Energy Forecasting with Building Characteristics Analysis

Shuang Dai
Department of Mathematical Sciences
University of Essex
Colchester, UK
sd19628@essex.ac.uk

Fanlin Meng
Department of Mathematical Sciences
University of Essex
Colchester, UK
fanlin.meng@essex.ac.uk

***Abstract*—With the installation of smart meters, high resolution building-level energy consumption data become increasingly accessible, which not only provides more accurate data for energy forecasting at the aggregated level but also enables data-driven energy forecasting for individual buildings. On the one hand, individual buildings exhibit high randomness, making the forecasting problem at the building-level more challenging. On the other hand, buildings usually have their own characteristics, therefore such valuable information needs to be considered in the forecast models at the aggregation level. In this paper we investigate how unique characteristics of buildings could affect the performance of forecasting models and aim to identify defining patterns of buildings. The usefulness of the proposed approach is demonstrated using data from three real-world buildings.**

***Keywords*—energy forecasting, building energy management, building characteristics, machine learning**

I. Introduction

Buildings account for up to 40% of the global energy consumption [1] as an indispensable role in day-to-day life. Building energy forecasting plays an important role in improving the flexibility and reliability of energy system operations [2]. With the installation of smart meters, building-level high-resolution energy consumption data become available, which provides more opportunities for energy forecasting applications at both the building-level (e.g. peer-to-peer energy trading) and aggregation level (e.g. distribution network operations).

Energy forecasting is often classified according to the forecast lead time, which can be roughly divided into long-term load forecasting, medium-term load forecasting and short-term load forecasting [3]. For practical operation and control applications, short-term load forecasting (e.g. one hour ahead or day-ahead forecast) is usually considered. In general, building energy forecasting models can be categorized into building physics-based approach and data-driven approach. The former requires detailed physical information (such as building material types and ventilation system parameters which are often difficult to obtain) to calculate the thermal dynamics and energy behaviours of individual buildings [4]. Moreover, the physics-based approach often leads to unsatisfactory forecasting results due to inevitable errors in the process of information collection [5]. In contrast, the data-driven approach such as machine learning-based forecast models, which uses the real-world data to find the hidden relationship between independent variables and response variable, has been widely emphasized in building energy usage forecasting during the past two decades for its superior performance [6].

In [7], a new integration model based on artificial neural network (ANN) was proposed, which performed better than the extension of autoregressive integrated moving average (ARIMA) model on day-ahead load forecasting and recursive least square on one-hour-ahead forecasting, respectively. In [8], a comparative analysis for short-term load forecasting for a building in Belgium was carried out where random forest (RF) was compared with multiple linear regression, support vector regression (SVR) and gradient boost machines. It demonstrated that RF has the best performance among the considered models. Similar conclusions were reached in [9] where a short-term load forecasting was investigated.

In recent years, energy forecasting at building-level becomes increasingly important with emerging applications in the demand response, microgrids and peer-to-peer energy trading. However, single building usually exhibits high randomness, thus making the forecasting more difficult. In addition, since each building has its own unique characteristics, it is difficult to find one forecast model that works for all different types of buildings. On the other hand, aggregation-level energy forecasting usually plays a significant role in applications such as community energy management and grid operations. However, most existing studies simply aggregate buildings directly without considering unique characteristics of buildings. Therefore, some useful information of individual buildings might not be well preserved in the data aggregation process.

Motivated by the above analysis, in this paper we investigate the relationship between building characteristics and forecasting performance considering three popular energy forecasting models at both the building-level and aggregation-level. In particular, we consider the one-hour-ahead load forecasting with SVR, RF and ANN on three real-world buildings. The forecast results are then compared and analyzed to investigate the impact of building characteristics on the performance of forecast models. In addition, we quantify the importance of predictors with the aim to identify key driving factors influencing the electricity consumption patterns. The importance of building characteristics in the aggregation-level forecasting

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

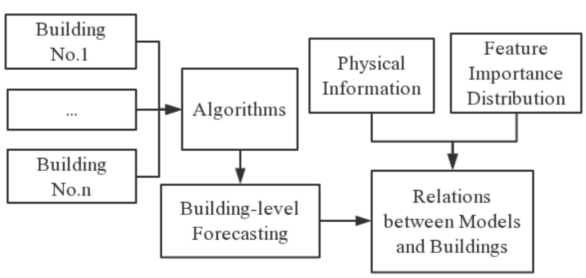
is also discussed through an comparative analysis.

