2021

A Data-driven Model Development for Generalized Building Energy Predictions

Tao Cao
University of Maryland, College Park, United States of America, taocao@umd.edu

Lei Gao

Vikrant C. Aute

Yunho Hwang

A Data-driven Model Development for Generalized Building Energy Predictions

Tao Cao[1]*, Lei Gao[1], Vikrant Aute[1], Yunho Hwang[1]

[1] 4164 Glenn Martin Hall Bldg., Center for Environmental Energy Engineering
Department of Mechanical Engineering, University of Maryland,
College Park, MD., 20742, United States

*: Corresponding author: Tel: (301) 405-8672, email: taocao@umd.edu

ABSTRACT

Building energy predictions are in critical need in many fields. The conventional physic model-based approach (via EnergyPlus or similar tools) does decent work to predict energy consumptions. However, it is limited to single predefined building analysis and requires an extensive amount of time and labor to build models. Nevertheless, decision-makers usually need to quantify the energy savings of large building clusters within a short time. The thriving of big data and machine learning techniques enables predicting energy consumptions accurately for different applications within reasonable time frames. This study aims at developing data-driven models for generalized building energy predictions. The models can be used for establishing counter-factual baselines to validate the efficacy of energy-saving measures and energy production and usage planning. The former is usually for medium to long-term durations, while the latter is for short-term durations. We used real-world open data sets from ASHRAE, which covers energy consumptions of about 1,500 buildings for two years. We then preprocessed the data following the industry's standard practice. Multiple approaches of missing values imputations, outlier detections, and feature engineering were explored, based on which the best methods are suggested for building energy predictions. Gradient boosting (GB) based model has been developed for medium to long-term predictions, while the long short-term memory (LSTM) based model has been developed for short term predictions. Hyperparameter tuning was performed on model structures and parameters. We used root mean squared error (RMSE) between the predicted and true energy consumptions to evaluate performances. The results show that the GB based model achieves RMSE of 0.49 for electricity, 1.10 for chilled water, 1.25 for steam, and 1.32 for hot water. The LSTM model performs better with shorter prediction days and longer input days. However, further increasing input days beyond a week does not increase the performance. The LSTM model has about 38% lower prediction errors than the baselines, which are averages of energy consumptions from similar historical days. The study demonstrates the development process of data-driven models for general purpose building energy predictions, from data preparations, model selections, development, and evaluations.

Keywords: Building Energy Prediction, Gradient Boosting, Long Short Term Memory, Root Mean Squared Error

1. INTRODUCTION

Energy consumption of buildings, including commercial and residential sectors, keeps around 38 Quads annually, which takes nearly 40% of the total energy consumption in the past two decades (*LLNL Flow Charts*, n.d.). The energy-saving is urgent, and a large portion of consumption would dramatically improve the total energy efficiency. Therefore, high energy-efficient buildings are critical to the efficiency and sustainability of society. High energy-efficient buildings rely heavily on optimum design and control of energy conversion systems, which are both rooted in predicted energy consumption. Therefore, it is essential to build an accurate energy prediction model to capture consumption changes.

The models for building energy prediction can be categorized into physical models, data-driven models, and hybrid models that combine the previous two models. Physical models are established based on modularized building sectors and heat and mass transfer mechanisms. Nevertheless, these models become incredibly complex when considering the complicated mechanism and the coupling characteristics of every single module in buildings. Only by simulating or simplifying every subsystem can the whole system performance be evaluated. In this genre, many commercial software packages have been developed to assess the energy consumption of buildings, such as DOE-2, EnergyPlus, and BLAST. Although these tools are well-developed, they require detailed building information and environmental parameters before being fed into the simulation system. Consequently, predicting the energy consumption of any

building would require an excessive amount of time, labor resources, and knowledge from experienced experts. These requirements make a comprehensive evaluation of building energy consumption cumbersome. On the contrary to the physical models, the data-driven models are sometimes deemed as black-box. Therefore, they have no specific physical meaning, such as machine learning (ML) approaches, which only implicitly extract features from data. These methods have advantages of robustness, flexibility, and rapidity when applying to prediction tasks compared with physical models. Concisely, the building energy prediction task is a type of supervised regression in the viewpoint of the ML field. 'Supervised' means the data used for constructing a prediction model has labels, and 'regression' means the prediction values are continuous. In this genre, any algorithms that are used to deal with supervised regression in the machine learning area have the potential to be shifted to building energy prediction tasks.

