End-use Disaggregation in Commercial Buildings with the Building Automation System Trend Data

January 29, 2021

Abstract

End-use submetering is essential for energy management in large commercial and institutional buildings. However, most existing buildings lack adequate submetering even for major end-uses. End-use disaggregation techniques offer an untapped opportunity to supplement deficiencies in a metering network. This study presents an end-use disaggregation method for commercial buildings by using building automation system (BAS) data. The BAS trend data provide contextual information about the operational state of major energy-consuming systems and equipment such as fans, pumps, air handling unit (AHU) heating and cooling coils, and chillers. The method applies a series of multiple linear regression models disaggregating bulk metered heating, cooling, and electricity use data into different end-uses by using BAS data as predictors. The results demonstrate that the method can accurately disaggregate hourly building-level electricity, heating, and cooling use into their end-use categories.

Key Innovations

- **Disaggregation**
- **Use of bulk metered data**
- **Leveraging building automation system**
- **Regression-based approach**

Practical Implications

The practical implication of the proposed end-use disaggregation method is that it enables building operations staff to understand energy flows in a large commercial building at a much higher spatial and categorical resolution than the physical energy meters available. It will provide better energy use benchmarks and lead to the detection and interpretation of energy use anomalies.

Introduction

One of the major obstacles envisaged in building energy management and retrofits is the nonstandard and insufficient end-use submetering. Installations of new submetering can be difficult and expensive in some cases. For example, heating, ventilation, and air conditioning (HVAC) electricity consumption is hard to measure because the data is always mixed with lighting and plug loads (Ji et al. (2015)).

Methods for appliance-level load disaggregation in residential buildings are widely studied by the industry. Recent articles showed over 50 different methods that attempt to improve the accuracy and practicality of load disaggregation methods in home energy management (Esa et al. (2016), Klemenjak and Goldsborough (2016)). It has been stated that there are over 18 companies with commercially available or advanced prototype devices for load disaggregation in residential buildings (Mayhorn et al. (2015)). On the other hand, disaggregation of loads in commercial buildings is more challenging and comparatively understudied.

Norford and Leeb (1996) were the first researchers who looked for the load disaggregation problem in commercial buildings by developing a prototype device to determine simultaneously scheduled equipment as well as devices with similar power level. Shao et al. (2012) also presented a temporal motif mining-based load disaggregation method. Their method applies clustering to detect a discrete number of power levels and identifies the periodic occurrence patterns of these power levels. They demonstrated the method with ten-second interval data collected individually from two fans, a blower, a pump, and an elevator.

In commercial buildings, there are many multi-stage variable load devices such as pumps and fans that serve air and hot or cold water distribution systems and the plant equipment while many occupants interact with lighting and plug-in office equipment. In addition, unlike residential buildings, energy systems in commercial buildings are utilized with a building automation and control network whereby individual pieces of equipment turn on/off at the same time (Shao et al. (2013)). For example, in a building with an air-based cooling system, only when the AHU fans are scheduled to start, the chillers turn on – also triggering the chilled water pumps, cooling tower fans, etc. This functional causality leads to near-simultaneous changes in energy consumption of different equipment and intensifies the difficulty to discern individual effects. While load disaggregation for energy management in commercial buildings is as important as it is for residential buildings, operations staff of large commercial buildings with limited submetering work without detailed knowledge of how the energy is used and distributed. High resolution information about how and where energy is used in a large commercial building can enable operators to pinpoint energy intensive operational anomalies, establish end-use level energy benchmarks, and guide ongoing commissioning.

Batra et al. (2014) introduced a combinatorial optimization-based disaggregation method which demonstrated AHUs as a two-state load. Although, the method was successful at disaggregating loads at the floor-level, it failed to obtain the continuously varying loads by multiple AHUs for the whole-building. Therefore, they represented that a disaggregation algorithm's accuracy is related to the number, diversity, and power draw characteristics of equipment. Ji et al. (2015) developed a Fourier series model-based method to accurately disaggregate hourly lighting and plug load data from HVAC energy use in commercial buildings. Aside from the studies that attempt to disaggregate the HVAC energy use in commercial buildings, Rogriguez et al. (2016) and Doherty and Trenbath (2019) studied disaggregation of plug-in office equipment loads. They both submetered a small number of office devices such as monitors, desktop computers, laptops, printers, microwave ovens, etc. They inspected the correlation and cross-correlation of their steady-state power draw patterns and explored the viability to disaggregate their effects. A load disaggregation algorithm has been developed to isolate the impact of occupant activities on aggregate electricity use data by Rafsanjani and Ahn (2016) and Rafsanjani et al. (2018). They applied Wi-Fi-based occupancy detection and a density-based clustering approach to identify arrival and departure events to the changes in the power draw level. Along with these two case studies with eleven occupants, their algorithm illustrated the importance of external data sources that provide context such as Wi-Fi to gain high disaggregation resolution.

