Proceedings of The ASME 2018 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2018
August 26-29, 2018, Quebec City, Canada

DETC2018-86084

SHORT-TERM LOAD FORECASTING WITH DIFFERENT AGGREGATION STRATEGIES

Cong Feng *
The University of Texas at Dallas
Richardson, TX 75080
Email: cong.feng1@utdallas.edu

Jie Zhang †
The University of Texas at Dallas
Richardson, TX 75080
Email: jiezhang@utdallas.edu

ABSTRACT

Effective short-term load forecasting (STLF) plays an important role in demand-side management and power system operations. In this paper, STLF with three aggregation strategies are developed, which are information aggregation (IA), model aggregation (MA), and hierarchy aggregation (HA). The IA, MA, and HA strategies aggregate inputs, models, and forecasts, respectively, at different stages in the forecasting process. To verify the effectiveness of the three aggregation STLF, a set of 10 models based on 4 machine learning algorithms, i.e., artificial neural network, support vector machine, gradient boosting machine, and random forest, are developed in each aggregation group to predict 1-hour-ahead load. Case studies based on 2-year of university campus data with 13 individual buildings showed that: (a) STLF with three aggregation strategies improves forecasting accuracy, compared with benchmarks without aggregation; (b) STLF-IA consistently presents superior behavior than STLF based on weather data and STLF based on individual load data; (c) MA reduces the occurrence of unsatisfactory single-algorithm STLF models, therefore enhancing the STLF robustness; (d) STLF-HA produces the most accurate forecasts in distinctive load pattern scenarios due to calendar effects.

Keywords: Load forecasting, machine learning, aggregation forecasting, hierarchical forecasting, gradient boosting machine, random forest

1 INTRODUCTION

With the development of smart meter techniques, massive amounts of data enable load forecasting (LF) to have a more critical impact on power system operations. In power systems, decision-making problems on di□erent time-scales and various hierarchies rely heavily on accurate load forecasts [1]. According to time-scales, LF can be classified into short-term LF (up to 1-week-ahead), medium-term LF (from 1-week-ahead to 1-year-ahead), and long-term LF (more than 1-year-ahead) [2, 3]. Based on power system hierarchy, LF can be categorized into transmission side forecasting, distribution side forecasting, and demand side forecasting.

Short-term load forecasting (STLF) is adopted to assist in a number of power system operations, such as ramp detection, generation scheduling, load switching, energy trading, etc [4]. With the rapid development of artificial intelligence (AI), various forecasting models have been developed in the literature [5]. Lusis *et al.* [6] developed 1-day-ahead (1DA) load forecasting models considering calendar e□ects and found regression trees outperformed artificial neural network (ANN) and support vector machine (SVM). Ahmad *et al.* [7] proposed an accurate and fast converging 1DA load forecasting model based on mutual information and ANN, which decreased the average execution time while en-

*PhD Student, Department of Mechanical Engineering, ASME Student Member.

†Assistant Professor, Department of Mechanical Engineering, ASME Professional Member. Address all correspondence to this author.

Copyright © 2018 by ASME

hanced the forecasting accuracy compared with benchmark methods. A more comprehensive review of AI methods for STLF can be found in recent review papers [2, 8].

A power system has a hierarchical structure that includes levels of appliance, customer, feeder, substation, district, region, and system. STLF at every hierarchical level is valuable therefore has been researched in the recent decades. For example, system-level STLF was improved by clustering customers based on similar load consumption patterns in [9]. Fan *et al.* [10] developed region-level STLF technologies for a power system that covers large geographical area. Factors that affect district-level STLF were analyzed and a framework of district level STLF was proposed in [11]. Goude *et al.* [12] forecasted short-term and medium-term load at substation-level for over 2,200 substations in the French distribution network using semi-parametric additive models, which achieved satisfying accuracy. Ledva *et al.* [13] developed a model for feeder-level load forecasting based on an online learning algorithm and obtained accurate disaggregated forecasts. To improve the customer-level STLF, Yu *et al.* [14] added sparse coding in the ridge regression model and obtained a 10% improvement. Appliance-level STLF was generated by a type of hidden Markov model from smart meter data in [15].