With the development of ML and data science, opportunities to leverage data for building energy prediction are manifold. Among all the paradigms of ML, neural networks (NN), support vector regression, decision tree, and Gaussian process regression are the most represented ones. These techniques and algorithms have all been intensively applied in this prediction task since 2000 (Zhong et al., 2019). Moreover, recent years have witnessed a similar migration of research interest as computer science marched in terms of ML algorithms. The trend has been alternated since the enlarged power of GPU makes the rapid auto gradient of a broader and deeper NN possible in the big data era. Taking enormous advantages of that, NN based ML algorithm deep learning (DL) has been reported to continuously outperform humans in many areas. DL's capability of dealing with the high-dimensional and large-scale data structure facilitated researchers probing into energy prediction tasks by this algorithm. Consequently, many investigations drew the support of various DL based algorithms for building energy prediction (Park et al., 2016). Among all the DL algorithms, long short-term memory (LSTM) showed its capability in building energy prediction tasks due to temporal characteristics along with the NN layers (Xu, 2019).

That being said, most studies are limited to developing models based on available data and particular buildings. There have been few investigations of generalized building energy prediction using different models based on various real-world settings. For instance, quantifying energy-saving measures' efficacy requires accurate and reliable predictions of baseline energy consumptions over medium to long time durations. The energy production and usage planning, however, focuses more on accurate predictions for short time durations. One needs to develop and optimize data-driven models for different applications, given the various available algorithms. For example, The LSTM may work well with short-term predictions, while the gradient boosting (GB) framework may work better with long-term predictions.

Our objective is to develop generalized building energy prediction methods for two applications mentioned above: short-term and medium- to long-term predictions. Our models were built upon the ASHARE dataset (*ASHRAE - Great Energy Predictor III*, n.d.). The paper is organized as follows. In section 2, we present the data preparations. In particular, we investigated multiple methods for missing value imputation, outlier detection, and feature engineering for building energy predictions. In section 3, we present the two models, the GB based model for medium- to long-term predictions and LSTM based model for short-term predictions. In section 4, we present each model's results, with a focus on their performance using root mean squared error (RMSE).

2. Data Preparation

2.1 Methodology

We followed the cross industry standard process for data mining (CRISP-DM) (“Cross-Industry Standard Process for Data Mining,” 2019). CRISP-DM (*CRISP-DM Help Overview*, 2014) defines the process of data mining into six major phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. This section presents our method of exploring the raw data and preparing them for model development.

2.2 Data Description

We used the open datasets released by ASHRAE (*ASHRAE - Great Energy Predictor III*, n.d.). The dataset covers about 15,00 buildings' energy consumptions over two years. Four types of energy consumption are covered: electricity, chilled water, steam, and hot water. Not all types of energy consumption are available to all buildings. All energy consumptions are metered real-world data. The datasets are organized into three sets. The first set provides building

id (a unique integer number linked to each building), meter type (0 for electricity, 1 for chilled water, 2 for steam, 3 for hot water), time step, and meter readings. The second set contains building information, including site id (a unique number for a geographic or climate zone), building's primary usage (such as education and office), building's gross floor area, the year the building was opened, and the number of floors of the building. The last set provides weather data. For each site id, the set contains air temperature, could coverage, dew temperature, precipitation depth, sea level pressure, wind direction, and wind speed at each time step. The dataset is the most comprehensive real-world metered data on building energy consumptions we have seen so far. While the site location and building locations are not provided (possibly due to privacy concerns), it does not affect the model development. The total amount of data is about 4 gigabytes.

2.3 Data Processing

We first combined three datasets into one by joining the dataset according to building identifier and climate zone identifier. The aggregated dataset has more than ten features and 80 million entries. Next, we checked the missing values. Table 1 shows the missing counts and percentages for the ten features. Based on the feature types and missing percentages, two approaches are adopted. For those features with high missing percentages (>50%) or without significant impacts on the building energy consumptions (e.g., cloud coverage and precipitation depth), we dropped the entire feature column. For those features with low missing percentages, we used either median or linear interpolations to fill in the missing entries. The interpolations methods are suitable as all features are weather-related. They are either constant over long periods (e.g., sea level pressure) or changing linearly (e.g., temperatures).