Virtual metering is also becoming more popular.It entails collecting long-term HVAC control system data to gain more understanding into unmetered energy end-uses and flows in the commercial building. Virtual meters are inverse models trained with HVAC control system data to compute unmeasured values. Unlike load disaggregation, virtual metering is a bottom-up technique generally relying on HVAC control system data. In this regard, Darwazeh et al. (2019) developed a virtual meter to compute the heat injected or removed by an AHU's heating and cooling coils, respectively. Utilizing supply, return, and outdoor air temperatures, supply airflow rate, outdoor air damper position, and heating and cooling coil valve position, the virtual meter computes the energy provided by the heating and cooling coils.

As stated earlier, in many commercial buildings, common electricity end-uses for lighting, plug loads, fans, pumps, and chillers are not metered separately. Further, even when meters for heating and cooling are present, distribution of heating into AHU heating coils and perimeter heating devices and cooling into AHU cooling coils are rarely submetered. Despite nonstandard and insufficient submetering in commercial buildings, the building automation system (BAS) trend data contain valuable insights into the operational state of energy systems and major equipment. For example, trend data from AHU fan schedules or variable frequency drive (VFD) states, and AHU supply air pressure sensors indicate when individual AHUs fans are operational. Similarly, even though a chiller's electricity use is not submetered, its effect on the building level hourly or sub-hourly electricity use data can be filtered out by looking at the state of the chilled water pumps. Wi-Fi device count data, which is a proxy for occupancy levels and accessible in an institutional IT network, can be a reasonable predictor for occupant-driven lighting and plug loads (Hobson et al. (2019) and Mahdavi et al. (2016)). Other than these electricity end-uses, trend data for AHU heating and cooling coils, and reheat and other perimeter heating device valves provide valuable insights regarding the distribution of heating and cooling into AHUs and thermal zones. Despite the potential of using BAS data for commercial building end-use disaggregation, only two studies developed end-use disaggregation models that use BAS data as regressors. Among them, Bansal and Schmidt (2017) used three-minute interval data for chiller flow rate, condenser water supply and return temperatures, and chilled water supply and return temperatures to disaggregate the electricity use of three large chillers. Burak Gunay et al. (2020) used hourly supply air pressure, fan VFD state, AHU fan schedule, chilled water pump state data extracted from an academic office building to disaggregate bulk electricity meter data into three end-uses: lighting & plug loads, fans & pumps, and chillers. To date, end-use disaggregation with BAS data has not been used to disaggregate heating and cooling meter data into individual AHUs and perimeter heating devices.

To this end, the contribution of this study is to demonstrate the potential of using previously listed meter and BAS data in disaggregating bulk electricity, heating, and cooling meters into end-uses. For electricity, these end-uses are fans and pumps, lighting and plug loads, and chillers. For heating and cooling, they are AHU heating and cooling coils, reheat coils, and other perimeter heating devices. The proposed method is demonstrated with measured data extracted from the BAS of an academic office building. The limitations of the method are discussed, and future work recommendations are developed for extensive testing and validation.

Methodology

The proposed method disaggregates building-level electricity, heating, and cooling energy use from meters into different subcategories. The cooling meter readings are disaggregated into individual AHUs, while heating meter reading are disaggregated into AHU heating coils, reheat coils and perimeter heating devices. Figure 1, 2 and 3 show the flow of electricity, cooling and heating in a commercial building (Burak Gunay et al. (2020)).