The remainder of this paper is organised as follows. Section II describes the methodology, which includes the proposed conceptual framework for energy forecasting with building characteristics analysis, the considered forecast models and evaluation metrics. Section III presents the dataset, data pre-processing and the pipeline of the experimental design. The results and comparative analysis are given in Section IV. Finally, Section V concludes the paper and points out potential future research directions.

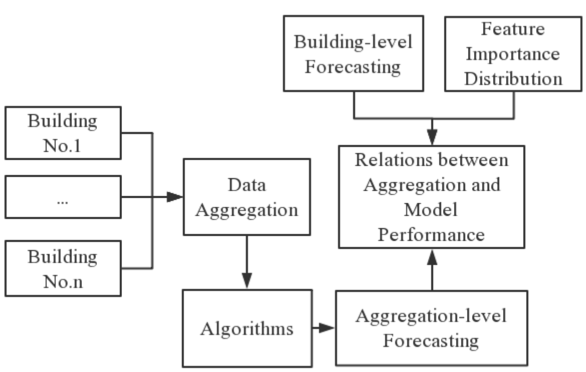
II. Methodology

A. General Framework

Each building has its own unique electricity usage pattern due to the distinct physical information such as different types of dwelling, occupants and heating types. Therefore, for different types of consumption patterns, it is desirable to choose appropriate forecast algorithms. This paper aims to explore the relationship between performance of forecast algorithms and different consumption patterns at both building-level and aggregation-level. The overall framework is given in Fig. 1.



(a) Building-level forecasting



(b) Aggregation-level forecasting

Fig. 1. The general framework.

At the building-level forecasting stage (Fig. 1 (a)), we explore the potential relationship between algorithms and buildings by comparing the performance of forecast models for different buildings. Considering the unique electricity usage patterns and physical characteristics of each building, we analyze the performance of each forecast model and recommend the suitable one. Moreover, We produce the feature importance distribution to analyze the reasons that lead to different energy usage patterns and further validate the assumption of potential connections between different performances of forecasting models and energy consumption patterns.

At the aggregation-level forecasting stage (Fig. 1 (b)), we also consider different forecast models based on the aggregated data, and analyze the results following the feature importance distribution. We further compare the performance of aggregation-level forecast with building-level forecast to explore the effect of the data aggregation process on the forecasting performance.

B. Forecast Models

1) Support Vector Regression: Support vector regression has been widely applied in energy forecasting applications for its high effectiveness in solving non-linear problems [10]. SVR adopts the structure risk minimization principle, which not only minimizes the training error, but also the upper bound of the generalisation error [11]. Given a training dataset $T = (\mathbf{x_1}, y_1), ..., (\mathbf{x_i}, y_i), ..., (\mathbf{x_m}, y_m)$, where $\mathbf{x_i} \in R^n$ denotes the i-th observation which is a n-dimensional input vector, $y_i \in R$ is the output corresponding to $\mathbf{x_i}$, and m denotes the size of training set. For non-linear SVR, the basic idea is to introduce kernel function to map the input space into a higher dimensional feature space efficiently, in which the problem becomes linearly separable [12]. The decision function of SVR can be represented in Eq. (1)

$$y = \langle w, \phi(x) \rangle + b \tag{1}$$

where $\phi(x)$ is the hypothetical higher dimensional feature space. Coefficients w and b need to be estimated based on the structure risk minimization principle.

A main advantage of SVR is that the loss function penalizes deviations that are greater than a threshold, which often leads to the sparse representation of the decision rule, thus brings major algorithmic and representational strength [13]. Hence in our paper, we selected SVR as one of the regression algorithms to perform energy forecasting.

2) Random Forest: Random forest [14] adopts the random method to establish a forest, which is an ensemble method. RF is often used for classification and regression problems. The RF for regression is detailed in Algorithm 1 [15].