Table 1: Missing counts and percentages of all features in the raw dataset

Feature	Missing Total	Missing Percentage (-)
Floor count	16,709,167	0.83
Year built	12,127,645	0.60
Cloud coverage	8,825,365	0.44
Precip. depth (1 hr)	3,749,023	0.19
Wind direction	1,449,048	0.07
Sea level pressure	1,231,669	0.06
Wind speed	143,676	0.01
Dew temperature	100,140	0.005
Air temperature	96,658	0.005
Meter reading	0	0

We then investigated outliers by investigating energy consumptions. Based on preliminary exploratory data analysis, there are two categories of outliers: extremely high energy readings and extensive long periods of zeros. For the first category, we found out that one building has several orders of magnitudes metered readings higher than all other buildings. As the building's features and energy consumptions are not representatives, it is removed from the dataset. For the zeros, we filtered out about 10% of them by determining whether they are true or false zeros. False zeros were when meters were not working correctly, or the facility was shut down for extensive periods (such as summer break for educational facilities). The false zeros were determined by comparing them against similar buildings at the same time.

Lastly, we performed feature engineering. Log transformation was applied to scale down metered energy readings. Further, new time features were added. We added months of the year, days of the week, hours of the day, and whether it is a holiday or not. Aside from this, we encoded cyclic features of hours, using a sine and cosine transformation, as shown in Eq. (1). Such feature transformations are widely used in time series regression problems.

$$x_{sin} = \sin\frac{2\pi x}{x_{max}}$$
$$x_{cos} = \cos\frac{2\pi x}{x_{max}} \quad (1)$$

3. MODELS

3.1 Model Overview

The model development lies in two focuses. First, we built a model for medium to long term predictions. The model is meant to establish baselines to evaluate the efficacy of energy-saving measures. The model is built upon the GB framework, using Light Gradient Boosting Machine (LightGBM) (*Welcome to LightGBM's Documentation! — LightGBM 3.1.1.99 Documentation*, n.d.). Hereafter, we refer to it as the GB model. The second model is meant to be used for energy predictions within short terms. Similarly, as the model is LSTM based, we refer to it as the LSTM model hereafter.

3.2 GB Model

As mentioned before, quantifying the efficacy of energy-saving measures requires accurate baselines. Counter-fractural models are needed to establish such baselines by assuming nothing changed in the building. These models were used for energy predictions of long periods, typically ranging from several weeks to years. The model takes building and weather information as inputs and outputs energy consumptions at corresponding timestamps. The data was first normalized by removing the mean value of each feature, then scale it by dividing non-constant features by their standard deviations. The data was then fed into the GB model. We trained one model for one type of energy consumption. For each model, we used five k-fold cross-validations (*3.1. Cross-Validation: Evaluating Estimator Performance — Scikit-Learn 0.23.2 Documentation*, n.d.) by splitting data into 80% testing set and 20% validation set. K-fold cross-validation was acceptable here as we did not shuffle data. Further, the model did not use metered energy consumptions as inputs. The final prediction results are averages of 5 models from the cross-validations.

Table 2 shows the parameter settings of the GB model. RMSE was used to evaluate the model's performance. It evaluates the overall differences between true and predicted energy consumptions. We set the early stopping criteria as no improvements of RMSE on the validation set in 20 rounds. Manual parameter tunings were performed to improve the performance over default settings.

Table 2: Parameter setting of the GB model

Parameter	Value
Boosting type	gbdt
Objective	regression
Metric	rmse
Learning rate	0.25
Feature fraction	0.9
Bagging fraction	0.8
Bagging frequency	5
Number of boost rounds	2000
Early stopping rounds	20

3.3 LSTM Model

The LSTM model aims for short-term energy predictions, contrary to GB models. It is meant to be used for energy production and usage planning within short-term, from hours to days at most. The LSTM network has been widely used for similar time-series predictions ("Long Short-Term Memory," 2020).

Figure 1 shows the overall flow chart of the LSTM model. The LSTM model takes historical energy consumptions and other features as input and output energy consumptions of the immediate future time steps. Hence, historical energy consumptions are critical inputs. For buildings without historical energy consumptions (such as new constructions), we proposed constructing a historical profile based on the most similar buildings from the 1500 building pool. The similarities are measured in Gower distance (*Gower.Dist Function | R Documentation*, n.d.), using building information including footprint, building types, year built, and floor count.