To achieve better forecasting performance, a number of methodologies have been reported in the literature, which could be generally divided into three categories. Research in the first category focuses on integrating more informative and better-organized data to enhance the forecasting accuracy, which is defined as *information aggregation (IA)* in this paper. For example, residents' life patterns were used to improve the customer-level STLF in [16]. The second category contains methodologies combining forecasts from individual models [17], which is defined as *model aggregation (MA)*. For example, Borges *et al.* [18] developed a model combining independent forecasts to improve STLF. The third category of research is called hierarchical forecasting, which we define as *hierarchy aggregation (HA)*. Forecasts of lower-level individuals are aggregated by a certain strategy to improve forecasting accuracy of upper-level individuals in the power system hierarchy in this category. For example, a hierarchical forecasting method was developed in [19] for STLF, which was more efficient in terms of the sparsity of the adjustments and the prediction accuracy. All these three kinds of aggregation forecasting methodologies have been proven to enhance the accuracy. However, the superiority of the three aggregation forecasting strategies has not been studied in the literature.

In an attempt to comprehensively compare the aggregation strategies at different stages in the forecasting process, STLF models with IA, MA, and HA are developed to aggregate inputs, models, and forecasts, respectively. A set of 10 machine learning models based on 4 AI algorithms are built to ensure the generality of this study. Performance of models in different groups are compared to show the pros and cons of the three aggregation strategies. The remainder of this paper is organized as follows. STLF models with IA, MA, and HA are developed in Section 2. Section 3 describes the data for case studies, benchmarks, and evaluation metrics. In Section 4, results of case studies are analyzed and compared. Section 5 concludes the paper.

2 SHORT-TERM LOAD FORECASTING METHODOLOGIES WITH DIFFERENT AGGREGATION STRATEGIES

Three types of aggregation strategy (IA, MA, and HA) are described and formularized in this section. The three aggregation strategies aggregate distinct objects at different stages in the forecasting process, as illustrated in Fig. 1.

2.1 Information Aggregation (IA)

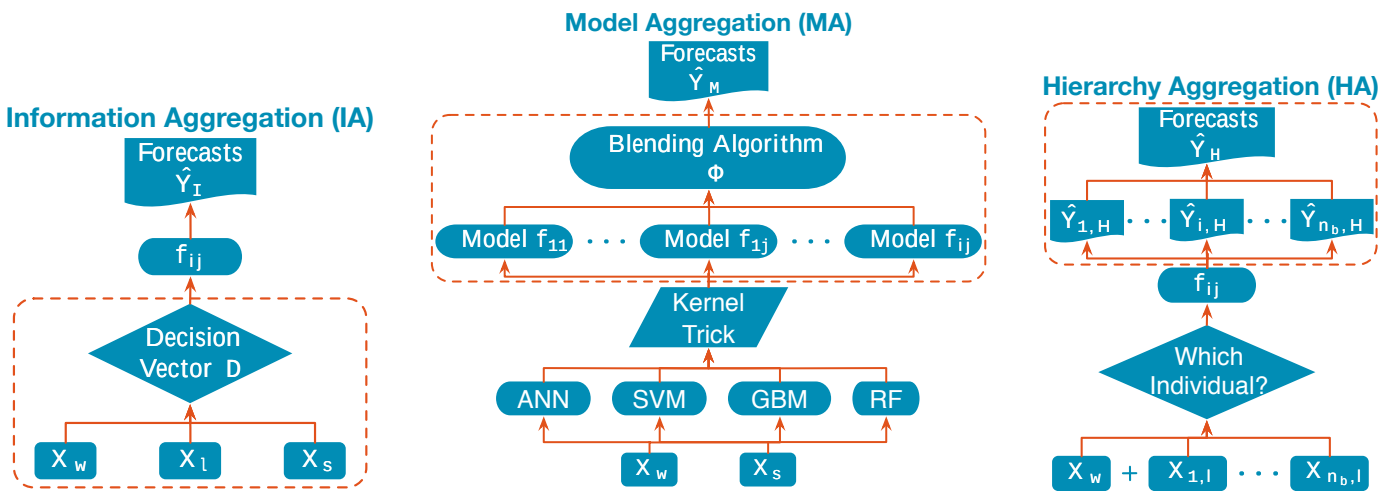
The first generation of STLF only depends on the load time series itself, which is called time series approach. External information, such as meteorological data and calendar data, is integrated into STLF. With the development of advanced metering infrastructure, smart meter data provides an opportunity to further improve STLF accuracy. In this paper, STLF with three sets of inputs is studied, which are: (i) weather data (X^w) and target variable (load at the top-level, which is denoted as X^s) data, (ii) individual load data (load data at bottom-level, which is denoted as X^l) and target variable data, (iii) weather data, target variable data, and individual load data. The STLF with IA (STLF-IA) conducts aggregation at the first step in the forecasting process, as illustrated in Fig. 1(a). STLF-IA is formularized as follows:

