Harnessing Machine Learning for Predicting Consumption

Jessica Thompson, Tetra Tech, Madison, WI
Jonathan Hoechst, Tetra Tech, Madison, WI

ABSTRACT

When analyzing customer billing records to estimate energy savings achieved through an efficient retrofit or installation, the gold standard has long been examining a full year of consumption data pre- and post-installation. This creates a considerable lag between the installation of the equipment and the ability to estimate savings and generate results. While in-situ metering can significantly reduce the time required before analysis, the method proves excessively expensive when analyzing results across entire energy efficiency programs or many individual sites. However, machine learning and advanced computing provide the potential to address this problem.

The results of this analysis will interest utilities, energy efficiency implementers, and evaluators. Utilities can harness machine learning and predictive analytics for load forecasting, both for longer-term integrated resource planning and the immediate effects of calling demand response events during times of peak system load. Implementers and evaluators could seize the possibilities of reducing the amount of post-installation data needed to accurately estimate energy savings produced through retrofits and other improvements, reducing costs of evaluations and implementation while providing faster feedback on projects.

Introduction

There are many existing incentive programs on the state, local and utility-level. These programs are big business, averaging $24 per customer in 2016 (EIA 2018). Given the amount of money involved, verification of these program results is of major importance. Results of these verification calculations are used to adjust the program parameters and allow program funders to get more energy savings per dollar spent. The International Performance Measurement and Verification Protocol (IPMVP) provides guidance on methods and best practices for these program evaluations (IPMVP 2002). According to the IPMVP, one or more full years of energy use and weather data should be used to construct regression models. Usually, to quantify savings due to an energy conservation measure, a full year of pre-installation and post-installation consumption data is necessary. This data is then weather-normalized and aggregated to a yearly level for comparison. This technique compares the known pre-installation consumption with the known post-installation consumption while accounting for differences in the weather and results in a lag of at least a year post-installation before verification can be run. If the verification indicates substantial differences in savings compared to the expected savings, the program administrator would prefer to know that they need to adjust their incentives as soon as possible. Given the scale of these incentive programs, there could be a significant impact.

Conventional physics-based models for energy consumption require many specific parameters and may be expensive to customize (Akbar 2020). Data-driven models can be based on a statistical approach or a machine learning algorithm. At a building level, various machine learning models have outperformed both physics-based and statistical data-driven models (Runge 2019). There has been research on using machine learning methods to predict energy consumption, showing promise for various models, including artificial neural networks (Hamdoun 2021). Machine learning models are helpful for energy consumption data because they do not require knowledge of the statistical distribution of the data. The assumption is that past observations include intrinsic patterns that will train the model to predict future values, implicitly handling trends, seasonality, and other data patterns. Many studies have been done on modeling commercial consumption data, while fewer specifically look at residential consumption (Román-Portabales 2021). These commercial studies show several algorithms with promise, including decision trees, random forest models, gradient-boosting machines, recurrent neural networks, and k-nearest neighbors (Seyedzadeh 2018). Among residential buildings, least squares support vector machines outperformed other algorithms (Edwards 2012). However, this result did rely on more detailed sensor data than is generally available. If these forecasting models perform well enough, it seems likely they can be used to reduce the amount of pre- and post-installation consumption data required for an energy conservation measure savings verification. With an increase in granularity of consumption data, it is also likely that less training data may be needed to fit an acceptable forecast model.

Scope

This study will explore the usage of machine learning models, using varying amounts of training data for two very different data granularities. Monthly billing records are still the most commonly available data used for these evaluations, but smart meter data is increasingly available. First, the data will be split into training and test portions, with the training datasets used to calibrate machine learning models. Given two years of these records, we will explore the accuracy of forecasting the second year of data based on models trained on the first year. Subsequently, we will reduce the number of months used to train the model. The calibrated model will then be used to predict the remaining year (or more) of consumption and the outcome compared to the test data. This will allow the researchers to evaluate the reductions (if any) in predictive power as the amount of data in the training dataset is reduced. Further, comparing the results across both high interval data (i.e., smart meter data) and monthly meter readings will provide insight into the additional predictive power provided by high volume, high-velocity energy use data. All consumption data (training and test) will be weather-normalized prior to modeling, so the predicted data is also weather-normalized. Because we won't know the weather for predicted data, this is the best way to account for its effect.

A critical facet of the analysis is nearly two years of consumption data in both analysis sets. This allows the trained model to estimate energy use and compare the predicted consumption against what indeed occurred – we know the "correct" answer the model should produce, but how close can machine learning bring us to predicting the future? Two machine learning models will be tested, a Random Forest regression and a Long Short-term Memory recurrent neural network model.

Methods

The evaluators trained machine learning algorithms on two separate datasets. The first dataset contains 24 months of monthly residential billing records for a southern utility. For utilities without smart meter data, this is a common granularity of available data. The dataset was imported and examined for evidence of outliers or erroneous values. Meters were removed if they had less than two full years of good records or were outside the geographic area of interest. Data was also weather-normalized, with each meter matched to an ASOS station[1]. The data for those ASOS stations were used to calculate daily heating degree days based on 65 degrees Fahrenheit and cooling degree days based on 75 degrees Fahrenheit. Heating degree days and cooling degree days were then used to calculate kWh per degree day for each record in the consumption dataset. The figures below show median monthly normalized kWh values and the distribution of readings for the modeled meters.