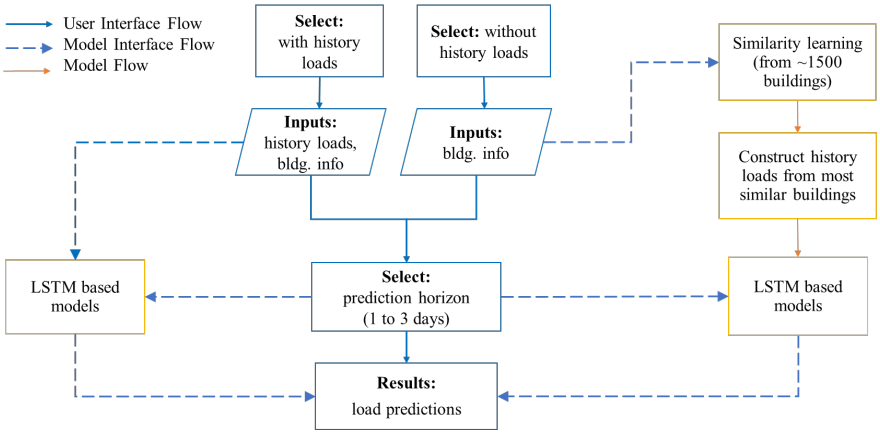


Figure 1: Flow chart of the LSTM model

The LSTM model is unique to each building. For each building, we used the first ten months of data as the training set and the remaining two months of data as the test set. The model accepts various historical hours to define the training set and different numbers of future hours to define the test set. For example, an LSTM model of 24 hours of inputs and 3 hours of outputs indicates the following. The inputs are energy consumptions and all other features (such as weather) in the past 24 hours, while the outputs are energy consumptions in the next three hours.

We conducted the hyperparameter tuning of the LSTM model using Keras Tuner (*Keras Tuner*, n.d.). Table 3 shows the parameter settings. The LSTM model structure is flexible with various layer combinations, optimizer learning rates, and early stopping rounds. The max epochs, batch size, and optimizer is fixed. We used the default loss function of mean squared error (MSE).

Table 3: Parameter setting of the LSTM model tuner

Parameter	Value (Range)
Layer	4~8
LSTM layer units	32 to 128 with a step of 16
Dropout rate	0.2 to 0.5 with a step of 0.1
Max epochs	1000
Batch size	32
Optimizer	Adam
Optimizer learning rate	1e-2, 1e-3, 1e-4
Loss function	mse
Early stopping rounds	20, 40, 60

Note that the Keras Tuner only performs preliminary training and validation. It behaves in such a manner to speed up the parameter tuning process. Therefore, we continued to train the model with the parameters selected by the Keras Tuner. The models were further trained with Keras from Tensorflow (Team, n.d.).

4. RESULTS

4.1 GB Model Results

The GB model was trained on the python 3.7 environments, on a windows machine with eight cores and 16 gigabytes of memory. The training process took about 30 minutes. Figure 2 shows the comparison between the true and predicted energy consumptions of four models for four meter types, with 0 for electricity, 1 for chilled water, 2 for steam, and 3 for hot water. The duration is an entire year. The energy readings are log-transformed. It is observed that the predictions agreed with the true values very well. The predictions have smoother trends than true values, especially when there are abrupt jumps. There could be several real-life operation mechanisms, which are difficult for the model to learn and capture. The power meter might have reading jumps when there were unstable voltage supplies. Also, the

building may have occasional significant occupant schedule changes. These might be some of the reasons contributing to deviations when there were abrupt jumps.