$$\hat{Y}_l = f_{ij}(X^D) \tag{1}$$

$$X^D = XD = \left[X^s_{n\times 1} \; X^w_{n\times d_w} \; X^l_{n\times d_l} \right] D \tag{2}$$

where n is the data length, d_w and d_l are the dimensions of weather data and individual load data, respectively. $f_{ij}(*)$ is the model using the ith AI algorithm with kernel j. $\hat{Y}_l$ is a forecasting vector produced by STLF-IA. X is a data matrix with all variables, and X^D is a selected input matrix. D is a decision vector that has three forms in terms

Copyright © 2018 by ASME



(a) Information aggregation (b) Model aggregation (c) Hierarchy aggregation (2-layer)

FIGURE 1. Frameworks of STLF with three different aggregation strategies

of the three IA scenarios:

$$D_{(1+d_w+d_l)\times(1+d_w)} = \begin{bmatrix} I_{(1+d_w)} & \vec{0}_{d_l\times(1+d_w)} \end{bmatrix}^T \tag{3a}$$

$$D_{(1+d_w+d_l)\times(1+d_l)} = \left[\begin{array}{c|c} 1 & \vec{0}_{(1+d_w)\times d_l} \\ \hline \vec{0}_{d_w\times 1} & \\ \vec{0}_{d_l\times 1} & I_{d_l} \end{array}\right] \tag{3b}$$

$$D_{(1+d_w+d_l)\times(1+d_w+d_l)} = I_{(1+d_w+d_l)\times(1+d_w+d_l)} \tag{3c}$$

where I and $\vec{0}$ are an identity matrix and a matrix of zeros, respectively.

2.2 Model Aggregation (MA)

Model Aggregation (MA) carries out aggregation at the model-building stage. MA is expected to take advantage of learning power from different algorithms. In the literature, averaging forecasts generated by several models is the first MA strategy, which is followed and advanced by a linear combination of models. Latest MA strategies seek to determine the weights of individual model forecasts dynamically by using AI algorithms, such as multi-model forecasting framework (MMFF) as shown in Fig. 1(b) [20, 21]. MMFF is a two-layer machine learning based method for short-term forecasting, which can be described as [17]:

$$\tilde{Y}_{ij} = f_{ij}(X^w) \tag{4}$$

$$\hat{Y}_{ij} = \Phi_{ij}(\tilde{Y}) \tag{5}$$

where $\tilde{Y}_{ij}$ is a forecast vector provided by the model f_{ij}, X_w is a input vector to the first-layer models, $\tilde{Y}$ is a combination of the first-layer forecast vector, and $\hat{Y}_{ij}$ is the final forecast vector by a blending model $\Phi_{ij}(*)$ in the second layer. Four AI algorithms with multiple kernels (shown as kernel trick box in Fig. 1(b)) are adopted in this paper, which are ANN, SVM, gradient boosting machine (GBM), and random forest (RF). Please note that all the models with various machine learning algorithms and kernels are used to construct the first layer, and only one of the models is selected as the blending algorithm in the MA framework.

2.3 Hierarchy Aggregation (HA)

Load data is hierarchically aggregated based on the electric connections and geographical distributions. STLF-HA forecasts load of bottom-level individuals (denoted as $\hat{Y}_{i,H}$ in Fig. 4) by using the weather data and specific individual load data (denoted as $X_{i,H}$ in Fig. 4) and then aggregates to the top-level load forecasts. For example, in a three-level hierarchy shown in Fig. 2, forecasting vectors at the bottom-level (level 3) are aggregated to the upper-level (level 2) until reaching the top-level (level 1). Numbers before the slash in subscripts indicate the upper-level component to which the lower-level individuals belong, and numbers after the slash in subscripts are used to identify individuals within the same aggregation group. For example, $\hat{Y}_{1/1,H}$ is a forecasting vector at the bottom-level (level 3), which is aggregated with another individual, $\hat{Y}_{1/2,H}$, in the same aggregation group to the upper-level component