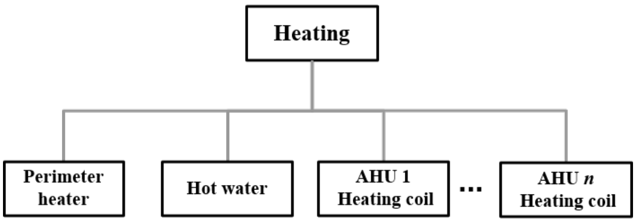

Figure 1: A schematic illustrating the flow of heating in a commercial building and the system-level submetering

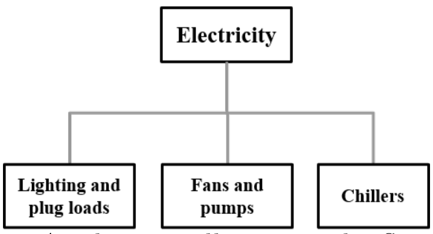

Figure 2: A schematic illustrating the flow of electricity in a commercial building and the system-level submetering

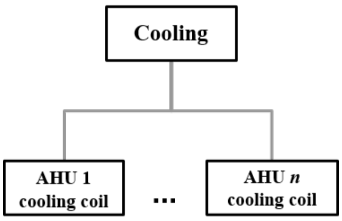

Figure 3: A schematic illustrating the flow of cooling in a commercial building and the system-level submetering

End-use Disaggregation Model

In this study, for end-use disaggregation, regression techniques are used with time-series data. For building energy disaggregation, the time series data can be either yearly, monthly, weekly, daily, hourly, or even smaller data resolutions such as 15 min. Regression analysis for energy disaggregation was used in a few papers (Kaselimi et al. (2019),(Chen et al. (2013), Schirmer and Mporas (2019) and Schirmer et al. (2019)).

Prior to model development, the observed data were preprocessed to identify whether it is needed to be cleaned. The data cleaning was based on two conditions, which are namely erroneous and missing data points. However, no missing data points were identified in this dataset. Occupancy data needed to be down-sampled from five minutes to one-hour intervals data. Based on Hobson et al. (2019)'s study, using ground truth data, the occupant counts are estimated by assuming 1.2 Wi-Fi-enabled device per person on average after removing background Wi-Fi device counts – which represent the peripheral Wi-Fi devices such as printers. Additionally, the heating and cooling valve position set to zero while metered heating/cooling is close to zero which means there is no hot/chilled water in pipes to respond to valve opening and closing. Moreover, AHU heating/cooling coils are set to zero when the AHUs are not running. In this study, the BAS data provide contextual information about the operational state of major energy-consuming systems and equipment therefore, regression analysis used to describe the relationship between a set of independent variables and the energy meter data.

Linear Models

Two regression models were developed for each of the three energy meters: Sci-kit-learn linear-model (Python library) and the genetic algorithm(GA). Equation 1 shows the formulation of a linear model where $\hat{y}$ is the predicted output and x_1 to x_p are inputs.

$$\hat{y}(\beta, x) = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p \tag{1}$$

Sci-kit-learn linear regression fits linear models that minimize the residual sum of squares between the observed targets in the dataset, and the targets predicted by the linear approximation. Learning the parameters of a prediction function and testing it on the same data is a methodological mistake, which may lead to over-parameterized models, also known as over-fitting. To avoid this problem, it is common practice when performing a supervised learning experiment to hold out part of the available data as a test set. In this study 80% of the data set was used for training the regression model and the rest was used for testing (sci-kit-learn developers (2020)).

The genetic algorithm minimized the root mean square error between the observed and predicted energy use. After a preliminary sensitivity analysis, the algorithm parameters were set as follows: Population size = 1000; Crossover probability = 0.5; Maximum number of iteration = 10 (Solgi (2020)).

The assessment results show the model accuracy for both methods is very similar but the computational time for Sci-kit-learn linear regression is significantly lower than the genetic algorithm.

The developed regression equations for each meter and the corresponding coefficients of determinations

(R^2) are presented in Equations 2 and 3:

$$Electricity = \beta_0 + \beta_1(Occupancy) + \beta_2(FanPressure\ AHU_1) + \beta_3(FanPressureAHU_2)$$

$$Heating = \beta_0 + \beta_1(HeatingCoilAHU_1) + \beta_2(Heating\ CoilAHU_2) + \beta_3(RadiatorStateforZones)$$

$$Cooling = \beta_0 + \beta_1(CoolingCoilAHU_1) + \beta_2(Cooling\ CoilAHU_2) \quad (2)$$

Case Study and Dataset

The end-use disaggregation methods are demonstrated on a dataset observed from an academic office building in the city of Ottawa, Canada. The six-story building has a floor area of 6000 m^2 and it was constructed in 2011, (Burak Gunay et al. (2020)). In this study, we used the hourly data gathered in 2019.