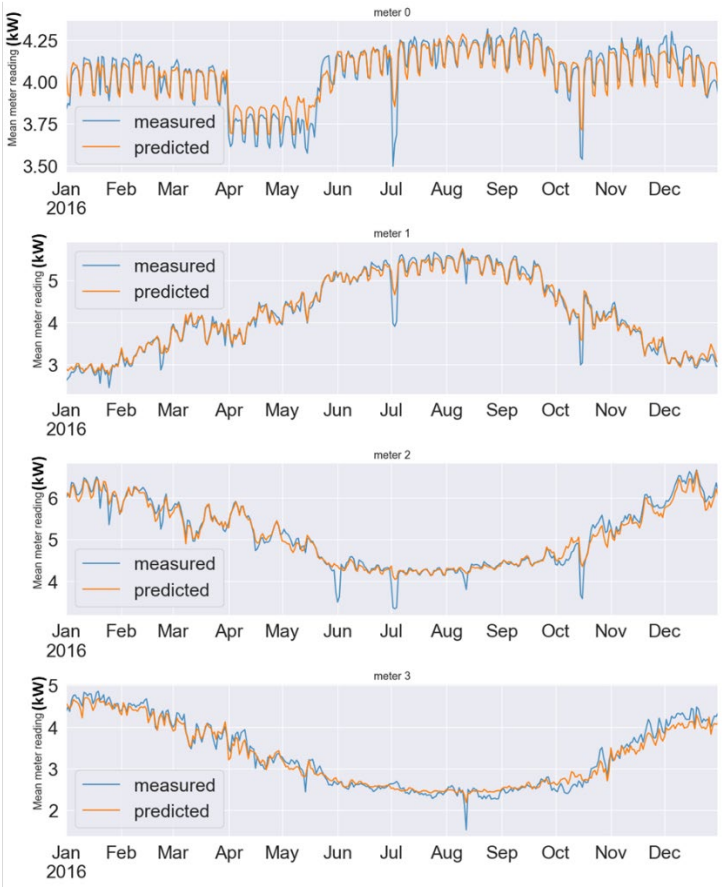


Figure 2: GB model predictions of four energy types

As mentioned in section 3, we used RMSE to evaluate model performance. Based on Figure 2, the models have an RMSE of 0.49 for electricity, 1.10 for chilled water, 1.25 for steam, and 1.32 for hot water. With more than 50% of entries are for electricity, we expect a significantly lower RMSE for electricity predictions.

4.2 LSTM Model Results

For the LSTM model development, we focused on the electricity data of several representative buildings for demonstration purposes. The model is trained on the python 3.7 environments, on a windows machine with eight cores and 16 gigabytes of memory. The training process took about 30 minutes for one building, including about 15 minutes for hyperparameter tuning. We monitored the losses of the training and validation set and determined the early stopping criteria to prevent overfitting. As an example (as shown in Figure 3), we stopped training until the validation set losses were not decreasing for minimal (e.g., 20) epochs. The model which presented minimal loss in the last 20 rounds was adopted. From Figure 3, epoch 87 was selected.

As mentioned in section 3.3, each model has the flexibility of various input and output periods. The results are presented by taking a look at the impacts of both variations. Figure 4 shows the prediction results of various prediction periods (1 to 3 days), with the same input periods (5 days). The red dots show the LSTM model's predictions, while the blue dots represent the baseline's predictions. The baseline is a simple average of similar three historical days of the week. We have observed that the LSTM model not only outperforms the baseline consistently but also agrees with the true values very well. The model's prediction starts to deviate more from the true further down the prediction period. It is expected as the model has less true historical information to rely on further into the future.

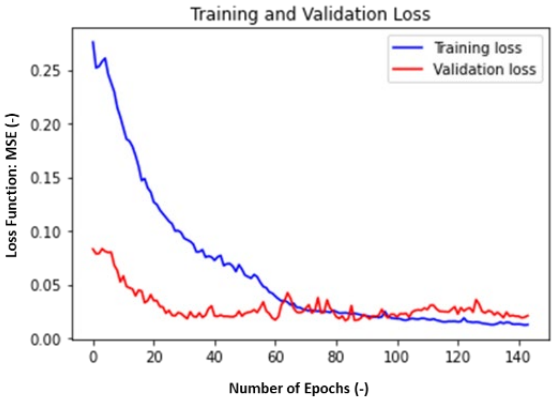
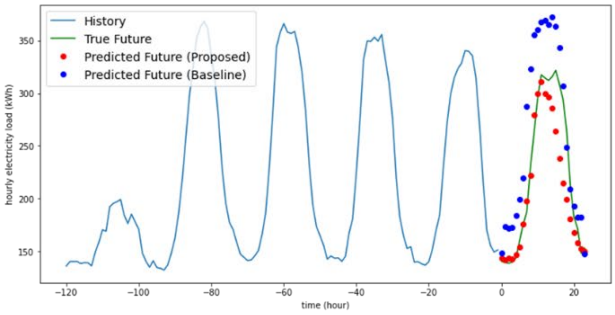
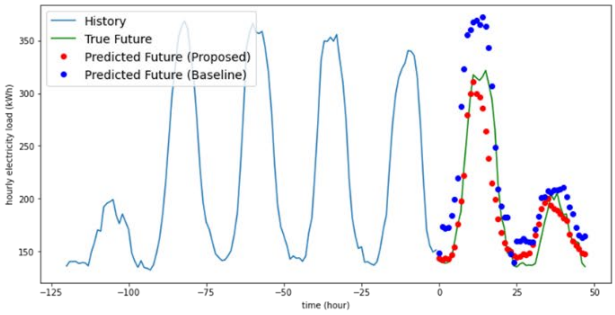