Copyright © 2018 by ASME

$\hat{Y}_{1,H}$ (which could be expressed as $\hat{Y}_{i,H} = \sum_j \hat{Y}_{i/j,H}$). This process can be expressed using matrix notation $\hat{Y} = S\,\hat{Y}_{i/j,H}$, which is further expanded as [22]:

$$\begin{bmatrix} \hat{Y}_H \\ \hline \hat{Y}_{1,H} \\ \hat{Y}_{2,H} \\ \hline \hat{Y}_{1/1,H} \\ \hat{Y}_{1/2,H} \\ \hline \hat{Y}_{2/1,H} \\ \hline \hat{Y}_{2/2,H} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ \hline & & I_4 & \end{bmatrix} \begin{bmatrix} \hat{Y}_{1/1,H} \\ \hline \hat{Y}_{1/2,H} \\ \hline \hat{Y}_{2/1,H} \\ \hline \hat{Y}_{2/2,H} \end{bmatrix} \tag{6}$$

where $\hat{Y}$ is a forecasting matrix containing all elements in the hierarchy, S is a summing matrix, and I_4 is a 4×4 identity matrix. Please note that in this paper, the objective is to forecast load at the top-level, $\hat{Y}_H$.

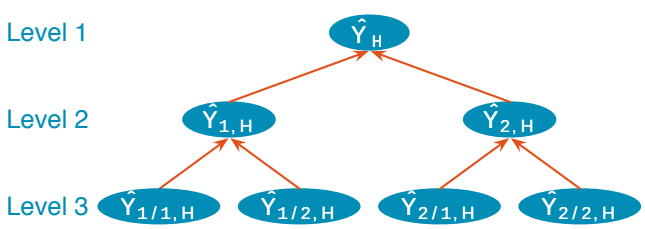


FIGURE 2. A three-layer hierarchical structure

3 EXPERIMENTAL SETUP

In this section, experimental setups for case studies are described including data description and pre-analysis, benchmarks and comparison settings, and evaluation metrics. While the methods can be applied to different forecasting horizons, the forecasting time horizon in this paper is 1-hour-ahead (1HA). 1HA load forecasting plays an important role in power system operations, such as helping decision-making of real-time dispatch and energy storage charging/discharging. 1HA load forecasting is also flexible and scalable to generate longer-term forecasts in a recursive manner.

3.1 Data Description and Pre-analysis

In this paper, hourly load data of 13 buildings (selected based on the data availability) at The University of Texas at Dallas (UTD) was used for case studies. The reasons to research with university campus load are threefold: (a) demand-side load forecasting is more challenging than upper-level load forecasting in power system hierarchy, (b) large electricity consumers, such as universities, are more critical in demand-side management, (c) a university campus has buildings with diverse load patterns that are interesting to explore. In addition, hourly weather information was retrieved from the National Solar Radiation Database (NSRDB) [1]. The weather features in NSRDB dataset include air temperature, relative humidity, air pressure, wind speed, wind direction, direct normal irradiance, global horizontal irradiance, and diffuse horizontal irradiance. Calendar features, i.e., hour of the day, day of the week, and month of the year, were extracted and included in the case studies. Both UTD load and NSRDB weather data have two-year length, spanning from January 1st 2014 to December 31st 2015.

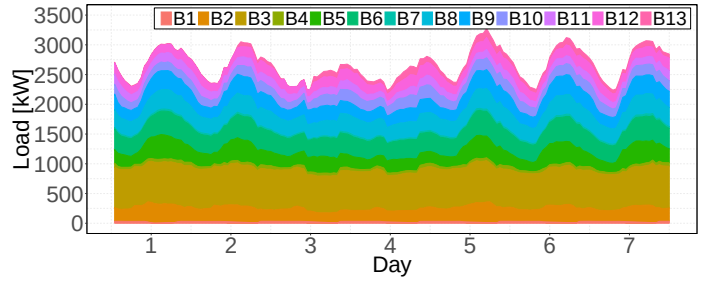


FIGURE 3. UTD campus and building load profile for seven days in spring

Figure 3 shows load profiles of the total 13 buildings (which is the top-level object in HA) and each individual (which is the bottom-level individuals in HA). It is observed that the load profiles have evident diurnal patterns. This is also proved by a time series analysis that shows all the load time series have the periodicity of 24 (1 day) [21, 23]. Moreover, load patterns of the 13 buildings are different, which could be further validated by load statistics shown in Fig. 4. Among the 13 buildings, B1 is a parking structure with photovoltaic panels, which may have negative netload during daytime, as shown in Fig. 4. B2 is an administration building that has large load and variance from 8am to 5pm. B3 is a library that has the largest and most stable load among all buildings. B4 is a lecture hall, which has relatively small but chaotic load. B5 - B9 are five classroom/lab buildings with similar patterns. B10 - B13 are four student residence halls that have diverse load patterns in contrast to other buildings. Compared to individual buildings, the whole campus load is relatively smoother.