In Algorithm 1, given a training dataset of size N, we build B regression trees on bootstrapped training samples of size N. When a split in a tree is considered, a random sample of m variables is selected as split candidates from the full set of variables. To make a prediction for a new point, it uses the average output of B regression trees, see Eq. (2).

RF achieves high prediction accuracy without increasing computation time. Moreover, RF is a robust algorithm when applied to data with missing values or imbalanced data [14].

Algorithm 1: Random Forest for Regression

for each $b \in [1, B]$ **do**
  select a bootstrap sample $\mathbf{Z}^*$ of size N from the training dataset
  set a minimum node number n_{min}, and grow a random forest tree T_b to the bootstrapped data
  repeat
    choose m variables randomly from the input variables
    select the best split point from m
    split the node into two sub-nodes
  until reach n_{min}
end for

Output: fully grown forest $\{T\}_1^B$

To make a prediction at a new point x:

$$\hat{f}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \tag{2}$$

Therefore, we consider RF as one of the comparison algorithms in our experiment.

3) Artificial Neural Network: The artificial neural network is a mathematical model inspired by the behaviour characteristics of biological neural network and performs distributed parallel information processing [16]. The backpropagation neural network (BP-NN) is the most widely used ANN. BP-NN is a forward multi-layer network, including the input layer, the hidden layer and an output layer. Connection weights of each layer and threshold values between each node are adjusted according to the error between the actual output and the expected output, which is repeated until the termination condition is satisfied. In this paper, BP-NN is selected for its strong non-linear fitting ability and low computing complexity [17].

C. Evaluation Indices

The error indicators of mean squared error (MSE), mean absolute error (MAE), mean absolute percentage error ($MAPE$), and R-squared (R^2) are considered in this paper to evaluate the forecast performance.

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(f_i - y_i)^2 \tag{3}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|f_i - y_i| \tag{4}$$

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}|\frac{f_i - y_i}{f_i}| \times 100\% \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(f_i - y_i)^2}{\sum_{i=1}^{m}(\bar{y}_i - f_i)^2} \tag{6}$$

where f_i, y_i and $\bar{y}_i$ are the actual electricity consumption, the predicted electricity consumption and the actual mean consumption, respectively. m denotes the size of the testing dataset, and i is the index of test observations.

Since MSE takes the square of the error, it will exaggerate the error caused by outliers. For MAE, it will be affected by the magnitude of the electricity consumption since it reflects the absolute error.

$MAPE$ and R^2 can overcome the aforementioned limitations and are suitable for forecast comparison among different forecast algorithms for different buildings of difficult electricity consumption levels. Specifically, $MAPE$ gives a percentage value, making the deviations comparable for different magnitudes of electricity consumption of different buildings. R^2, which has a range from 0 to 1, is also a desirable eveluation indicator.

III. Experimental Design

A. Data set and Preprocessing

The data used in this paper are collected from Cardiff Council [18], which includes different types of buildings such as office buildings, community facilities, schools, and cultural buildings. The data of each building includes half-hourly energy consumption and the physical information such as building type, year built, number of occupants, number of floors, and heating types. In this study, energy consumption data of three buildings from 2015 to 2016 are used. The detailed physical information of the considered buildings is shown in the Table I.

TABLE I
Physical information of each building

Type	ID	Year Built	Number of Floors	Usable Floor Area/m^2	Main Heating Type	No. of Occupants
Primary School	918	1967-1976	1	811.0	Electricity	192
Primary School	1022	1967-1976	1	1201.0	Gas	230
High School	1227	2013	2	14475.0	Biomass	1400

Although the above three buildings are all educational buildings from a macro perspective, they have different physical information at a high subdivision level. For example, the heating type of building 918 is electricity, while building 1022 and 1227 use gas and biomass respectively for heating. Moreover, for buildings with electricity as the main heating source, there are usually storage heater/hot water tank alongside to fully take advantage of time-of-use electricity prices (e.g. Economy 7 in the UK) where radiators/water tanks are heated and stored during the off-peak time periods (usually from around midnight to early morning) and are used later (e.g. during daytime). This leads to different electricity peak demand periods between building 918 and other buildings during cold seasons. In addition, building 1227 has the biggest usable floor area with the largest number of occupants among all three schools, therefore its electric consumption is likely to have higher magnitude compared with other buildings.