Exploratory Data Analysis

Since energy consumption is affected by various features in a complicated manner, the results of statistical analysis between energy meter and corresponding BAS data variable presents the strength of variables selected for estimating each end-use energy consumption (Lee et al. (2019)). However, the reliability of the linear regression model also depends on how many observed data points are in the sample. In this section, the Pearson correlation coefficient and P-value are used for testing non-correlation. The Pearson correlation coefficient measures the linear relationship between two groups of variables. In addition, the calculation of the P-value relies on the assumption that each dataset is normally distributed. Kowalski (1972) shows the effects of non-normality of the input on the distribution of the correlation coefficient. Similar to other correlation coefficients, Pearson varies between -1 and +1 with 0 indicating no correlation. Correlations of -1 or +1 show an exact linear relationship. Positive correlations imply that as x increases, so does y. Negative correlations imply that as x increases, y decreases.

The P-value indicates the probability of an uncorrelated system producing data sets that have a Pearson correlation at least as extreme as the one computed from these data sets.

According to the correlation analysis for the cooling meter in Table 1 and the scatter matrix plot in Figure 5 (a), the R-value for cooling valve position in both AHUs is 0.85 the test concludes that there is sufficient evidence that there is a significant linear relationship between these features and cooling meter because the coefficient is significantly different from zero. Also, the P-value for cooling valve position for both AHUs is less than a significance level of 5% , $\alpha = 0.05$, therefore, both variables can be used for linear prediction. For heating energy use, the R and P-value corresponded to radiator state for zones has a very strong linear relationship with meter data in comparison to heating valve position. The correlation coefficients are shown in Table 2 and scatter matrix plot in Figure 5 (b). For electricity use, the fan pressure for AHU 1 and occupancy data play a strong role in predicting the output while the fan pressure for AHU 2 has a moderate relationship for linear regression prediction, the correlation results for electricity is shown in Table 3 and scatter matrix plot in Figure 4.

Table 1: Correlation analysis for cooling.

Variable	R	P-value
Cooling	1	0
Cooling Coil AHU 1	0.850	0
Cooling Coil AHU 2	0.849	0

Table 2: Correlation analysis for heating.

Variable	R	P-value
Heating	1	0
Heating Coil AHU 1	0.114	0
Heating Coil AHU 2	0.1012	0
Radiator State	0.896	0

Table 3: Correlation analysis for electricity.

Variable	R	P-value
Electricity	1	0
Occupancy	0.689	0
Fan Pressure AHU 1	0.760	0
Fan Pressure AHU 2	0.541	0

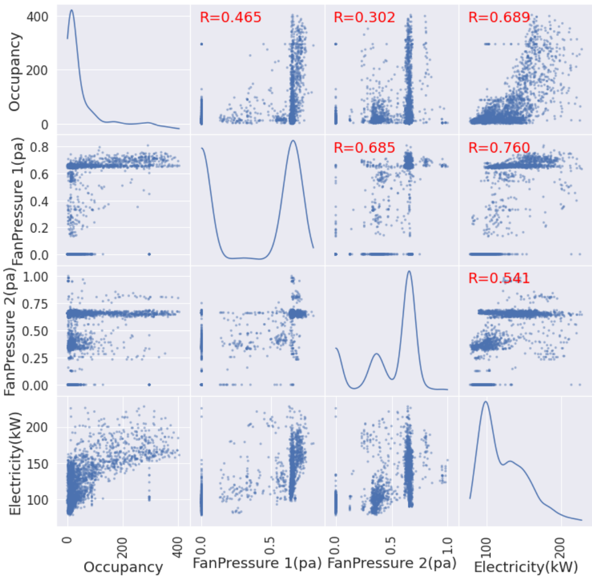

Figure 4: The scatter matrix plot of occupancy, fan pressure for AHU 1 and AHU 2 and electricity meter. The scatter matrix and R values show occupancy has a good linear relationship with electricity use. Also, the AHU 1 has more impact on electricity use compared to AHU 2.

End-use Disaggregation Model Validation and Error Calculation

For the evaluating process, Sci-kit-learn has a train and test helper function which can randomly split the collected data into training and test sets. The evaluation is performed by dividing the dataset into two