[1] The Automated Surface Observing System (ASOS) program is a joint effort between the National Weather Service, the Federal Aviation Administration, and the Department of Defense. ASOS serves as the nation's primary surface weather observing network.

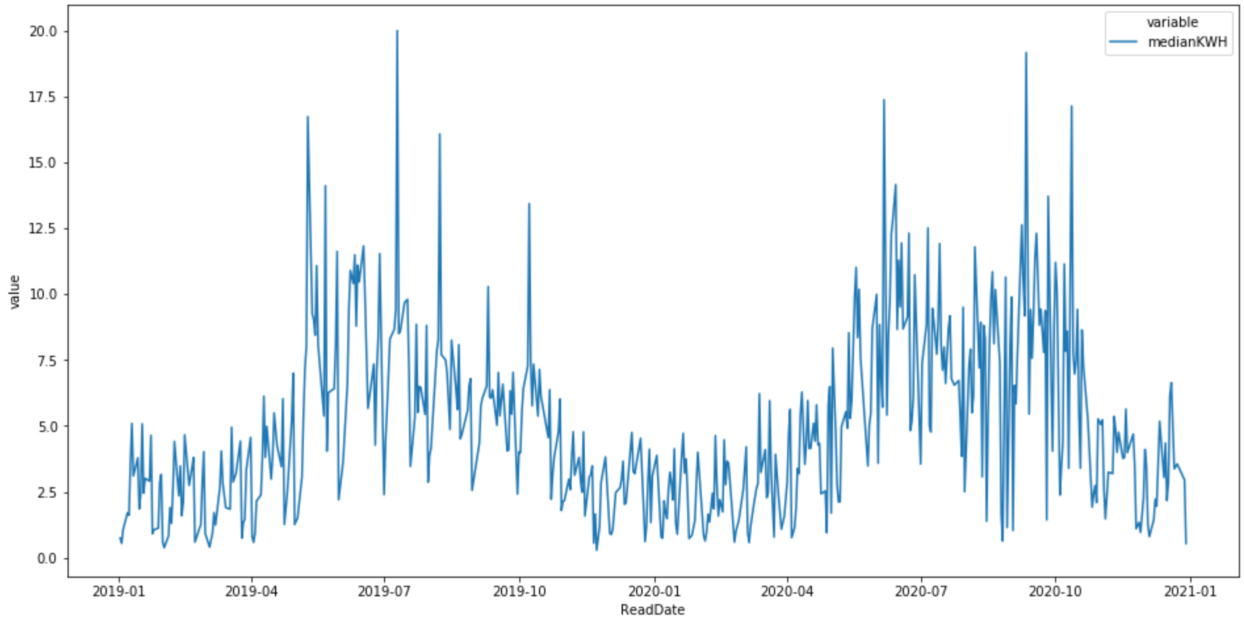

Figure 1. Median Normalized kWh for Monthly Meter Model Set

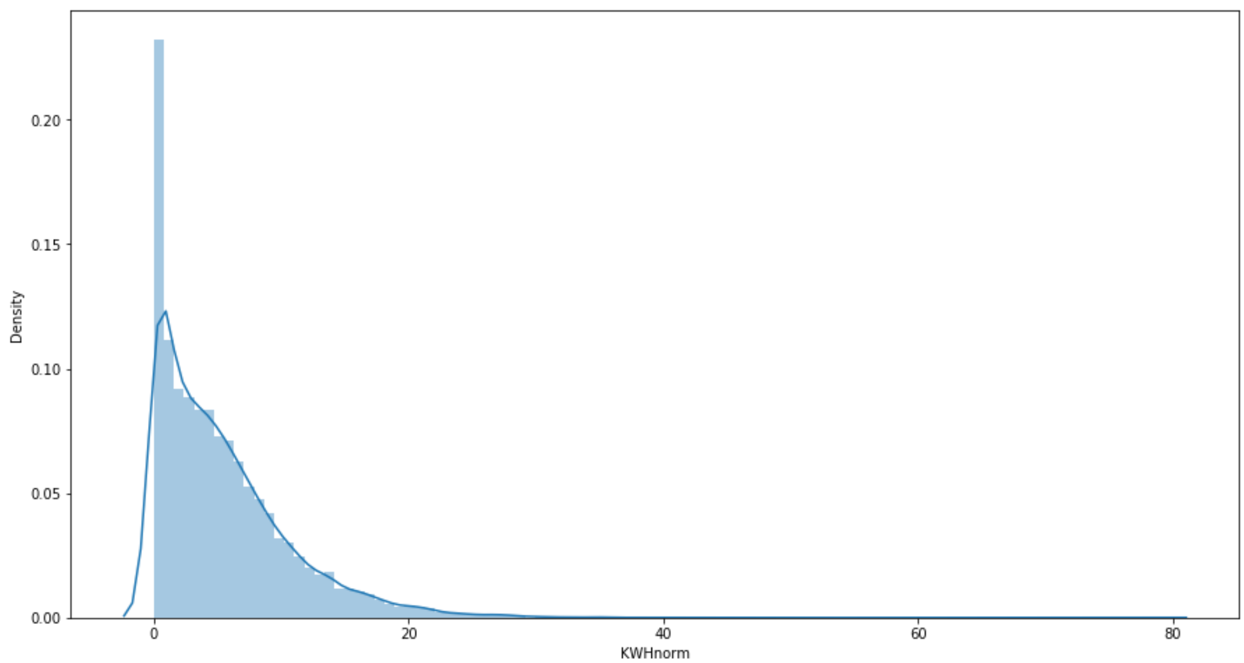

Figure 2. Distribution of Monthly Normalized kWh for Modeled Meter Set

Mean	Std Dev	Min	25th Perc	50th Perc	75th Perc	Max
5.5	5.3	0	1.4	4.2	7.9	79

Table 1. Characteristics of Model Set Monthly Normalized kWh Readings

The second dataset contained twenty months of 15-minute interval data from residential meters in a separate southern utility. Meters were removed from the dataset if they were missing readings during their range of data. This data was weather-normalized following the same procedure used for the monthly billing data but on an hourly granularity (due to the availability of ASOS weather data). Heating and cooling degree days were calculated using the same setpoints used to calculate the normalized kWh per degree day for each record in the consumption dataset.