3.2 Benchmarks and Comparison Settings

In this paper, forecasting methods with three kinds of aggregation strategies are investigated and compared, which are IA, MA, and HA. First, STLF-IA is compared with

[1] https://nsrdb.nrel.gov

Copyright © 2018 by ASME

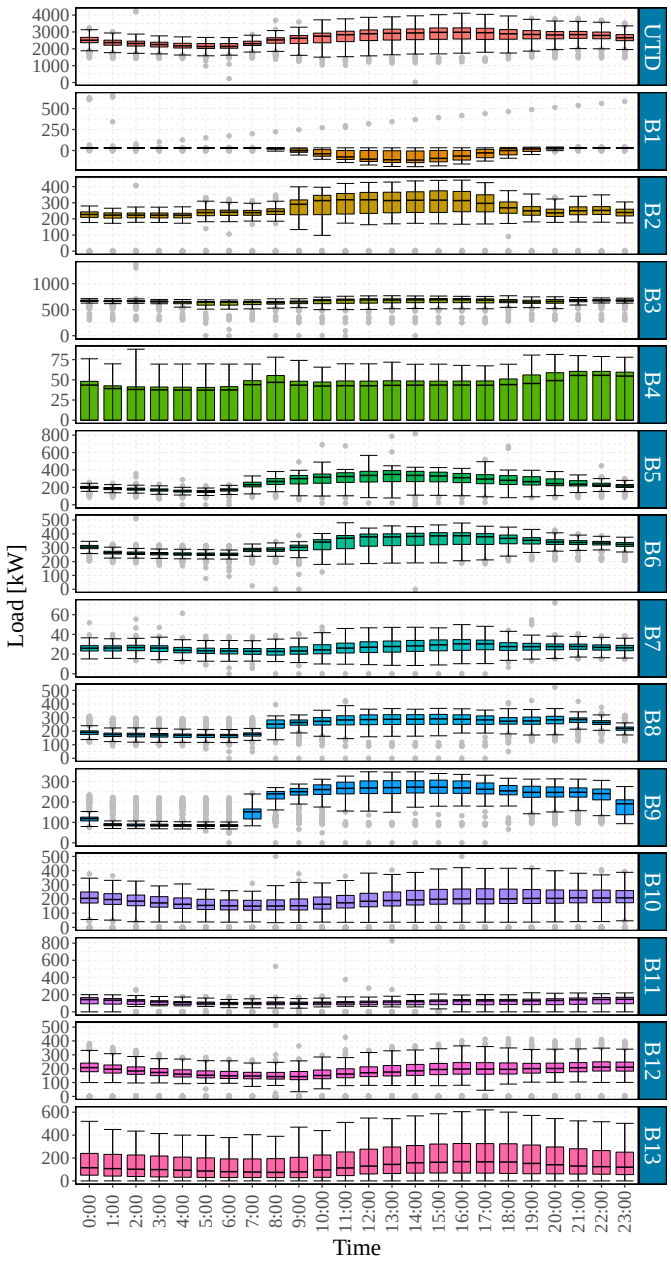


FIGURE 4. Hourly statistics of buildings' and whole campus' load. The line in the box is the median. The interquartile range box represents the middle 50% of the data. The upper and lower bounds are maximum and minimum values of the data, respectively, excluding the outliers. The outliers are data outside two and half times of the interquartile range

STLF using weather data (STLF-W) and STLF using individual buildings' load data (STLF-L). Note that the historical whole campus load data is included in STLF-IA, STLF-W, and STLF-L. The second comparison is made between STLF-MA and STLF with single-algorithm machine learning models (STLF-S), both of which use weather data and the historical campus load data (so STLF-S is the same as STLF-W). At last, STLF-HA is compared with STLF-IA, STLF-MA, and STLF-W to show the e□ectiveness of aggregation strategies in STLF.