The considered dataset has some missing data. In general, missing values in the data can be handled using methods such as mean, hot deck and multiple imputations, etc. [19]. In this paper, we adopted mean imputation considering that only less than 10% data is missing.

B. Model Development

The whole development procedure of the forecast model is given in Fig. 2: firstly, the key influential usage features are determined and extracted from the preprocessed data; then three algorithms (SVR, RF, ANN) are applied to build forecast models with grid search being used to find the optimal combination of parameters for each model; finally, the performance of each model is evaluated through 5-fold cross-validation.

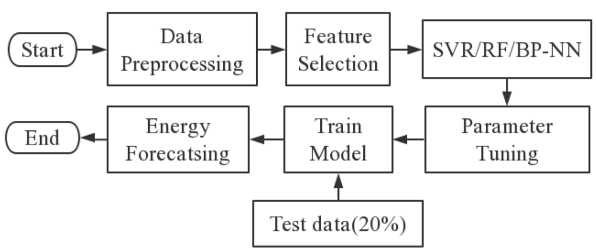


Fig. 2. Development pipeline for forecast model.

1) Feature Selection: Feature selection involves different factors that affect prediction performance. The selection of relevant, informative, different and independent high-quality features [20] is a crucial step in successful developing the forecast model.

Feature selection methods can be roughly divided into automatic and manual selection. The commonly used automatic feature selection methods include filter, wrapper and embedded methods [21]. For mannual feature selection, it often relies on expert domain knowledge and experience and is commonly used in the energy and power applications. In this study, we consider manual feature selection motivated by existing studies e.g. [22].

The features selected for predicting the energy consumption of h-th hour in this paper are listed in the Table II.

The temperature used in the experiment were retrieved from the world weather online [23]. One-hot encoding is performed to represent festival/non-festival[1], year, month of a year, day of a week and hour of a day.

2) Parameter Tuning: In the experiment, five models are established by using SVR, RF and BP-NN, among which three models are built by SVR using linear kernel function, polynomial kernel function, and radial basis function (RBF) respectively. We carry out grid search for parameters tuning, and the implementation steps are as follows:

1) Given a range of possible parameters, for the five models, their parameter ranges are determined via a trial and error manner. The range of parameters are shown in the Table III.
2) Parameters are paired to form parameter grids.
3) Relevant parameters are successively substituted into the model for the network nodes, and the optimal parameter combination is selected according to the best results.

Based on above steps, the best parameter combinations for models of each building are listed in Table IV.

[1]In the experiment, we considered the school holidays according to the historical school calendars of local schools in Cardiff, including four seasonal half-semesters, winter vacation, summer vacation and bank holidays.

TABLE II
SELECTED FEATURES FOR HOURLY LOAD FORECASTING

Input	Size	Description
L_h^{lw}	1	h-th hour load on the same day of last week
L_h^{yd}	1	h-th hour load of yesterday
L_{h-1}	1	Load of the last hour
T	1	Temperature of h-th hour (of today)
F	2	One-hot code for festival/non-festival day
Y	2	One-hot code for year index
M	12	One-hot code for month index
W	7	One-hot code for day index of a week
H	24	One-hot code for hour index of a day

TABLE III
OPTIMAL PARAMETERS FOR EACH FORECAST MODEL

Model	Parameters and range
SVR(Linear)	C:[0.01, 0.1, 1, 10]
SVR(RBF)	C:[0.01, 0.1, 1, 10], γ:[0.001, 0.01, 0.1, 1]
SVR(Poly)	C:[0.01, 0.1, 1, 10], degree:[3, 4, 5]
RF	max_depth:[5, 10, 15, 20], min_sample_split:[2, 5, 10, 15], n_estimators:[10, 100]
BP-NN	activation:['relu', 'tanh', 'sigmoid'], optimizer:['SGD', 'Adam'], epochs:[1, 5, 15], batch-size:[5, 20]