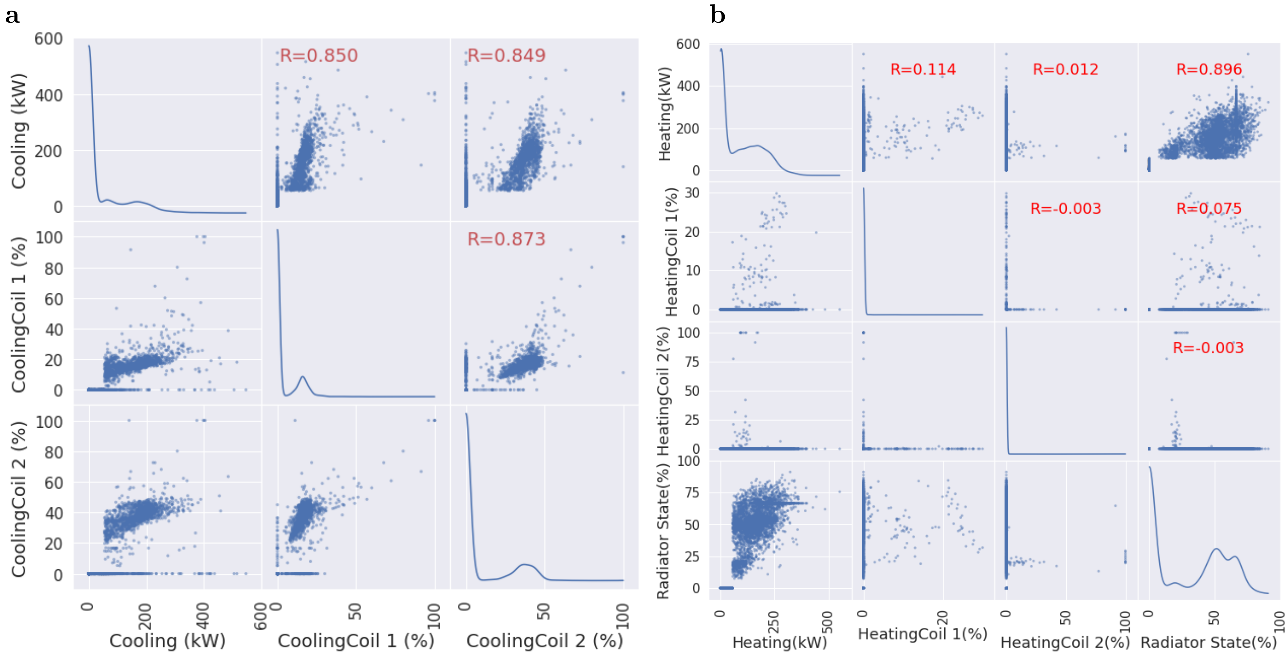

Figure 5: The scatter matrix plot of independent variables and energy meters data.The scatter matrix plot shows all pair-wise correlation between BAS data and measured energy used for (a) cooling and (b) heating. According to scatter matrix for heating energy use radiator state for zones has a strong linear relationship to heating energy use compared to heating valve position.

parts, where 20% of data is kept out of the training process. This part is used to evaluate the established model. The developed regression equation for predicting energy use for each meter and the corresponding performance criteria are presented in Table 5.The regression coefficients presented here show the impact of inputs on the hourly energy use in the building for each meter. The accuracy of the regression models was assessed using the coefficient of determination and coefficient of variance of the root mean square error (CV-RMSE).
Coefficient of determination, R^2 is the proportion of the variance in the dependent variable that is predictable from the independent variable(s). In general, the higher the R^2, the better the model performance (Fan and Ding (2019)). In Equation 3, $\hat{y}$, y and $\bar{y}$ present the predicted, measured energy and the mean of the real observation set (Kneifel and Webb (2016), Walker et al. (2020)).

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} \tag{3}$$

Coefficient of variance of the root mean square error (CV-RMSE), represents the ratio of root mean square error (RMSE) to the mean of observation, as shown in Equations 4, 5 and 6 (Taebi and Mansy (2017)). According to the ASHRAE Guideline 14 for the measurement of energy demand, CV-RMSE is a suitable tool to calculate the performance of the model for engineering applications. For hourly data predictions, a CV-RMSE value below 0.3 (= 30%) is sufficiently close to physical reality and adequate for engineering purposes (Fan et al. (2017)).

$$CV\!-\!RMSE = \frac{RMSE}{\frac{\sum_{i=1}^{N}(\bar{y}_i)}{N}} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}} \tag{5}$$

$$Mean = \frac{\sum_{i=1}^{N}(y_i)}{N} \tag{6}$$

Results and Discussion

After creating the models, the next step is to perform the prediction. The results associated with the training and testing regression models using 2019 electricity, cooling, and heating energy are shown in Table 5. It illustrated the CV-RMSE and R^2 value variation of the predictions using the trained models for each meter of the building. For CV-RMSE and R^2 the performance matrices below 0.3 and above 0.7 respectively, are chosen as the desired values as described in the previous section. Regarding the more challenging hourly, daily weekly, and monthly level, the Sci-kit-learn method shows the best performance, reaching a minimum R^2, 0.75 and accuracy of 86% for all scenarios in the electricity meter. The best performance results for cooling energy use have been reached in weekly and monthly level and for heating meter, the best performance reached the monthly level. One explanation of this behaviour is that linear regression

Table 4: Regression coefficients for electricity, heating and cooling energy.