TABLE 1. Machine learning forecasting models with different kernel/distribution functions/training algorithms

Algorithm	Model	Function/Algorithm
ANN	M1	Resilient back-propagation (BP)
	M2	Momentum-enhanced BP
	M3	Standard BP
SVM	M4	Linear kernel
	M5	Polynomial kernel
	M6	Radial basis function kernel
GBM	M7	Squared loss
	M8	Laplace loss
	M9	T-distribution loss
RF	M10	CART aggregation

The forecasting models (f_{ij} in Fig. 1) adopted are tabulated in Table 1, including four machine learning algorithms with di□erent kernel functions/distribution functions/training algorithms [21]. It is important to note that all these models are used in the first-layer and only one of them is used in the second-layer in MA. The training data was randomly selected from each month, and the remaining data was used for testing. The ratio of training samples to testing samples was 4:1. The experiment was carried out on a laptop with an Intel Core i7 2.6 GHz processor and a 16.0 GB RAM.

3.3 Forecasting Accuracy Assessment

To assess the forecasting accuracy, four evaluation metrics are used, which are normalized mean absolute error (nMAE), mean absolute percentage error (MAPE), nMAE improvement (Imp^A), and MAPE improvement (Imp^P). The mathematical expressions of the four metrics

Copyright © 2018 by ASME

are respectively shown as [24, 25, 26]:

$$nMAE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_{max}}\right| \times 100\% \quad (7)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100\% \quad (8)$$

$$Imp^{A}_{ab} = \frac{nMAE_{M_b} - nMAE_{M_a}}{nMAE_{M_b}} \quad (9)$$

$$Imp^{P}_{ab} = \frac{MAPE_{M_b} - MAPE_{M_a}}{MAPE_{M_b}} \quad (10)$$

where $\hat{y}$, y, and y_{max} are the forecast value, actual value, and maximum actual value, respectively; i is a sample index and n is the number of samples; M is the model name; the parameters of a and b are group indices to which a model belongs. Specifically, a and b could be selected from I, M, H, W, L, and S, which represent the groups of STLF-IA, STLF-MA, STLF-HA, STLF-W, STLF-L, and STLF-S, respectively. It is important to note that both Imp^{A} and Imp^{P} are calculated based on the same model M, because the focus of this paper is to compare STLF with different aggregation strategies, instead of comparing STLF using different machine learning models.

4 RESULTS AND DISCUSSION

Tables 2 and 3 list the four evaluation metrics of STLF benchmarks and STLF with three aggregation strategies. Comparisons are made and discussed from different perspectives.

The effectiveness of STLF-IA is evaluated by the first six columns in Table 2 and the first four rows in Table 3. It is found that STLF-IA models reduce forecasting errors significantly and consistently, compared with STLF-W and STLF-L models. The accuracy improvements are more evident by aggregating weather information data into forecasting models. Regarding to different models, M3 and M9 (which are an ANN and a GBM forecasting model, respectively) are enhanced the most by IA. *It is concluded that IA improves STLF forecasting accuracy significantly and consistently, especially when aggregating weather information.*

MA forecasting evaluation results are listed in the 7th and 8th columns in Table 2. The comparisons of MA with STLF-S are shown in the 5th and 6th rows in Table 3. It is found that the performance of relatively less-accurate STLF-S models are improved more significantly by MA, such as M7 and M8. However, the best two models in STLF-S, i.e., M9 and M10, deteriorate in STLF-MA, which is due to the unsatisfactory forecasts ($\tilde{Y}$) from part of the first-layer models. Regarding to models using different machine learning algorithms, all the ANN models (i.e., M1 - M3) and SVM with linear and polynomial kernels (i.e., M4 and M5) perform relatively better in STLF-MA. Among the four different ensemble learning algorithm models (M7 - M10), two of them (i.e., M7 and M8) have increasing accuracies while the other two (i.e., M9 and M10) have decreasing accuracies by using the MA strategy. Even though three models (i.e., M6, M9, and M10) produce worse forecasts, their forecasting accuracies are still competitive. *Therefore, it is concluded that STLF-MA enhances STLF robustness by improving the unsatisfactory single-algorithm machine learning models.*