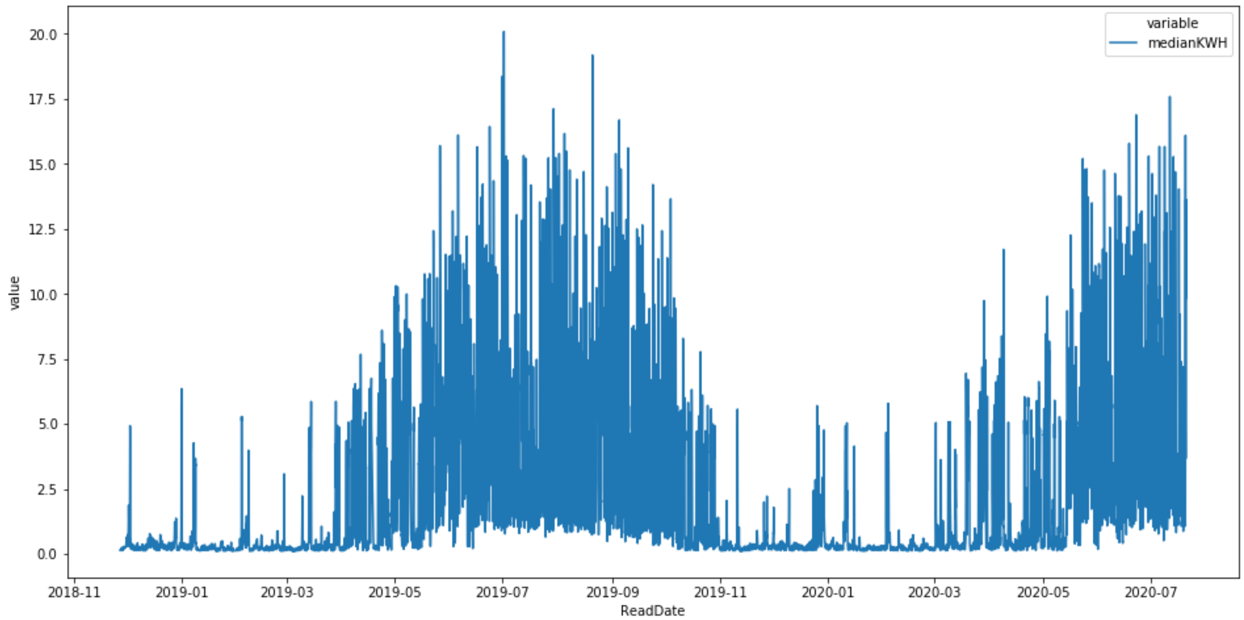

Figure 3. Median Normalized kWh for Interval Meter Model Set

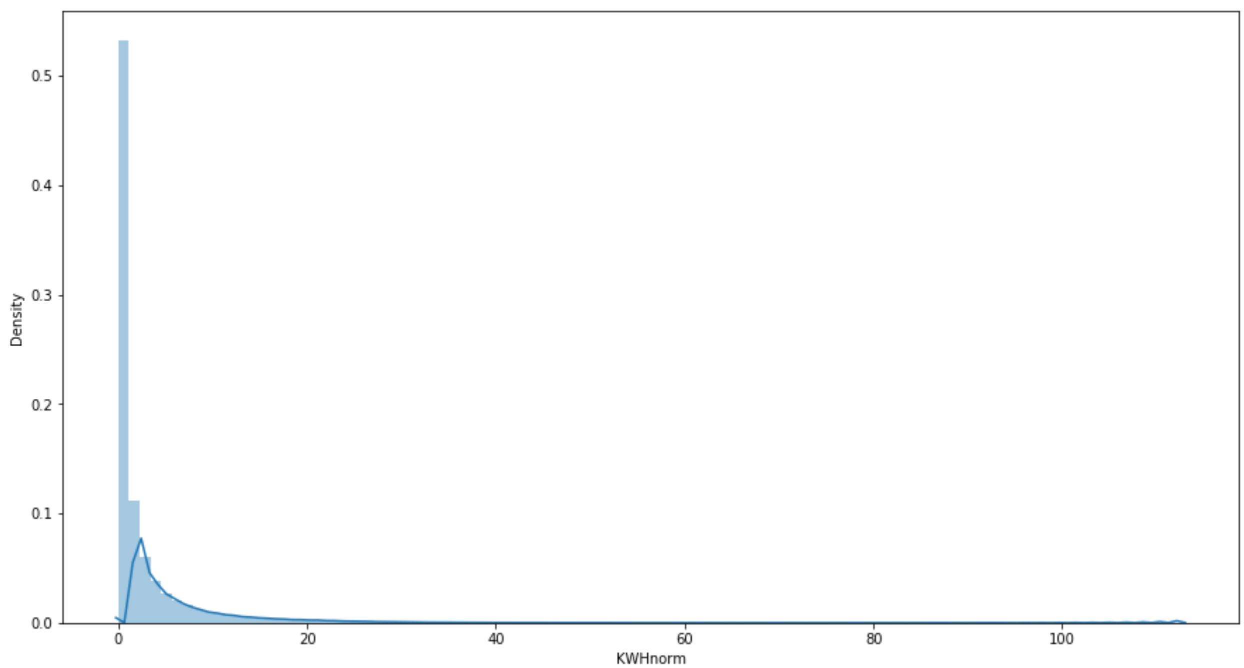

Figure 4. Distribution of Normalized kWh for Interval Modeled Meter Set

Mean	Std Dev	Min	25th Perc	50th Perc	75th Perc	Max
2.6	4.9	0	0.22	0.66	2.6	110

Table 2. Characteristics of Interval Meter Model Set Normalized kWh Readings

Model Calibration

The first type of model tested was a Long Short-term Memory (LSTM) model based on two lagged values. For each meter, the data had two lagged values added to enable a supervised model. The data was then split into test and training datasets, based on the number of months to use for training. The test and training datasets were then scaled to [-1,1] range, and a subset of meters was used to determine the best values