Regression Coefficients	Electricity		Heating		Cooling	
	GA	LR	GA	LR	GA	LR
β_0	93.34	93.85	10.07	9.98	0.00	7.01
β_1	0.14	0.15	4.08	4.14	4.43	4.09
β_2	49.57	47.92	0.53	0.59	2.32	2.01
β_3	6.27	6.61	2.43	2.44	-	-

Table 5: Performance of the proposed regression models for electricity, heating and cooling demands. Green shadings indicate the best performance for hourly, daily, weekly and monthly.

Electricity								
	R^2				CV-RMSE			
Method	Hourly	Daily	Weekly	Monthly	Hourly	Daily	Weekly	Monthly
LR	0.78	0.98	0.99	0.99	0.12	0.10	0.07	0.01
GA	0.76	0.89	0.95	0.98	0.14	0.12	0.10	0.07
Heating								
LR	0.75	0.77	0.79	0.81	0.70	0.50	0.45	0.30
GA	0.76	0.79	0.80	0.81	0.70	0.51	0.46	0.31
Cooling								
LR	0.78	0.90	0.91	0.94	0.55	0.40	0.31	0.18
GA	0.77	0.89	0.90	0.94	0.60	0.39	0.31	0.19

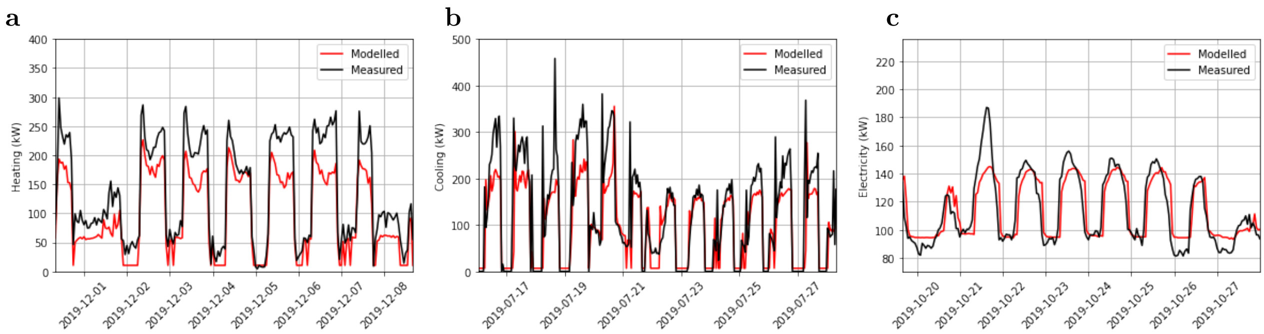

Figure 6: Comparison of observed and predicted data by the model trained by the Sci-kit-learn with hourly data for (a) Heating, (b) Cooling and (c) Electricity.

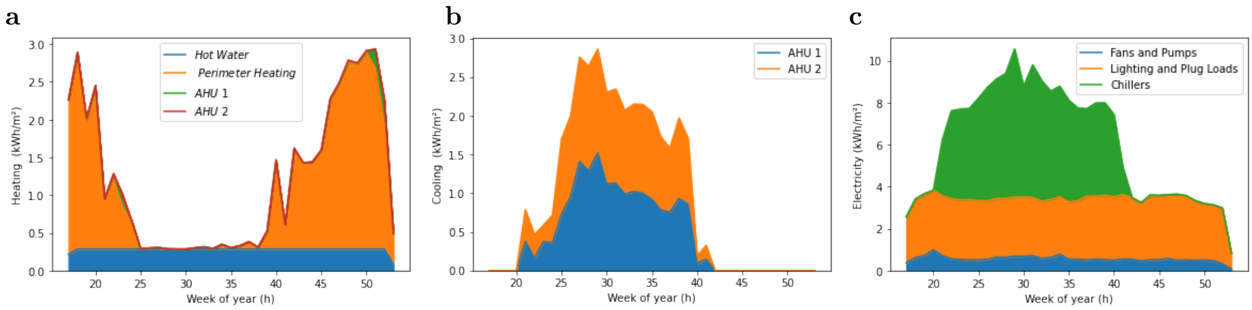

Figure 7: Weekly (a) Heating use by AHU 1, AHU 2, perimeter heaters and other; (b) Cooling energy use by chillers, AHU 1 and AHU 2; (c) Electricity use by chillers, fans and pumps, lighting and occupants predicted by the model trained by the Sci-kit-learn with hourly data.