TABLE IV
BEST PARAMETERS FOR EACH MODEL

Model	Building / Parameters	918	1022	1227
SVR(Linear)	C	0.1	1	10
SVR(RBF)	C	10	10	10
	γ	0.001	0.01	0.001
SVR(Poly)	C	10	10	10
	degree	3	3	3
RF	max_depth	20	20	20
	min_sample_split	5	5	5
	n_estimators	100	100	100
BP-NN	activiation	'tanh'	'tanh'	'relu'
	optimizer	'Adam'	'Adam'	'Adam'
	epochs	15	15	15
	batch-size	20	20	20

IV. ANALYSIS OF EXPERIMENTAL RESULTS

A. Comparison between Models

In this paper, the network structure of BP-NN is decided by trial and error, and 20-10-1 is finally selected in this study. That is, the number of nodes in the input layer, hidden layer and output layer are 20, 10 and 1 respectively.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

TABLE V
PREDICTION SCORE FOR MODELS OF EACH BUILDING

	SVR (Linear)	SVR (Poly)	SVR (RBF)	RF	BP-NN
			Building 918		
R^2	0.823	0.860	0.870	0.886	0.880
$MAPE$ [%]	20.76	22.76	16.38	16.93	28.72
			Building 1022		
R^2	0.946	0.966	0.979	0.984	0.982
$MAPE$ [%]	6.52	8.68	4.53	4.04	9.15
			Building 1227		
R^2	0.953	0.827	0.882	0.979	0.865
$MAPE$ [%]	5.81	11.27	4.32	3.97	10.61

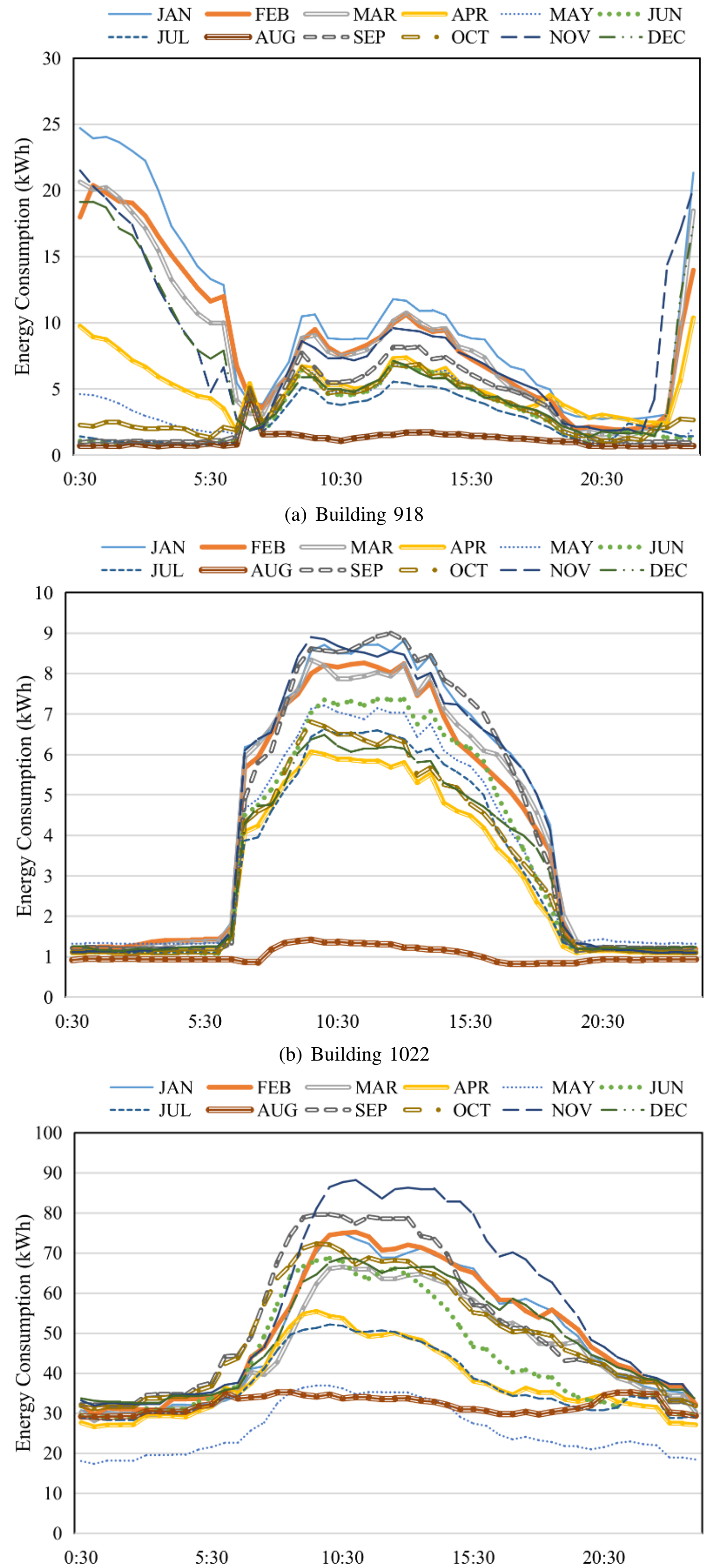
As mentioned previously, since $MAPE$ and R^2 give relative metrics to quantify the performance of different forecast models across different buildings, this paper only reports these two metrics. For each building, the forecast accuracy of five forecast models is shown in Table V.