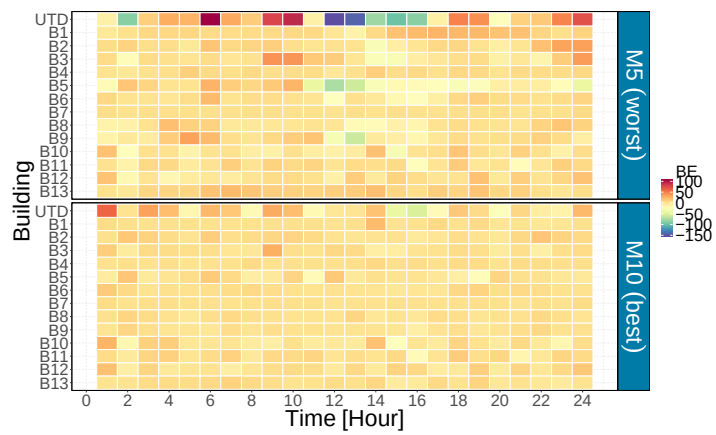


FIGURE 5. UTD campus and building forecasting bias errors (BEs) [kW] for STLF-HA

The forecasting errors of STLF-HA models are listed in the last two columns in Table 2 and their comparisons with STLF-S, STLF-IA, and STLF-MA are shown in the last six rows in Table 3. As opposed to STLF-S, the HA strategy improves STLF stably by up to 21.57% and 22.47% based on Imp^{A} and Imp^{P}, respectively. Most STLF-HA models outperform their counterparts in STLF-IA and STLF-MA groups, including ANN, GBM, and RF models. However, STLF-HA with SVM models (M4 - M6) is beat by the same models in the STLF-IA group. Moreover, four models (M4, M5, M7, and M8) produce worse forecasts with the HA strategy than those with MA. The forecasting accuracy deterioration of STLF-HA is due to the individual forecasting error accumulation effect, which is illustrated in Fig. 5. Two contrasts shown in Fig. 5 are STLF-HA with M5 and M10, which are the worst and the best STLF-HA models, respectively. It is observed that M10 generates forecasts with smaller bias for each building than M5, such as B5 and B9.

Copyright © 2018 by ASME

TABLE 2. Forecasting nMAE [%] and MAPE [%] of different STLF groups

Model	STLF-W/STLF-S		STLF-L		STLF-IA		STLF-MA		STLF-HA	
	nMAE	MAPE	nMAE	MAPE	nMAE	MAPE	nMAE	MAPE	nMAE	MAPE
M1	1.32	2.16	1.60	2.63	1.24	2.04	1.19	1.91	1.11	1.83
M2	1.34	2.20	1.57	2.56	1.28	2.11	1.17	1.89	1.11	1.83
M3	1.52	2.53	1.58	2.58	1.25	2.07	1.28	2.03	1.11	1.83
M4	1.45	2.34	1.57	2.54	1.31	2.12	1.11	1.79	1.36	2.18
M5	1.83	3.01	1.72	2.82	1.53	2.50	1.27	2.06	1.58	2.58
M6	1.34	2.17	1.43	2.34	1.18	1.94	1.39	2.30	1.20	1.97
M7	1.64	2.65	1.82	2.96	1.64	2.65	1.18	1.93	1.46	2.37
M8	1.67	2.71	1.83	2.96	1.67	2.70	1.23	2.03	1.48	2.38
M9	1.07	1.73	1.88	3.05	1.05	1.71	1.20	1.95	0.89	1.46
M10	1.06	1.74	1.30	2.14	1.06	1.74	1.11	1.81	0.83	1.35

Note: The forecasting nMAE and MAPE of the persistence method are 2.01% and 3.21%, respectively. Bold values indicate the best results within the same group, while bold navy blue values indicate best results among all models.