cannot properly capture the non-linear relation between the demand and the modeled data for heating energy use since the linear correlation relationship could also explain this accuracy result. When comparing the results, the model accuracy for both methods (a linear regression model trained with the GA and Sci-kit-learn) is very similar. With R^2 assessments for all three energy meters, the desired margin was met but the CV-RMSE for electricity with high accuracy in all levels of sampling. Heating was above the target threshold stated in ASHRAE Guideline 14 (ASHRAE (2014)). One possible reason for this could be non-linear behaviour when applying the models to untrained data. CV-RMSE provides a better indication of the model's usefulness on some occasions.

Additionally, when comparing the computation times needed in creating these models, the Sci-kit-learn completed the task in less time compared to GA. Considering all these facts, the linear regression model for electricity and cooling end-use disaggregation using both methods performs better than the heating energy meter. Based on disaggregation analysis in this work, 46.1% of the total electricity was used by lighting and plug-in equipment, 44.7% was used by the chillers, and 9.2% was consumed by the fans of two AHUs. For heating, 75.6% was used for perimeter heating devices, 22.4% was used for hot water, and 17.3% consumed by the two AHUs. In addition, for cooling, 48.3% of energy consumed by AHU 1 and 51.7% was used by AHU 2, Figure(7).

Conclusion and Future Work

In this study, linear regression models that disaggregate bulk electricity, cooling, and heating meter data into end-uses of higher spatial and categorical resolution were developed. The models use BAS trend data that indicate the operational state of the major energy using equipment (e.g., AHU supply fan state). The models were trained with the Sci-kit-learn and the genetic algorithm. Their accuracy and ability to disaggregate end-uses were demonstrated upon data from an academic office building in Ottawa, Canada. For electricity, the end-uses considered were fans and pumps, lighting and plug loads, and chillers. For heating and cooling, they were AHU heating and cooling coils, reheat coils, and other perimeter heating devices.

From the results of this study, it was observed that low-frequency meter data could be disaggregated at a reasonable accuracy with provided BAS data.

Future research will expand the findings of this study to disaggregate energy flow by using BAS data with more accurate regression models at zone-levels with different commercial buildings data-sets. The coefficients of the linear regression model could give us an opinion about the impact of independent variables but that would fail for non-linear models. Developing a decision tree-based model and feature importance analysis in each node can explain the impact of inputs to predict the target values and non-linear models and help us to investigate more for disaggregation purposes which will be studied as future works. In addition, cluster decomposition property (CDP) would be a potential field of study to find the clusters that determine the strength of the correlations between each region for the selected variable to estimate more accurate regression models.

Acknowledgment

Omitted for review.

References

ASHRAE (2014). Ashrae guideline 14 measurement of energy, demand, and water savings.

Bansal, S. and M. Schmidt (2017). Energy disaggregation methods for commercial buildings using smart meter and operational data. In *AAAI Workshops*.

Batra, N., O. Parson, M. Berges, A. Singh, and A. Rogers (2014). A comparison of non-intrusive load monitoring methods for commercial and residential buildings. *arXiv preprint arXiv:1408.6595*.

Burak Gunay, H., Z. Shi, I. Wilton, and J. Bursill (2020). Disaggregation of commercial building end-uses with automation system data. *Energy and Buildings 223*, 110222.

Chen, H., P. Wang, C. Sung, Y. Yeh, and Y. Lee (2013). Energy disaggregation via clustered regression models: A case study in the convenience store. In *2013 Conference on Technologies and Applications of Artificial Intelligence*, pp. 37–42.

Darwazeh, D., B. Gunay, and J. Duquette (2019). Development of a virtual metering method for characterizing energy flows in air handling units. In *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*, pp. 24–29.

Doherty, B. and K. Trenbath (2019). Device-level plug load disaggregation in a zero energy office building and opportunities for energy savings. *Energy and Buildings 204*, 109480.

Esa, N. F., M. P. Abdullah, and M. Y. Hassan (2016). A review disaggregation method in non-intrusive appliance load monitoring. *Renewable and Sustainable Energy Reviews 66*, 163–173.