for epochs and neurons. For an LSTM model, an epoch represents a complete pass through the training dataset, while the number of neurons indicates the number of memory cells in the hidden layer of the model. These inputs were used to fit the model, optimizing mean squared error. These tuning runs were performed on both the monthly and the interval data. The monthly data showed minimal effect from changes in the number of epochs or neurons. Based on this finding and the interval data tuning results shown below, the model was set to use 1 neuron and 100 epochs. The second type of model was a Random Forest regression, also using two lagged values as features.

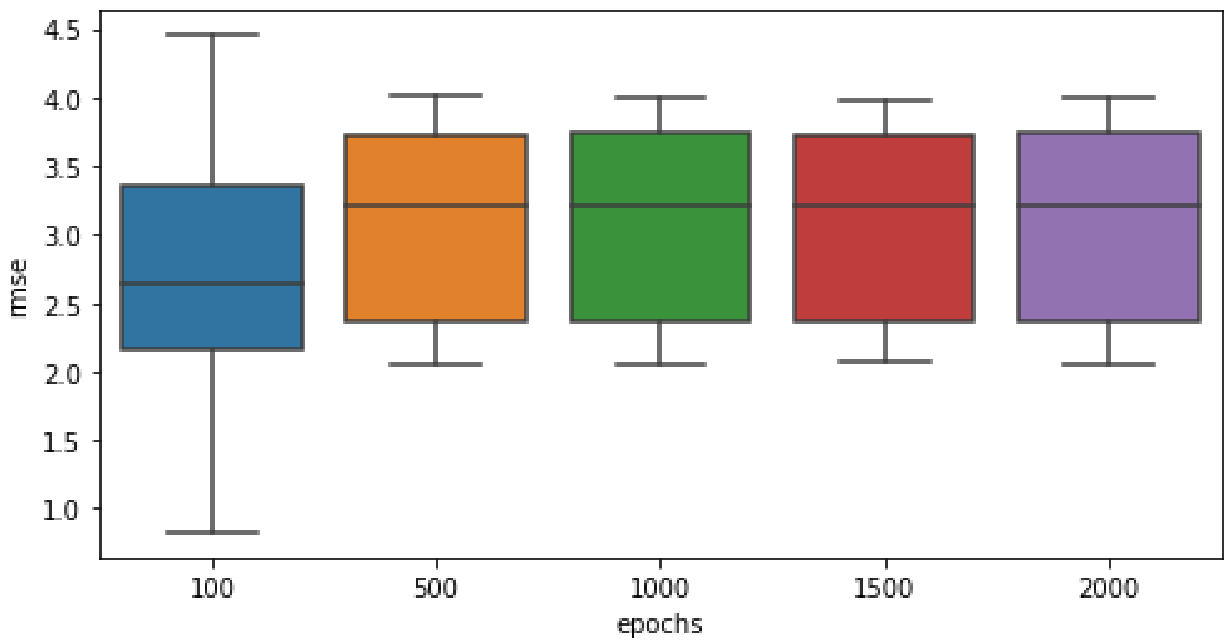

Figure 5. Root mean squared error for models grouped by number of epochs for Interval Modeled Meter Set