As can be seen from the above table, in terms of model accuracy, RF and SVR (RBF) perform better than other models in general. It is worth mentioning that the forecast accuracy of building 1022 and 1227 is significantly higher than building 918. An intuitive explanation is that the data distribution of building 918 is very different from the other two buildings. As aforementioned, the heating type of building 918 is electricity whereas heating types of 1022 and 1227 are gas and biomass respectively, which is likely to lead to totally different electricity consumption patterns for building 918 and is more difficult to forecast. For building 918, both SVR (RBF) and RF produce comparable performance. Similarly, both SVR (RBF) and RF produce comparable performance for building 1022 with RF giving a better forecast accuracy in terms of R^2 and $MAPE$. As for building 1227, SVR (Linear), SVR (RBF) and RF all produce comparable performance with RF recording the best forecast accuracy on R^2 and $MAPE$.

To further investigate the relationship between model performance and building characteristics, the average daily electricity consumption of three buildings in each month over two years (2015 and 2016) are shown in Fig. 3. It can be found that building 918 has significantly higher consumption during late night and early morning in months with cold weather (from November to March). It also has two peak consumptions during daytime (around breakfast time and lunch time). Recall that the heating type of building 918 is electricity, and it is likely that the building has adopted times-of-use tariffs and electricity storage heater/hot water tank. In order to reduce the energy cost, take the storage heater for example, it charges over the night (i.e. store heat) when the electricity price is low and discharges (i.e. release heat) during the day when price is high. Compared with the other two buildings which only have peak consumptions in the day time, the energy consumption pattern of building 918 is much different.

B. Feature importance

In this subsection, the feature importance distribution based on RF is reported and analyzed. For each building, the first



(a) Building 918

(b) Building 1022

(c) Building 1227

Fig. 3. Average daily curves of each building in each month over two years.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

TABLE VI
FIRST 10 MOST IMPORTANT FEATURES

Building	918	Building	1022	Building	1227
Feature	Feature Importance	Feature	Feature Importance	Feature	Feature Importance
Last hour	0.737677669	Last hour	0.9200781	Last hour	0.938889721
Last week	0.121576466	Hour 7	0.0226223	Last day	0.017619565
Last day	0.064341659	Last day	0.0161636	Last week	0.015196115
Hour 23	0.009931366	Last week	0.0092828	Hour 7	0.007937783
Hour 6	0.009241794	Hour 6	0.0086995	Hour 8	0.006680166
Saturday	0.00772556	Hour 18	0.0047512	Hour 9	0.002967179
Hour 8	0.00627628	Hour 8	0.0026321	Temperature	0.001843714
Temperature	0.005447298	Hour 9	0.0023456	Hour 6	0.000879435
Friday	0.005053048	Temperature	0.0019667	Hour 12	0.000462711
Hour 0	0.002883764	Hour 19	0.0014589	Hour 10	0.000454611

10 most important features are selected and displayed in the Table VI.