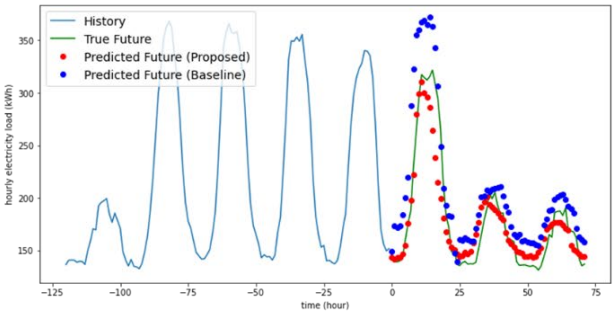
Figure 3: LSTM model loss function versus training epochs



(a) 5 days to 1 day



(a) 5 days to 2 days



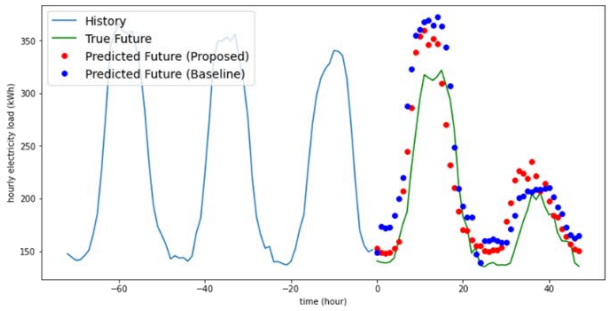
(a) 5 days to 3 days

Figure 4: LSTM model results with various prediction periods

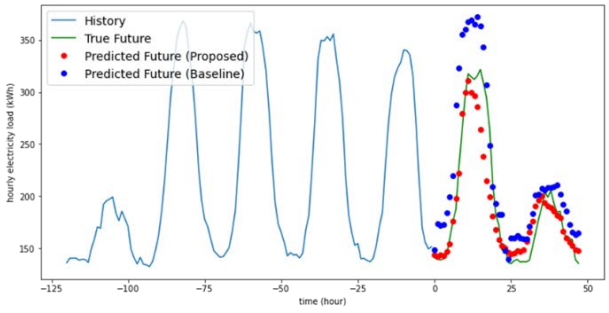
The model achieved RMSE of 30.5, 23.5, and 20.5 for 1 day, 2 days, and 3 days prediction, respectively. The baseline has RMSE of 48.6, 37.7 and 33.2 for similar prediction days. Hence the LSTM model has about 38% lower prediction errors than the baseline. The lower values for the longer day predictions are due to how RMSE is calculated. Deviations at higher absolute values contribute more to higher RMSE.

Figure 5 presents the prediction results of various input periods (3 to 7 days), with the same prediction periods (2 days). The LSTM model performs better with increasing input periods, as observed from better agreements between the red dots and green line. Also, the model generally performs better than the baselines, with similar performances under three input days. The LSTM model achieved RMSE of 33.1, 23.5, 22.6 for 3 days, 5 days, and 7 input days. It is worth noting that increasing input days does not necessarily result in better model performance. We further increased the input days up to 14 days but did not observe increased model performance. The LSTM model performed worse when the input days approached 14 days. LSTM has been known to struggle to work with extensive long sequences.

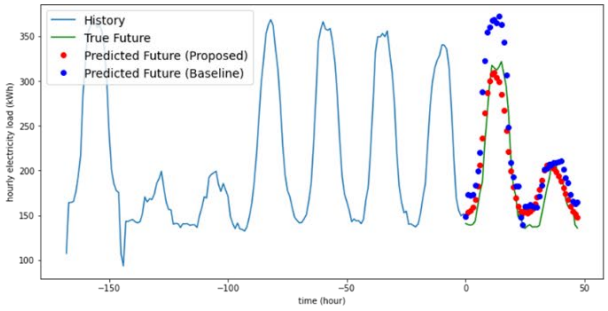
It is still an open research topic (Trinh et al., 2018). Regarding the LSTM model's structure, we have found 4 to 6 layers are enough for all cases. The LSTM layer's units usually range from 64 to 128. The drop out rate is mostly 0.2.