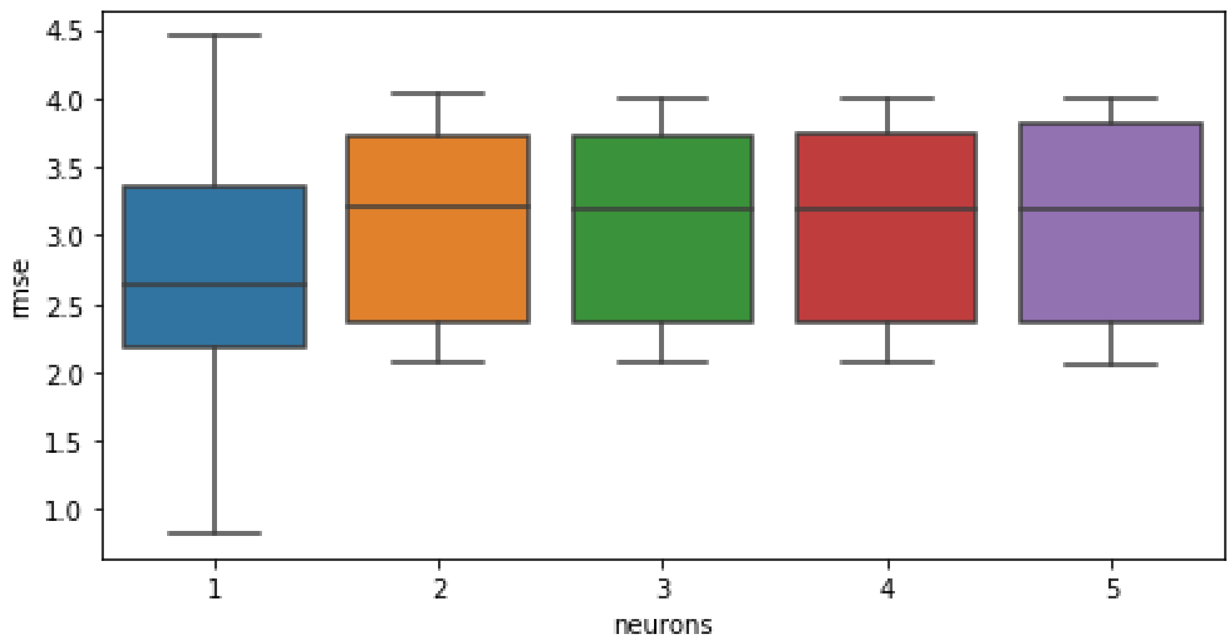

Figure 6. Root mean squared error for models grouped by number of neurons for Interval Modeled Meter Set

Models were fitted to meters in the monthly, monthly disaggregated to daily, and interval data sets for both types. The fitted models were then used to forecast predictions for comparison with the actual values. The root-mean-squared error (RMSE), mean bias error (MBE), r-squared, and mean absolute percentage error (MAPE) were recorded for each model run. This method was repeated, varying the number of months used to train the model.

Model Results

For the LSTM models, we can see that the fit improves with more months of training, as expected. We also see that the monthly disaggregated data performs the poorest of the three.

Table 3. Characteristics of LSTM Models Grouped by input data type and months of training

Model data	Months Training	Coefficient of Variation	Mean Bias Error	Mean Absolute Percentage Error	RMSE	R-Squared
Monthly	6	80	-13	-55	2.0	0.39
Monthly	10	73	-3.6	-38	1.9	0.54
Monthly	12	68	1.8	-29	2.0	0.52
Monthly disaggregated	6	120	-0.11	-58	2.9	0.38
Monthly disaggregated	10	120	3.5	-65	3.0	0.39
Monthly disaggregated	12	110	7.8	-56	2.9	0.44

Interval hourly	6	96	9.7	-23	1.8	0.66
Interval hourly	10	101	0.32	-45	1.7	0.68
Interval hourly	12	112	-3.3	-57	1.6	0.68

For the interval data, we can see that the model fit does not decline much with fewer months of training. Is the fit good enough for our purposes? In this case, what we are interested in is not the accuracy of individual kWh predictions but the prediction of overall energy use in the entire time period. When we sum the kWh for the forecast period on a per-meter basis, the plot of actual versus predicted values fits quite well for all models.

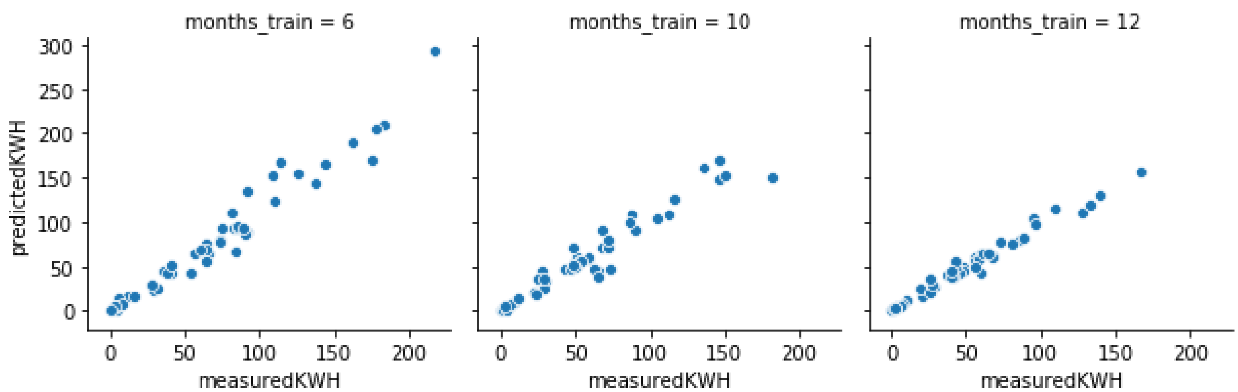

Figure 7. Plots of summed raw versus predicted kWh values for the monthly billing data, grouped by months of training