The results showed that key influential features vary from different buildings although the most critical feature (electricity consumption of last hour) is the same for all three buildings. In addition, last day and last week are another two important common features for three buildings. For building 918, hour 23 and hour 6 are also important features. As can be seen from the Fig. 3 (a), these hours are turning points of different electric usage patterns. For buildings 1022 and 1227, the electricity consumption changes sharply (especially for building 1022) at 7:00 AM, which represents the turning point for working/non-working hours. As a result, hour 7 forms an essential feature for those two buildings.

By looking at the distribution of feature importance for the three buildings, it can be found that the distribution of building 918 is relatively more even than buildings 1022 and 1227. This is possibly due to the fact that building 918 utilizes electricity for heating. Therefore, features such as temperature and months have greater impact on its electricity consumption. In contrast, seasonality and temperature have a much less impact on the electricity consumption of the other two buildings.

C. Aggregated Analysis

In this subsection, we investigate the energy forecast performance based on the aggregation data of three buildings. The forecast accuracies of different models are reported in the Table VII while the first 10 most important features are shown in the Table VIII. By comparing with the forecast performances of individual buildings, we aim to investigate: 1) whether some patterns/features that are important to individual buildings are still influential in energy forecasts based on the aggregated data, i.e. information lose in the data aggregation process; 2) whether the aggregation can mitigate high randomness in electricity consumption of individual buildings.

As can be seen in the Table VII, most forecast models produce acceptable performance. In other words, the forecast based on the aggregated data is more stable among different forecast models compared with the forecast of individual buildings.

TABLE VII
SCORES OF DIFFERENT MODELS

Model \ Score	R^2	$MAPE$
SVR(Linear)	0.9503	6.50%
SVR(Poly)	0.9036	14.14%
SVR(RBF)	0.9036	4.86%
RF	0.9787	4.49%
BP-NN	0.9506	9.67%

TABLE VIII
FIRST 10 MOST IMPORTANT FEATURES

Feature	Feature Importance
Last hour	0.93416817
Last day	0.023045343
Last week	0.017150075
Hour 8	0.00527625
Hour 9	0.003624921
Hour 7	0.002727989
Temperature	0.001958506
Saturday	0.001322719
Hour 12	0.000650304
Hour 15	0.000587423

Based on the feature importance reported in Table VIII, features that are unique to individual buildings (e.g. hour 23 for building 918, and hour 7 for the other two buildings) are not as important in the aggregated data. In addition, the feature importance distribution is similar to that of building 1227 (see Table VI). This might be because building 1227 has the highest magnitude of electricity consumption among three buildings, and therefore is more influential in the aggregated data.

Although the data aggregation only considers three buildings via a simple aggregation, there are some insights we can still get from the above analysis. First, it is clear that some important features for individual buildings will be lost in the data aggregation process. Therefore, there is a need to adopt a better aggregation approach rather than the simple aggregation (e.g. clustering-based approach [24]) to preserve as many important features as possible. Second, the data aggregation will reduce variations of individual buildings and create more stable forecast performance among different forecast models. In contrast, the forecast performance among different models for individual buildings varies much (see Table V). Therefore, model selection is particularly crucial for energy forecast of individual buildings.

V. CONCLUSION

In this paper, we proposed a framework of energy forecasting with building characteristics analysis to study how different building characteristics affect the performance of forecast models at both the individual building level and the aggregation level. To be more specific, SVR, RF and BP-NN were used to conduct short-term load forecasting by using real-world data. At the individual building level, following the comparative analysis of different forecast models, the feature importance was analyzed to further understand the defining patterns of each building. The results show that the forecast performance was closely related to the data distribution and electricity consumption patterns associated with each building. At the aggregation level, through an comparative analysis of forecast performance and feature importance, we find that some important features of individual buildings are lost in the data aggregation process. Nevertheless, the aggregation reduces the randomness and variations of individual buildings and produce more stable forecast performance among different models. In our future work, we will investigate the problem considering more buildings with different types (e.g. commercial, residential, etc.) and develop a better data aggregation process to preserve as many important features as possible to further improve the forecast performance.