TABLE 3. Forecasting improvements according to Imp^{A} [%] and Imp^{P} [%] of different comparisons

Group		M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
STLF-IA	Imp^{A}_{IW}	5.83	4.74	18.00	9.69	16.03	11.51	0.00	0.03	1.61	0.00
	Imp^{P}_{IW}	5.73	3.89	18.25	9.49	16.84	10.67	0.23	0.23	1.08	0.28
	Imp^{A}_{IL}	22.35	18.41	21.01	16.63	10.94	17.12	10.15	8.62	22.96	18.05
	Imp^{P}_{IL}	22.65	17.48	19.99	16.68	11.33	17.16	10.49	8.67	23.49	18.61
STLF-MA	Imp^{A}_{MS}	10.03	12.62	16.16	39.52	12.33	-4.05	27.67	26.49	-12.06	-4.68
	Imp^{P}_{MS}	11.79	13.98	19.57	40.48	11.83	-6.04	27.14	24.89	-13.17	-3.93
STLF-HA	Imp^{A}_{HS}	15.97	17.48	27.10	6.25	13.28	10.05	10.95	11.38	16.54	21.57
	Imp^{P}_{HS}	15.13	16.80	27.68	6.76	14.25	9.31	10.85	11.87	15.35	22.47
	Imp^{A}_{HI}	10.77	13.37	11.09	-3.81	-3.27	-1.66	10.96	11.35	15.18	21.84
	Imp^{P}_{HI}	9.97	13.43	11.54	-3.01	-3.11	-1.52	10.64	11.66	14.43	22.25
	Imp^{A}_{HM}	6.60	5.56	13.05	-6.94	-43.38	13.55	-23.12	-20.56	25.53	25.07
	Imp^{P}_{HM}	11.99	10.42	12.91	-3.19	-32.92	33.66	-11.30	-7.72	31.46	28.34

Note: Bold values indicate the most improvements within the same comparison, while brick red values indicate accuracy deteriorations in comparisons.

Moreover, the individual building load forecasting errors of M5 accumulate to larger values in contrast with those of M10, which is illustrated by the darker color of the whole campus' load forecasting in Fig. 5. Even though there are some unsatisfactory models compared with other two aggregation strategies, the overall improvement of STLF-HA is significant. *Additionally, STLF-HA produces the most accurate forecasts (0.83%* nMAE *and 1.35%* MAPE *) among all models.*

The best model in each group is picked out to make

Copyright © 2018 by ASME

FIGURE 6. Forecasting errors by month

FIGURE 7. Forecasting errors by day of the week

FIGURE 8. Forecasting errors by hour of the day

further comparisons, which are M9 in the STLF-W/STLF-S group ($M9_W$), M10 in the STLF-L group ($M10_L$), M9 in the STLF-IA group ($M9_{IA}$), M4 in the STLF-MA group ($M4_{MA}$), and M10 in the STLF-HA group ($M10_{HA}$).

To characterize forecasting performance of the ve models, forecasting errors with respect to calendar units (i.e., month, day of the week, and hour of the day) are shown in Figs. 6 - 8. One interesting nding is that the calendar e ect has considerable impacts on forecasting errors. For example, errors in January, August, and September are much larger than those in other months. This is possibly due to the load pattern variation caused by university holidays. The calendar e ect on forecasting errors is even more evident by hour of the day, as shown in Fig. 8. Forecasts deviate the most from 6:00 to 8:00, during which load patterns change more signi cantly. However, no signi cant calendar e ect is found on forecasting errors by day of the week, as shown in Fig. 7. This is possibly due to the diverse building load of the university, for example, classroom and library buildings have higher load during weekdays and residential halls have higher load during weekends. Even though the load pattern varies signi cantly, it is observed that $M10_{HA}$ presents superior performance in every month, every day of the week, and at every hour of the day than the best models in other groups.

5 CONCLUSION

This paper investigated and compared short-term load forecasting (STLF) with di erent aggregation strategies, including information aggregation (IA), model aggregation (MA), and hierarchy aggregation (HA). The three aggregation strategies integrated distinct objectives at di erent stages in the forecasting process. STLF-IA aggregates more informative and better-organized data. STLF-MA aggregates forecasts of di erent machine learning models nonlinearly and dynamically. STLF-HA aggregates lower-level forecasts into higher level forecasts in the hierarchical structure. Case studies based on 2-year of university campus hourly load and weather data with 13 individual buildings showed that:

1) STLF with three aggregation strategies improved forecasting accuracy, compared with benchmarks without aggregation.
2) STLF-IA presented superior behavior than STLF with weather data and STLF with individual load data consistently.
3) MA improved the accuracy of unsatisfactory single-algorithm STLF models, therefore enhancing the STLF robustness.
4) STLF-HA produced the most accurate forecasts in distinctive load pattern scenarios caused by calendar effects.

Future work includes exploring di erent methods in each aggregation strategy, and investigating the exibility and

Copyright c 2018 by ASME