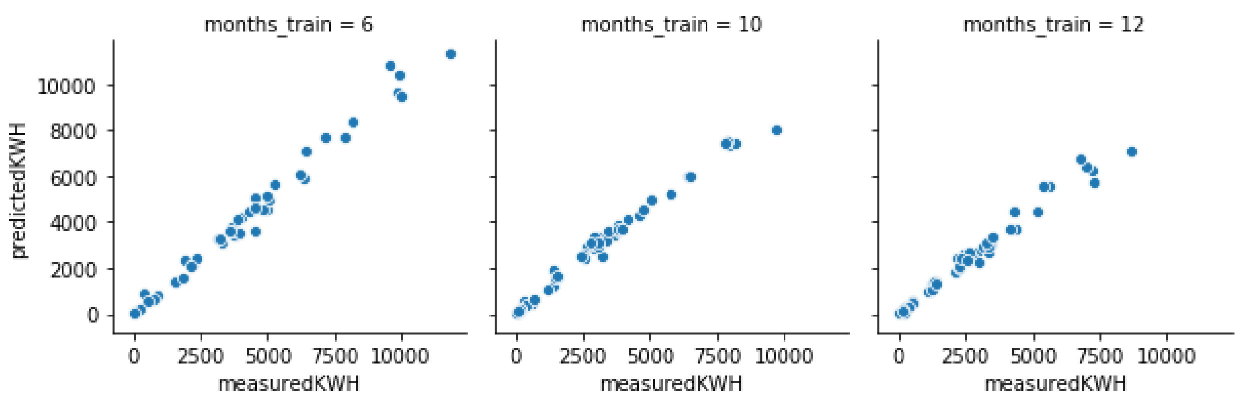

Figure 8. Plots of summed raw versus predicted kWh values for the monthly disaggregated billing data, grouped by months of training

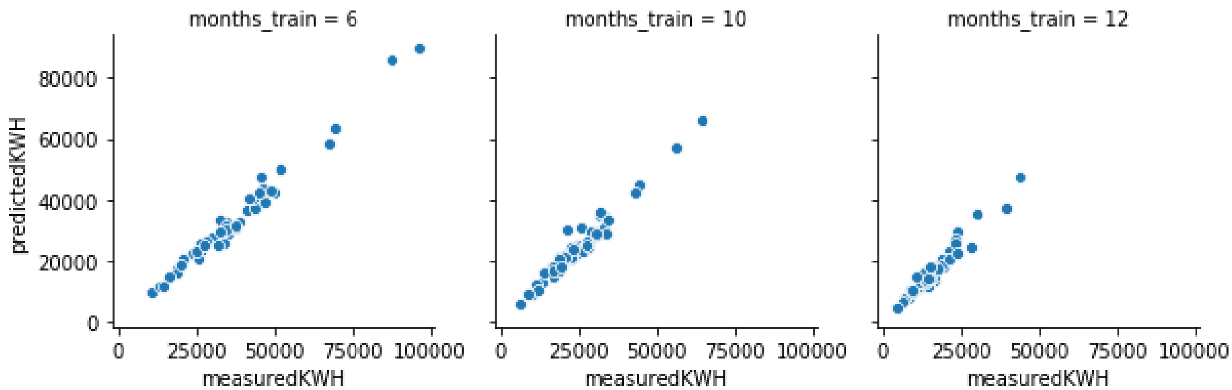

Figure 9. Plots of summed raw versus predicted kWh values for the interval data, grouped by months of training

Examining the modeled interval data, for a given meter, the percentage of difference in the totaled measured kWh and predicted kWh (100*normalized residual) is usually less than 10 percent. There are some meters with significant differences; however, that would be unacceptable as estimates of savings.