(a) 3 days to 2 days



(a) 5 days to 2 days



(a) 7 days to 2 days

Figure 5: LSTM model results with various historical periods

4. CONCLUSIONS

We have developed data-driven models for general building energy predictions using metered energy consumptions of about 1,500 buildings for two years. The data covers eight typical climate regions and a dozen types of buildings. The data is preprocessed accordingly to the industry's standard practice. In particular, we performed outlier detections to handle excessive false zeros in energy consumptions. New features were added to handle time series related predictions. We have developed two models for two types of applications. The GB model is suited for medium- to long-term predictions, used to establish counter-factual baselines to evaluate energy-saving measures' efficacy. The LSTM model is meant for short-term predictions of energy production and usage planning. RMSE is used to evaluate models' performances. For the GB model, we performed parameter tuning and used 5 fold cross-validations. The model achieved an RMSE of 0.49 for electricity, 1.10 for chilled water, 1.25 for steam, and 1.32 for hot water predictions (after log-transformed) over an entire year. For the LSTM model, we have developed the model to be flexible for buildings with or without historical metered energy consumptions. The model is capable of generating historical profiles from similar buildings. Further, the model is flexible with various input days and prediction days. We have further tuned the model structure and parameters. Results show that the LSTM model performs better with shorter prediction days and longer input days. Though further increasing input days beyond a week does not increase the performance. The LSTM model has about 38% lower prediction errors than the baselines, which are averages of energy consumptions from similar historical days.

NOMENCLATURE

ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
CRISP-DM	cross industry standard process for data mining
DL	deep learning
GB	gradient boosting
LightGBM	Light Gradient Boosting Machine

LSTM	long short term memory
ML	machine learning
MSE	mean squared error
NN	neural network
RMSE	root mean squared error

REFERENCES

3.1. Cross-validation: Evaluating estimator performance—Scikit-learn 0.23.2 documentation. (n.d.). Retrieved December 15, 2020, from https://scikit-learn.org/stable/modules/cross_validation.html

ASHRAE - Great Energy Predictor III. (n.d.). Retrieved March 31, 2020, from https://kaggle.com/c/ashrae-energy-prediction

CRISP-DM Help Overview. (2014, October 24). www.ibm.com/support/knowledgecenter/ss3ra7_15.0.0/com.ibm.spss.crispdm.help/crisp_overview.htm

Cross-industry standard process for data mining. (2019). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Cross-industry_standard_process_for_data_mining&oldid=930958276

Gower.dist function | R Documentation. (n.d.). Retrieved December 15, 2020, from https://www.rdocumentation.org/packages/StatMatch/versions/1.2.0/topics/gower.dist

Keras Tuner. (n.d.). Retrieved December 15, 2020, from https://keras-team.github.io/keras-tuner/

LLNL Flow Charts. (n.d.). Retrieved December 16, 2020, from https://flowcharts.llnl.gov/

Long short-term memory. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Long_short-term_memory&oldid=993897583

Park, H. S., Lee, M., Kang, H., Hong, T., & Jeong, J. (2016). Development of a new energy benchmark for improving the operational rating system of office buildings using various data-mining techniques. *Applied Energy*, *173*, 225–237. https://doi.org/10.1016/j.apenergy.2016.04.035

Team, K. (n.d.). *Keras documentation: About Keras*. Retrieved December 15, 2020, from https://keras.io/about/

Trinh, T. H., Dai, A. M., Luong, M.-T., & Le, Q. V. (2018). Learning Longer-term Dependencies in RNNs with Auxiliary Losses. *ArXiv:1803.00144 [Cs, Stat]*. http://arxiv.org/abs/1803.00144

Welcome to LightGBM's documentation! —LightGBM 3.1.1.99 documentation. (n.d.). Retrieved December 15, 2020, from https://lightgbm.readthedocs.io/en/latest/index.html

Xu, L. (2019). Probabilistic load forecasting for buildings considering weather forecasting uncertainty and uncertain peak load. *Applied Energy*, 16.

Zhong, H., Wang, J., Jia, H., Mu, Y., & Lv, S. (2019). Vector field-based support vector regression for building energy consumption prediction. *Applied Energy*, *242*, 403–414. https://doi.org/10.1016/j.apenergy.2019.03.078

ACKNOWLEDGEMENT

We gratefully acknowledge the support of the Center for Environmental Energy Engineering (CEEE) at the University of Maryland.