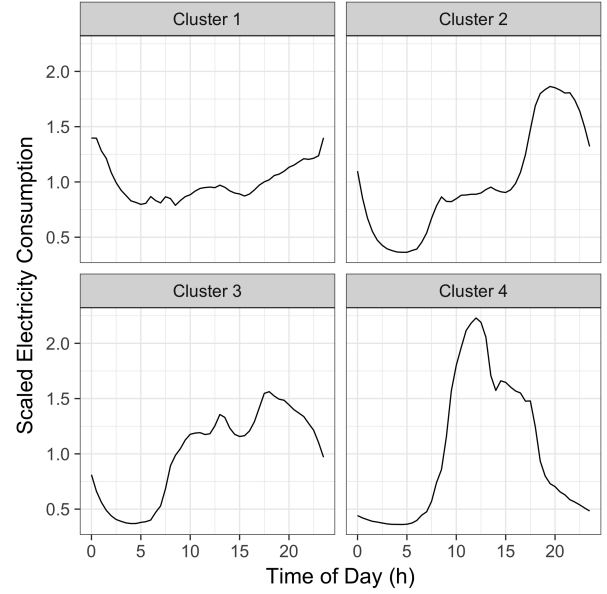


Figure 5: Average load profile for each cluster.

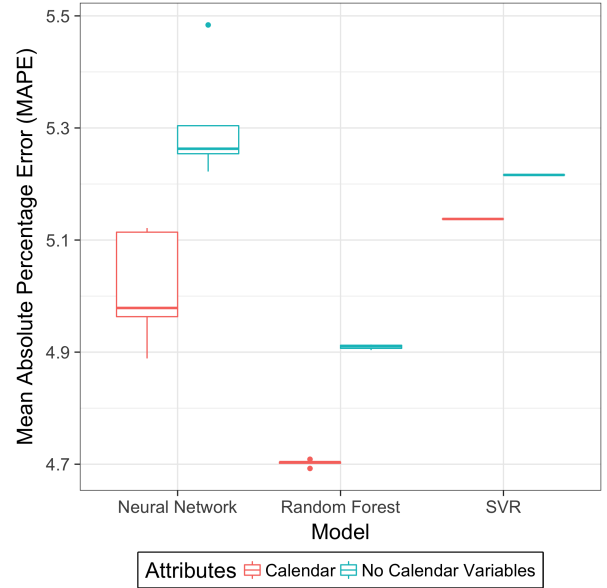


Figure 6: Comparison of accuracy of models with and without calendar attributes.