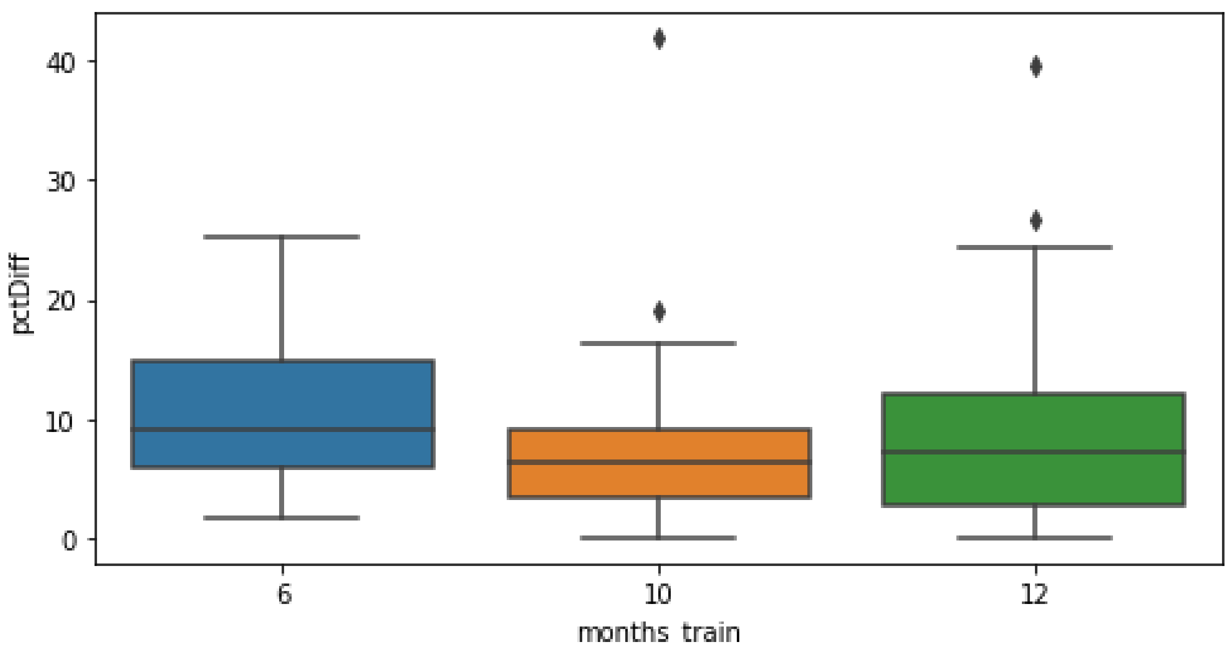

Figure 10. Plots percent difference between total measured kWh and predicted kWh, grouped by months of training

There is not a strong correlation between premises with large percent differences and total measured kWh for the prediction period. It is possible they have other commonalities that would allow for more confidence in forecasting for prediction.

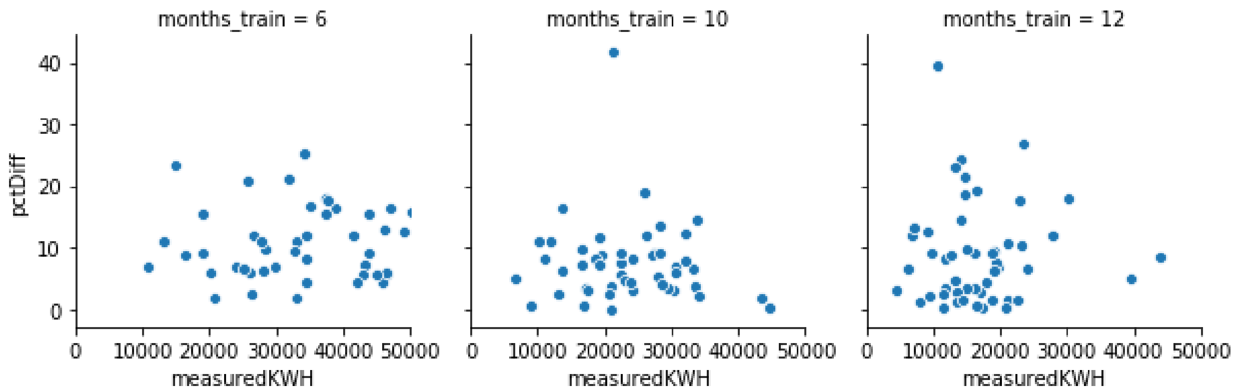

Figure 11. Plots of pct difference vs. total measures kWh for the interval data, grouped by months of training

The second type of model tested was a Random Forest regression. We can see that the fit improves with more months of training for these models, as expected. Overall, the random forest regression performed similarly to the LSTM method.

Table 4. Characteristics of random forest models grouped by input data type and months of training

Model data	Months Training	Coefficient of Variation	Mean Bias Error	Mean Absolute Percentage Error	RMSE	R-Squared
Monthly	6	91	51	59	2.3	0.02
Monthly	10	72	-20	102	2.0	0.45
Monthly	12	61	2.6	69	1.9	0.59
Monthly disaggregated	6	118	-0.6	82	3.2	0.27
Monthly disaggregated	10	119	4.4	85	3.2	0.34
Monthly disaggregated	12	109	6.9	77	3.2	0.39
Interval hourly	6	105	3.8	69	1.8	0.63
Interval hourly	10	117	-2.8	79	1.6	0.67
Interval hourly	12	113	-1.1	75	1.6	0.67

Like the LSTM model results, the percentage of difference in the totaled measured kWh and predicted kWh is usually less than 10%. There are, however, a larger portion of outlier percentage difference values using random forest regression.

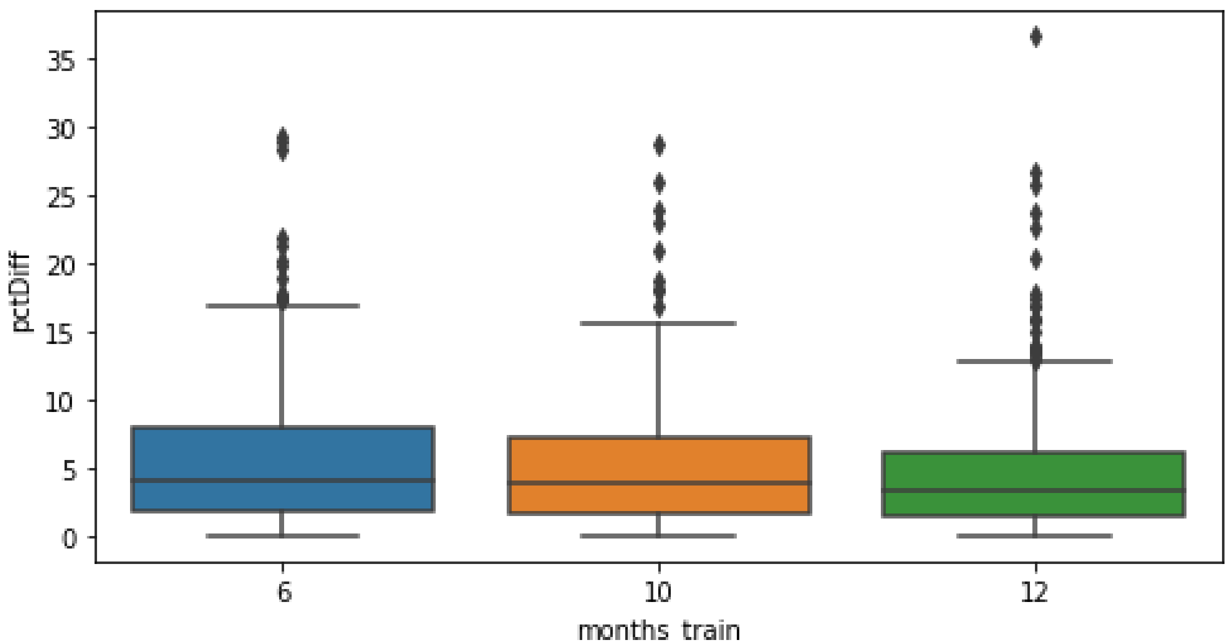

Figure 12. Plots percent difference between total measured kWh and predicted kWh, grouped by months of training

The percent differences for random forest models using the hourly interval data do not show any correlation with total measured kWh. It would be worthwhile to explore other possible correlations that may predict which meters may be forecast with high confidence.

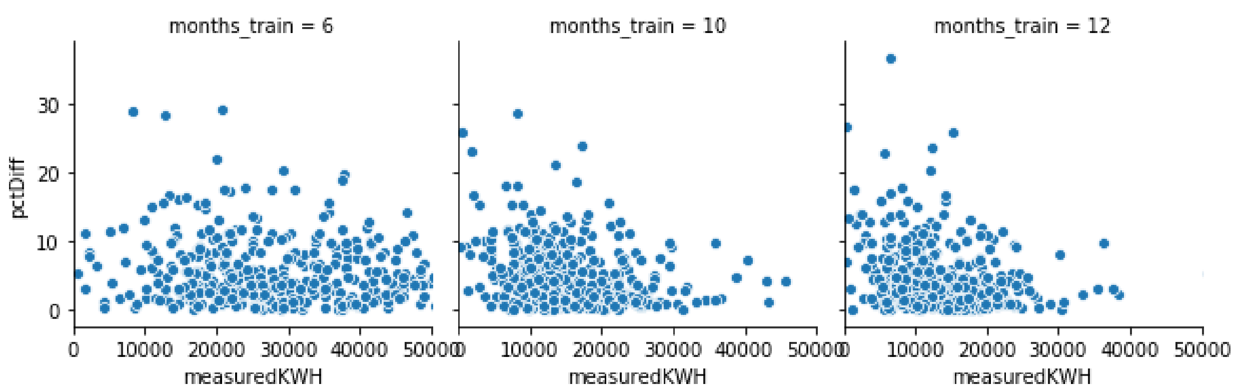

Figure 13. Plots of pct difference vs. total measures kWh for the interval data, grouped by months of training

Conclusions

Monthly data is not sufficient to forecast kWh usage with sufficient precision for pre- and post-installation purposes. Disaggregating monthly data to a daily cadence did not improve the forecasts, and in some cases, produced worse predicted values. Using interval data aggregated to an hourly cadence to fit models, using random forest regression and long short-term memory methods, shows promise in estimating pre- and post-installation usage. Even with only six months of training data, the predicted values were usually within 10% of the measured values. For both machine learning methods explored in this study, the only features used were two lagged kWh readings. Adding other features, such as the annual mean and median for the account or additional distinguishing information, would be an interesting extension of this work. Another variation would be to use the interval data at its original 15-minute cadence. For this study, we aggregated the data to an hourly cadence to match available ASOS weather information. We could have chosen to disaggregate the weather data to match the 15-minute cadence, which may have affected the outcomes.

References

Akbar, Bilal, Khuram Pervez Amber, Anila Kousar, Muhammad Waqar Aslam, Muhammad Anser Bashir, and Muhammad Sajid Khan. "Data-driven predictive models for daily electricity consumption of academic buildings." AIMS Energy 8, no. 5 (2020): 783-801.

Concepts and Options for Determining Energy and Water Savings Volume I, International Performance Measurement & Verification Protocol (International Performance Measurement & Verification Protocol Committee, 2002), DOE/GO-102002-1554.

Edwards, Richard E., Joshua New, and Lynne E. Parker. "Predicting future hourly residential electrical consumption: A machine learning case study." Energy in Buildings 49 (2012): 591-603.

Hamdoun, H., A. Sagheer, and H.Youness. "Energy Time Series Forecasting-analytical and empirical assessment of conventional and machine learning models." Journal of Intelligent & Fuzzy Systems 40, no. 6 (2021): 12477-12502.

Román-Portabales, Antón, Martín López-Nores, and José Juan Pazos-Arias. "Systematic Review of Electricity Demand Forecast Using ANN-Based Machine Learning Algorithms." Sensors 21, no. 13 (2021): 4544.

Runge, Jason, and Radu Zmeureanu. "Forecasting Energy Use in Buildings Using Artificial Neural Networks: A Review." Energies 12, no. 17 (2019): 3254.

Seyedzadeh, Saleh, Farzad Pour Rahimian, Ivan Glesk, and Marc Roper. "Machine learning for estimation of building energy consumption and performance: a review." Visualization in Engineering 6, no. 5 (2018).

U.S. Energy Information Administration (EIA), "Annual Electric Power Industry Report (EIA-861)", 2018.