Fan, C. and Y. Ding (2019). Cooling load prediction and optimal operation of hvac systems using a multiple nonlinear regression model. *Energy and Buildings 197*, 7–17.

Fan, C., F. Xiao, and Y. Zhao (2017). A short-term building cooling load prediction method using deep learning algorithms. *Applied Energy 195*, 222 – 233.

Hobson, B., D. Lowcay, B. Gunay, A. Ashouri, and G. Newsham (2019, 05). Opportunistic occupancy-count estimation using sensor fusion: A case study. *Building and Environment 159*.

Ji, Y., P. Xu, and Y. Ye (2015). Hvac terminal hourly end-use disaggregation in commercial buildings with fourier series model. *Energy and Buildings 97*, 33 – 46.

Kaselimi, M., N. Doulamis, A. Doulamis, A. Voulodimos, and E. Protopapadakis (2019). Bayesian-optimized bidirectional lstm regression model for non-intrusive load monitoring. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2747–2751.

Klemenjak, C. and P. Goldsborough (2016). Non-intrusive load monitoring: A review and outlook. *arXiv preprint arXiv:1610.01191*.

Kneifel, J. and D. Webb (2016). Predicting energy performance of a net-zero energy building: A statistical approach. *Applied energy 178*, 468–483.

Kowalski, C. J. (1972). On the effects of non-normality on the distribution of the sample product-moment correlation coefficient. *the Royal Statistical Society 21*(11), 1–12.

Lee, S.-J., Y.-J. Kim, H.-S. Jin, S.-I. Kim, S.-Y. Ha, and S.-Y. Song (2019). Residential end-use energy estimation models in korean apartment units through multiple regression analysis. *Energies 12*(12).

Mahdavi, A., F. Tahmasebi, and M. Kayalar (2016, 08). Prediction of plug loads in office buildings: Simplified and probabilistic methods. *Energy and Buildings 129*, 322–329.

Pacific Northwest National Lab.(PNNL), Richland, WA (United States) (2015). *Characteristics and performance of existing load disaggregation technologies*.

Norford, L. K. and S. B. Leeb (1996). Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms. *Energy and Buildings 24*(1), 51 – 64.

Rafsanjani, H. N. and C. Ahn (2016). Linking building energy-load variations with occupants' energy-use behaviors in commercial buildings: Non-intrusive occupant load monitoring (niolm). *Procedia Engineering 145*, 532 – 539. ICSDEC 2016 – Integrating Data Science, Construction and Sustainability.

Rafsanjani, H. N., C. R. Ahn, and J. Chen (2018). Linking building energy consumption with occupants' energy-consuming behaviors in commercial buildings: Non-intrusive occupant load monitoring (niolm). *Energy and Buildings 172*, 317 – 327.

Rogriguez, A., S. T. Smith, A. Kiff, and B. Potter (2016). Small power load disaggregation in office buildings based on electrical signature classification. In *2016 IEEE International Energy Conference (ENERGYCON)*, pp. 1–6.

Schirmer, P. A. and I. Mporas (2019). Statistical and electrical features evaluation for electrical appliances energy disaggregation. *Sustainability 11*(11), 3222.

Schirmer, P. A., I. Mporas, and M. Paraskevas (2019). Evaluation of regression algorithms and features on the energy disaggregation task. In *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pp. 1–4. IEEE.

(2007-2020). *Cross-validation: evaluating estimator performance*. `https://scikit-learn.org/stable/modules/cross_validation.html#cross-validation`.

Shao, H., M. Marwah, and N. Ramakrishnan (2012). A temporal motif mining approach to unsupervised energy disaggregation. In *Proceedings of the 1st International Workshop on Non-Intrusive Load Monitoring, Pittsburgh, PA, USA*, Volume 7.

Shao, H., M. Marwah, and N. Ramakrishnan (2013). A temporal motif mining approach to unsupervised energy disaggregation: Applications to residential and commercial buildings. *Power (W) 250*(1700), 2250.

Solgi, R. M. (May 4, 2020). geneticalgorithm python. `https://pypi.org/project/geneticalgorithm/`.

Taebi, A. and H. A. Mansy (2017). Time-frequency distribution of seismocardiographic signals: A comparative study. *Bioengineering 4*(2), 32.

Walker, S., W. Khan, K. Katic, W. Maassen, and W. Zeiler (2020). Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings. *Energy and Buildings 209*, 109705.