REFERENCES

[1] L. Pérez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy and buildings*, vol. 40, no. 3, pp. 394–398, 2008.

[2] E. Mocanu, H. P. P. Nguyen, M. Gibescu, and W. L. Kling, "Deep learning for estimating building energy consumption," *Sustainable Energy, Grids and Networks*, vol. 6, no. 6, pp. 91–99, 2016.

[3] D. Srinivasan and M. Lee, "Survey of hybrid fuzzy neural approaches to electric load forecasting," in *1995 IEEE International Conference on Systems, Man and Cybernetics. Intelligent Systems for the 21st Century*, vol. 5. IEEE, 1995, pp. 4004–4008.

[4] Z. Wang and R. S. Srinivasan, "A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models," *Renewable and Sustainable Energy Reviews*, vol. 75, pp. 796–808, 2017.

[5] E. M. Ryan and T. F. Sanquist, "Validation of building energy modeling tools under idealized and realistic conditions," *Energy and Buildings*, vol. 47, pp. 375 – 382, 2012.

[6] M. Yalcintas and U. A. Ozturk, "An energy benchmarking model based on artificial neural network method utilizing us commercial buildings energy consumption survey (cbecs) database," *International Journal of Energy Research*, vol. 31, no. 4, pp. 412–421, 2007.

[7] J. G. Jetcheva, M. Majidpour, and W. Chen, "Neural network model ensembles for building-level electricity load forecasts," *Energy and Buildings*, vol. 84, pp. 214–223, 2014.

[8] L. M. Candanedo, V. Feldheim, and D. Deramaix, "Data driven prediction models of energy use of appliances in a low-energy house," *Energy and Buildings*, vol. 140, pp. 81–97, 2017.

[9] A. Lahouar and J. B. H. Slama, "Day-ahead load forecast using random forest and expert input selection," *Energy Conversion and Management*, vol. 103, pp. 1040–1051, 2015.

[10] M. W. Ahmad, M. Mourshed, and Y. Rezgui, "Tree-based ensemble methods for predicting pv power generation and their comparison with support vector regression," *Energy*, vol. 164, pp. 465–474, 2018.

[11] B. Dong, C. Cao, and S. E. Lee, "Applying support vector machines to predict building energy consumption in tropical region," *Energy and Buildings*, vol. 37, no. 5, pp. 545–553, 2005.

[12] Q. Li, Q. Meng, J. Cai, H. Yoshino, and A. Mochida, "Applying support vector machine to predict hourly cooling load in the building," *Applied Energy*, vol. 86, no. 10, pp. 2249–2256, 2009.

[13] K. Roy, S. Kar, and R. N. Das, *Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment*. Academic press, 2015.

[14] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[15] D. Ruppert, "The elements of statistical learning: Data mining, inference, and prediction," *Journal of the American Statistical Association*, vol. 99, no. 466, pp. 567–567, 2004.

[16] S. Haykin, *Neural Networks and Learning Machines, 3rd Edition*. McMaster University, 2009.

[17] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," in *North American Power Symposium 2010*. IEEE, 2010, pp. 1–8.

[18] "Cardiff council carbonculture," https://platform.carbonculture.net/communities/cardiff-council/19/, [Accessed: 06-11-2019].

[19] A. Perez, R. J. Dennis, J. F. Gil, M. Rondon, and A. Lopez, "Use of the mean, hot deck and multiple imputation techniques to predict outcome in intensive care unit patients in colombia," *Statistics in Medicine*, vol. 21, no. 24, pp. 3885–3896, 2002.

[20] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[21] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.

[22] "World weather," https://www.worldweatheronline.com, [Accessed: 17-11-2019].

[23] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and J. He, "Short-term load forecasting with deep residual networks," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3943–3952, 2019.

[24] X. Fu, X. Zeng, P. Feng, and X. Cai, "Clustering-based short-term load forecasting for residential electricity under the increasing-block pricing tariffs in china," *Energy*, vol. 165, pp. 76–89, 2018.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE