Statistical learning to estimate energy savings from retrofitting in the Norwegian food retail market

Severinsen, A.* Myrland, Ø.†

2022-05-17

Abstract

Buildings worldwide consume about 40% of all produced energy and are major contributors to GHG emissions. Hence, to reach the 2030 European energy efficiency target it is vital to reduce the energy consumption in buildings. An important barrier that hinders renovation projects is uncertainty regarding the expected savings. The main objective of this paper is to present two different statistical methods to estimate energy savings. The two methods are easy to implement for practitioners within the energy retrofitting industry, and at the same time has acceptable precision and reliability. The two methods are applied at 5 different food retail stores that undertook renovation in 2019. The models are trained on data from 2018 (one whole year before any of the retrofitting's took place) and are further applied to estimate the energy savings in 2021. The first method is the Tao Vanilla benchmarking method (TVB). The TVB model predict energy consumption in buildings on an hourly level. The model has received a lot of attention within the load forecasting literature and has previously proved its performance in machine learning competitions. The TVB has a straightforward specification, and the model parameters are easily understood. This is the first study that apply the TVB to estimate energy savings in a large retrofitting project within the energy and building sector. The second method relies on a more common industrial approach, which is to use weekly data and energy temperature curves to document energy savings. In addition, we demonstrate a novel approach of using broken line (BL) models to estimate energy savings. The suggested BL approach can simultaneously estimate all the model parameters and yield a full covariance matrix within a standard linear regression framework. The results from the

*Corresponding author, Norwegian University of Life Sciences, School of Economics and Business, Campus Ås, Universitetstunet 3, 1433 Ås, Norway, alexander.severinsen@nmbu.no

†UiT The Arctic University of Norway, School of Business and Economics, Breivangveien 23, 9010 Tromsø, Norway, oystein.myrland@uit.no

retrofitting projects demonstrates considerable energy savings between 25% and 55%. Furthermore, both the TVB and the BL models deliver reliable precision. The estimated energy savings from both models are coinciding. This indicates that they could jointly be used to gain insight that may lead to more informed decisions for energy saving projects. The TVB model proves to be a proficient benchmarking model that can give detailed hourly information about the savings. The BL model is used to gain intrinsic details about the buildings varying cooling and heating needs depending on the outside temperature during the year.

Keywords— Energy savings evaluation, Building energy retrofitting, Measurement and verification, Data driven models, Broken line models, Tao Vanilla Benchmark model

Word count: 7436

Abbreviations— ASHRAE;American Society of Heating; Refrigerating and Air-Conditioning Engineers, MV; Measurement and Verification; BL; Broken line model, TVB; Tao Vanilla Benchmarking model, ESCO; Energy Service Company, GHG; Green house gas, HVAC; Heating, ventilation, and air conditioning, ECM; Energy Conservation Measures, CPT; Changing point temperature, CW-GB; Component-wise gradient boosting, CV-RMSE; coefficient of variation root mean square error

1 Introduction

Globally, the building sector use about 32% of all generated energy, 51% of the global electricity use and accounted for 19% of all energy-related GHG emissions [1]. Within the different building categories food retail stores are one of the largest consumers of energy. For instance, the EIA's latest commercial buildings energy consumption survey finds the average energy use for food stores are 524 kWh/m^2; the highest energy intensity of any of the building types [2]. Hence, to reach the 2030 European energy efficiency targets it is vital to reduce the energy consumption of buildings, and retrofitting is known as an important driver to improve energy efficiency [3]. Nonetheless, one important obstacle that hinders renovation projects is uncertainty regarding the expected savings [4].

The work presented in this paper is in close collaboration with a medium sized Norwegian energy service company (ESCO) that has specialized in retrofitting food retail stores, and the current research focus the attention to energy conservation measures (ECMs) that was conducted in 5 different stores during autumn 2020. Thus, one set forth to measure the effect of the energy savings as a result of the ECMs for the year 2021. All the buildings got new LED-lightning and refrigeration systems, additionally one store got a new

Table 1: Yearly energy consumption (2018), gross area (m^2) and energy intensity (kWh/m^2)[a]

Store-id	Gross area (m^2)	Yearly kWh[b]	kWh/m^2
4 103	1 409	431 989	307
4 097	1 066	545 159	511
4 479	356	207 713	583
4 396	1 514	1 061 682	701
4 391	1 412	1 062 906	753

[a] Data collected from meter readings for the whole of the buildings

[b] Energy use one whole year before any retrofittings

HVAC system installed. The control systems were also renewed and optimized across the stores. The average cost of the ECMs was: lighting NOK 100 000, the HVAC system NOK 800 000, and the refrigeration system NOK 4 000 000. Also, the project incurred administrative cost of NOK 350 000, including an energy audit of the different buildings.

In table 1 the size and the electricity consumption for the 5 different food retail stores that was retrofitted is presented. Note that the actual store names are anonymous in agreement with the ESCO and the building owners. The size of some of the stores are quite different with a range from 356 m^2 to 1514 m^2. The store with the largest yearly energy consumption is store-id 4291 with a yearly consumption of 1 062 906 kWh. Also, note that there is a substantial difference in the energy intensity measured through kWh/m^2. For instance, store-id 4391 has 753 kWh/m^2, while store-id 4103 has 307 kWh/m^2, thus a 59% lower energy intensity. These measures are often used by practitioners to rank the potential energy savings in buildings, hence, one expects the estimated savings to correlate with these measures.

An important concern for practitioners in the retrofitting industry is to have reliable methods to document energy savings. Several approaches exist, both deterministic and data-driven. However, many of the data-driven methods are quite complex and time consuming to conduct. Further, during this research project we have seen that the ESCO prefer methods that are easy to understand and to communicate to clients - a vital aspect of reporting the savings. For instance, current research from field experience show that interpretability of models may even keep the clients from accepting complicated models such as artificial neural networks [5]. The ESCO that we worked together with used weekly data and linear regression as a basis of modeling what the energy consumption would have been without the ECMs. While this method worked relatively fine this paper will demonstrate other methods that may improve the reliability. Additionally, new advanced metering systems with high-frequency data have led to advanced approaches that give new and detailed insights into the effect of retrofitting's. Since this research has been in close collaboration with the ESCO, one of the main objectives was to improve their way of estimating energy savings. As such, it is difficult to approach commercial actors that over many years have been accustomed to their own preferred set of solutions; solutions that have served them well. Thus, these issues are approached carefully.

This paper demonstrates two different methods to estimate energy savings, the broken line model (BL) and the Tao Vanilla benchmark model (TVB). The BL model use weekly data and is relatively close to the ESCO's established method. Furthermore, the research demonstrate that the BL model was uncomplicated to implement, reliable, enhance understanding, and the methods resemblance to the ESCO's current workflow eased the uptake. Second, the TVB model is based on hourly data and demonstrates the added benefits and insights that can be gained by higher frequency data.

The BL and the TVB model are then compared in terms of reliability, advantages and disadvantages. Previous experiences have shown that there is a large energy efficiency potential in the existing building stock, and that the potential is mainly untapped. One important reason for this is the lack of reliable methodologies to evaluate the effect of energy efficiency measures [6]. In that respect one of the objectives of this research is to fill that gap.

1.1 Novelty of the paper

First, the TVB-model, published in [7], has been used frequently as as benchmarking model within the load forecasting literature [8–10], and has previously proven to be among the top performers within machine learning competitions. For instance, the model was ranked among the best 25 of 100 teams in the GEFCom2012 [11]. In previous research the TVB was applied in [12] to estimate energy savings from ECMs with small expected effects. However, in the present paper the TVB is used to document energy savings for food retail stores that has implemented extensive retrofitting. Given the models previous prediction performance, easy implementation, and the lack of use to estimate savings in retrofitting projects the present paper promotes the novelty of the method, and adds to the already established data-driven tools within the M&V industry.

Second, we use the BL model to estimate the changing point temperature (CPT) value, and the cooling and heating slopes. Standard methods to estimate non-linear effects, such as regression splines, polynomial regression, and non-parametric smoothing are not relevant because the CPT values are fixed a priori, and the regression parameters are not directly interpretable [13–16]. The BL model is estimated within a linear framework and is accessible through the *R* package *segmented* [17]. The package is easy to implement for practitioners and use of the method has previously, as far as the authors has been able to find, not been published within the M&V literature.

Third, the data used in this project is unique and it is the first research that document the potential energy savings from the above mentioned ECMs in the food retail sector in Norway. Furthermore, lack of reliable information may be a barrier for new renovation projects [18], and as such, the results from this paper is a novel contribution and may advance interest in similar projects.

This rest of this paper is structured in the following way. First, the relevant data is presented. Second, an exposition of the measurement and verification (M&V) industry and related research is offered. Third, the methods section presents the two models used to estimate the savings. Fourth, the results are presented, implications discussed, limitations and suggestions for future research, and at last the conclusion is offered.

1.2 Data - electric load and weather data

In recent years there has been several breakthroughs in advanced metering infrastructure systems, and easier access to high-frequency data has even transitioned and renamed the Measurements and Verification (M&V) industry into M&V 2.0. New metering systems allow for energy savings being estimated close to real-time [6].

For instance, Statnett, the system operator of the Norwegian power system, owns and runs, Elhub AS. Elhub is a central IT system to support and streamline market processes in the Norwegian electricity market, but they also support the distribution and aggregation of metering values for all consumption and production in Norway. Their system daily collects energy use on an hourly level. It is obligatory for all the Norwegian grid operators to update the Elhub repository each day. The service was launched in February 2019. All the energy data from February 2019 and on wards for the 5 food retail stores in this paper stems from Elhub[1]. The energy data from 2018 up until January 2019 is collected from the building energy management system (EMS) that previously collected data from the grid operators. Temperature data is downloaded from the Norwegian Meteorological Service (www.met.no). Each stores position (longitude and latitude) is mapped against a 2.5km x 2.5km grid of Norway. Further, the temperature data gathered is modeled weather data that use several of the closest weather stations to set the temperature.

Figure 1 shows an example of a typical hourly electricity consumption for one of the food retail stores in this paper. As can be seen the consumption follow the same pattern every day depending on the opening hours, except for Sunday when the store is closed. During the night the consumption fluctuates around 150 kW, and when the store opens the kWh shift to around 200. Also, note the extra peak (often referred to as the “morning ramp”) when the store opens at 07:00. This is a feature seen in many food retail stores and is attributed to the shift from night to day-mode for the refrigeration and HVAC system (which is on

[1]the data gets pulled each morning from Elhub and stored in a database (postgreSQL).

"stand-by" when the store is not open).

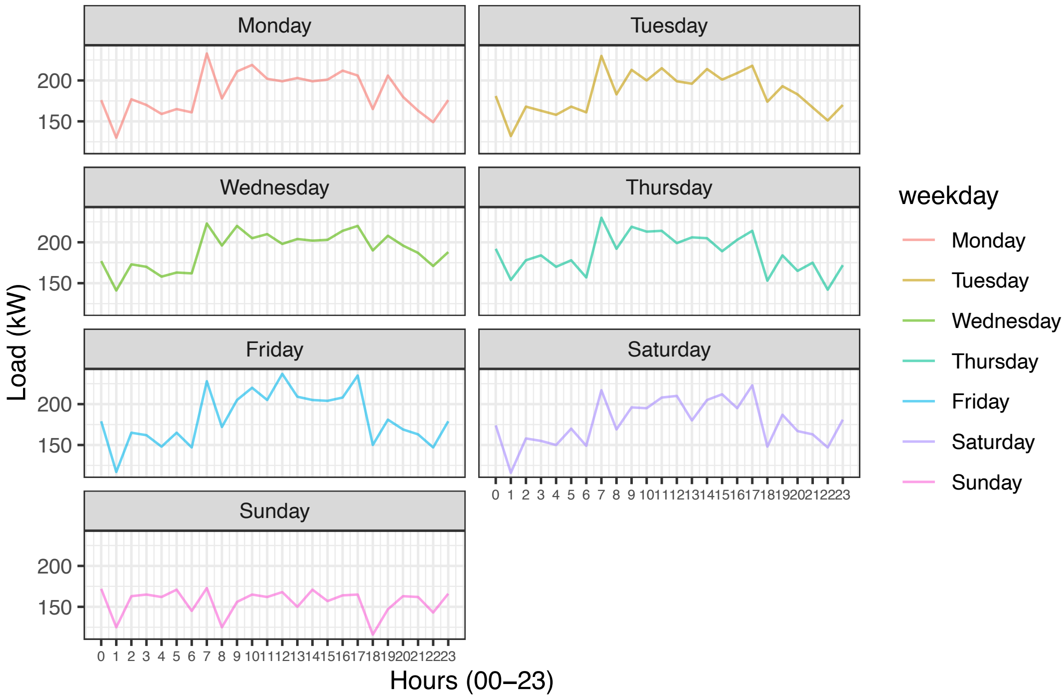

Figure 1: Hourly loads (kW) throughout a week

Furthermore, figure 2 shows the same data for the same food retail store, but aggregated to a weekly energy use level together with the average outside temperature for one whole year. This figure clearly demonstrates the relationship between outside temperature and the electricity use. In winter the electricity consumption increases due to heating demand, and on contrary the energy consumption is much lower during the summer months, though there are some "spikes" here and there during summer which can be explained by cooling needs in the warmest summer days. These features are important to take into consideration when building models to predict the energy use in the buildings.

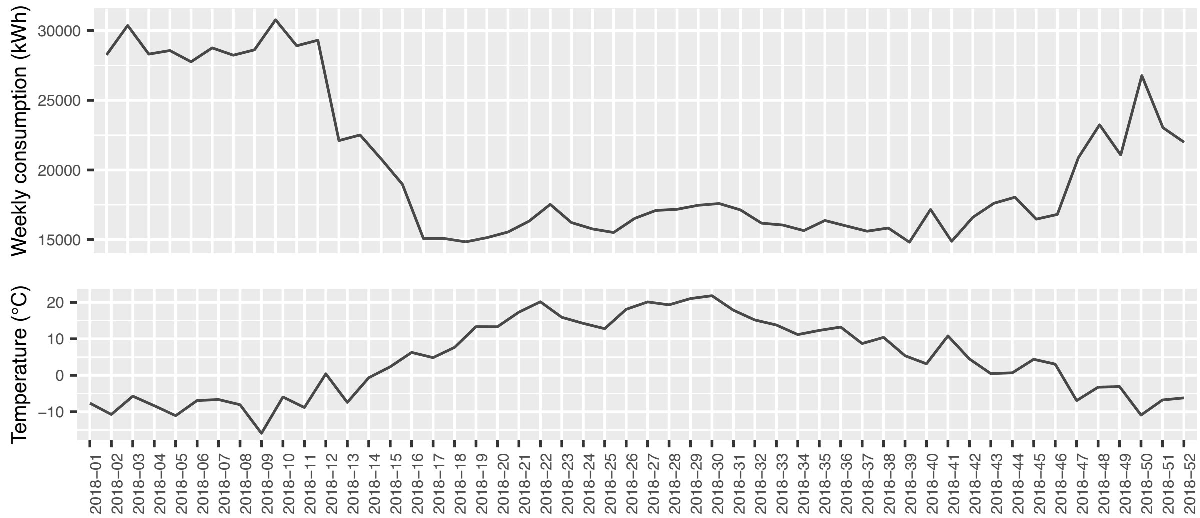

Figure 2: Weekly energy consumption (kWh) and site spesific weekly average temperature (°C) for a food retail store

1.3 Measurement and verification

Measurement and verification (M&V) is the process of using measurements to accurately estimate energy savings generated in a building as a result of implementation of an energy management strategy [19]. In order to compare the energy usage before and after the implemented retrofitting, a model of the consumption prior to the retrofitting needs to be developed. This model is often referred to as the baseline energy model.

Figure 3 is an illustration of the measurement and verification process. The y-axis represents energy consumption and the x-axis time. The vertical dotted line represents the implementation of an energy retrofitting; let's say the change of coolers and freezers in a store. The expectation is that one will find a substantial decrease in the energy consumption as this new equipment is much more energy efficient than the old coolers and freezers. The baseline period represents how the building consumed energy before the ECMs. In this paper this period is the year 2018, decided after a review of the data together with the building owners. It is important that the baseline period is representative of the energy consumption in the building, otherwise the measurements will not be correct. The solid line represents the actual energy use in the building. Note that once the ECMs were implemented the energy consumption decrease. Now the question is whether this decrease was due to the ECMs or other external factors? For instance, the outside temperature might have been very different in the baseline period compared to after. This is the reason that a model is needed to estimate the energy savings. Imagine that no ECMs were implemented. This is represented by the dotted line and is the potential energy use if there were no retrofitting's. The difference between the actual energy use and the dotted line is the energy savings, illustrated by the red arrow in the graph.

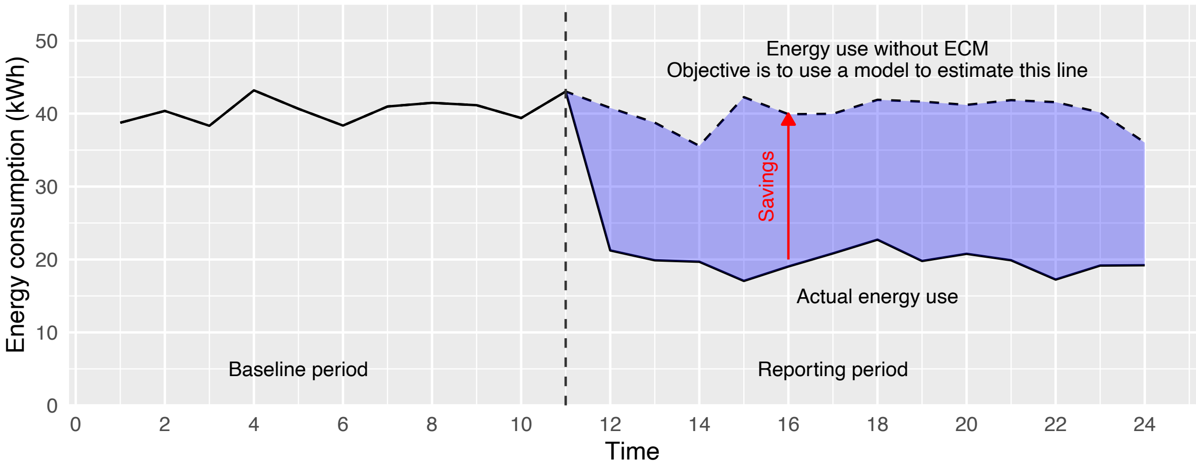

Figure 3: Illustration of measurement and verification. Baseline = before the retrofitting, Reporting period= after the ECM

Within M&V there are various methods and best practices, and there also several standards that have been suggested [6]. This paper follows the ASHRAE Guideline 14 for measurement of Energy, Demand and Water Savings [20] . This protocol suggests best practices to quantify energy savings, including metrics to evaluate the validity of the models. The protocol has three different options to determine energy efficiency savings.

- *Retrofit Isolation*: No estimation is allowed. For example, if you install a new refrigeration system in a food retail store you need energy data on that particular system before and after. This often requires sub-meters (sensors) that can collect these data and was not available in this study.
- *Whole facility.* The present research use meter readings (from Elhub.no) to evaluate the energy performance of the whole building. This option determines the savings of all the implemented ECMs. This option is recommended for projects where the expected savings are substantial and is the approach followed in this paper.
- *Whole building Calibrated simulation.* Using building energy modeling software that allows the prediction of energy consumption. Often requires extensive physical data.

To conform with the ASHRAE protocol the research literature has suggested several useful approaches. In the next section an overview of relevant research is offered to set the suggested modeling approaches into context.

1.4 Baseline models to estimate energy savings

There are two different main classes to estimate energy savings: data-driven and deterministic models. Data-driven models are statistical models that find relationships between a dependent variable (energy consumption) and feature variables (air temperature, like in this paper, or wind speed, solar irradiance or other external factors that may impact energy consumption). The other class is deterministic: typically, a detailed simulation model based on the energy transfer flow within the building. For example, one established and well-known tool for building modeling and simulation is EnergyPlus, which is a freely available energy modeling software. The software has been used to simulate energy performance and savings in buildings [21,22], however, since the results are based on simulations, and not actual conducted retrofitting's, the savings are theoretical. In the retrofitting business the actors must document actual savings. Nonetheless, it is possible to adapt energy modeling software into prediction tasks, however the software typically requires extensive physical building data, something that may be difficult to acquire, and if possible, may complicate model training [23]. Based on this it is attractive to investigate simpler models without a strong dependence on physical data, and as such data -driven methods may be a useful candidate to simplify prediction.

Several recent reviews find data-driven methods scalable and more effective than traditional approaches [24–27,27–30]. Hence, this paper focuses the attention on data-driven methods. Figure 4 presents an overview of the different data-driven baseline energy modeling approaches, and was the starting point of a recent review of data driven methods by [6]. They separate data-driven methods into three main paths: statistical learning, machine learning and Bayesian methods.

1.4.1 Statistical learning

The two approaches that is presented in this paper fits within the statistical learning path 'linear and nonlinear regression.' This path has a long history within the M&V industry. In 1986 the PRInceton Scorekeeping Method (PRISM) was proposed as the standard method to measure energy conservation savings [31]. The PRISM is a piece-wise linear regression model with monthly electricity consumption, using heating degree-days for weather normalization. The PRISM has been a popular approach both with academia and industry and has over the years received more than 450 research citations. However, as energy data became more available, models using weekly, daily and hourly data was introduced, both using multiple linear regression [32] and change-point models [23]. In a more recent study [33] use multiple regression with a time-of-week indicator variable (similar to what the TVB model described in the next section use) and a piece-wise linear and continuous outdoor air-temperature dependence.

Furthermore, using monthly data [34] used linear regression to estimate the energy savings of changing the HVAC system in an office building, and [35] used both linear and nonlinear regression models to document the energy savings as a result of mechanical system retrofitting in a healthcare facility.

As can be seen from figure 4 transfer functions and kernel regression have their own paths within statistical learning. Transfer function has been deployed to estimate energy savings in a building of the University of Granada [36], however the method requires the internal temperature of the building, and that was not available for any of the buildings in our study. Kernel regression was initially proposed by [37] to improve the accuracy of standard linear regression, however there are some concerns regarding the methods ability to take into account seasonal variations [6].

1.4.2 Machine learning

The second main path of data-driven models is 'machine learning.' For instance, [12] compares the TVB model and gradient boosting in 9 different food retail stores that had implemented ECMs with low expected savings. They found that gradient boosting did perform somewhat better in terms of accuracy, but both models performed well below the ASHRAE CV-RMSE limits set for reliable estimates of energy savings for all the buildings. Furthermore, one advantage of the gradient boosting approach was that the model enabled to identify a unique feature set of the best explanatory variables for each of the buildings. One the other hand, the tuning of the model was time consuming, and the approach was not easy to communicate to the ESCO. Further, Artificial neural networks (ANN) have seen several applications to estimate energy savings. The ANN are easy to implement, but on the contrary are not that easy to interpret. Another drawback is that ANN need large sets of training data. In [38] they document energy savings in two hotels using ANN models with Levenberg-Marquardt back-propagation. To develop the ANN baseline model they used weather, occupancy and building operation schedules.

1.4.3 Bayesian methods

Bayesian statistics is an approach to parameter estimation based on Bayes' theorem and is quite different from the frequentist approaches presented so far in this section. For instance, in the frequentist approach you only use the actual data to estimate the energy savings, but in the Bayesian approach you integrate prior information about the expected savings from the retrofitting's. In many cases prior information may be an advantage for the analysis, however it is also a known limitations of Bayesian statistics that the priors may be challenging to justify and can be a source of inaccuracy. We are not aware of any Bayesian studies that estimate energy savings from the whole-building perspective similar to this paper. However, [39] applied Bayesian statistics to estimate savings of a model-predictive controller for space heating for a Swiss office building. They argue that the traditional statistical approach is expensive, however, easy access to data through Elhub.no and open meteorological data through services such as met.no makes today's model building using frequentist approaches quite inexpensive.

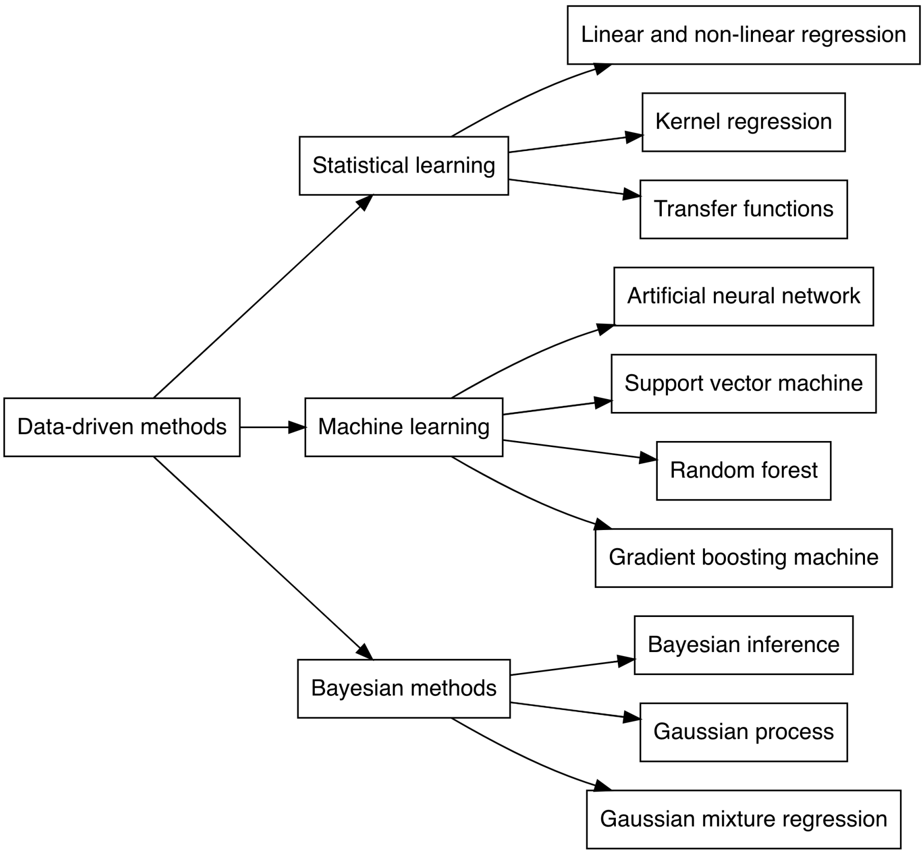

Figure 4: Overview of baseline modeling approaches to estimate energy savings

2 Methods

This section presents the two methods that will be used to estimate the energy savings: broken line models using weekly data, and the TVB model using hourly data.

2.1 Broken Line Models

Broken line (BL) models are common in many different fields, such as toxicology, ecology, epidemiology, and medicine [40,41]. These models are used to estimate two straight lines connected at unknown values, often referred to as change-points or breakpoints. For example, when the outside temperature is cold this typically leads to an increase in a building's energy consumption due to increased demand of heating. In the same way the energy consumption increases due to hot temperature in the summer when coolers in the stores are in use. The change point, changing point temperature (CPT) is the point at which no heating or cooling is required.

The classical methods used to take into account non-linear effects, such as polynomial regression, non-parametric smoothing, and regression splines are not applicable because the change-points are fixed a priori. Further, regression parameters obtained in regression splines or polynomial regression approach are not directly interpretable [42]. When the CPT parameters must be estimated, standard likelihood-based inference is convoluted by the fact that the log-likelihood is only piecewise differentiable and the classical regularity conditions are not met [14–16]. In this paper the problem is reduced to a linear framework. The CPT relationship between the mean response $\mu = E[\Upsilon]$ and the variable Z is modeled by adding in the linear predictor for

$$\beta_1 Z_i + \beta_2 (Z_i - \psi)_+ \tag{1}$$

Where $(Z_i - \psi)_+ = (Z_i - \psi) \times I(Z_i > \psi)$ and $I(\cdot)$ is the indicator function equal to one when the statement is true. Accordingly, β_1 is the left slope, β_2 is the difference-in-slopes, and ψ is the CPT value. Several challenges have previously been described by [43]. For instance, grid-search algorithms have been used to estimate broken-line models, for example fitting several linear models and searching for the value that corresponds to the model with the best fit. Despite, this is not an optimal approach when there is more than one changing point or a large dataset. Also, estimating models with fixed changing point may lead the parameters to have to narrow standard errors.

The *R* package *segmented* can estimate and summarize generalized linear models with broken line relationships. The package uses a method that simultaneously estimate all the model parameters and yields the approximate full covariance matrix [17]. For example, [44] shows that the nonlinear term in equation (1) has an approximate intrinsic linear representation. Thus, given an initial guess for the breakpoint (the CPT value), ψ , a standard linear framework can be utilized to solve the problem. Previous research has established the CPT value for food retail stores to be around 7, consequently, $\psi = 7$ [45].

The *segmented* package estimate model (1) by iteratively fitting the linear model

$$\beta_1 Z_i + \beta_2 (Z_i - \tilde{\psi})_+ + (z_i > \tilde{\psi})\tilde{\psi}^- \tag{2}$$

where $I(\cdot) = -I(\cdot)$ and γ is the parameter to be interpreted as a re-parameterizaton of ψ, thus accounts for the breakpoint estimation. At each iteration, a standard linear model is fitted, and the breakpoint value (CPT) is updated through $\psi = \psi + \tilde{\gamma}/\tilde{\beta}_2$.

2.2 The Tao Vanilla Benchmarking model - Estimating the energy savings

In the previous section an exposition of the BL model, that will be applied on weekly aggregate data, was given. For the same purpose of estimating energy savings on an hourly level the Tao Vanilla benchmark (TVB) model will be used. The TVB model has proven easy to implement and produce accurate results [7]. Previously, the TVB model has been used for load forecasting by grid operators, a noteworthy exception is found in [12] that estimate energy savings using the TVB model to benchmark against component-wise gradient boosting with p-splines (CW-GB), in a context where the expected savings target was below 10%, for instance in smaller implemented ECMs. Both the TVB and the CW-GB was found to produce reliable results. The TVB model has the following specification:

$$\begin{aligned} Y_t = \beta_0 + \beta_1 M_t + \beta_2 W_t + \beta_3 H_t + \beta_4 W_t H_t + \beta_5 T_t + \beta_6 T_t^2 + \\ \beta_7 T_t^3 + \beta_8 T_t M_t + \beta_9 T_t^2 M_t + \beta_{10} T_t^3 M_t + \beta_{11} T_t H_t + \beta_{12} T_t^2 H_t + \beta_{13} T_t^3 H_t \end{aligned} \tag{3}$$

where Y_t is the actual load for hour t, β_i are the estimated coefficients from the least squares regression method; M_t, W_t and H_t are month of the year, day of the week and hour of the day. Further, T_t is the outside temperature for time t. Note that the original TVB model includes trend and past loads. However, in this paper the TVB model will reflect the energy consumption in food retail stores based on a reference period, thus trend and lagged variables are not included as predictors.

2.3 Model accuracy

To measure the accuracy of the TVB and the BL models the *coefficient of variation root mean square error* (CV-RMSE) is calculated. The CV-RMSE is computed in the following way,

$$CV - RMSE = \frac{\frac{\sum\left(\hat{Y}_i - Y_i\right)^2}{n-k-1}}{\bar{Y}} \tag{4}$$

where $\bar{Y}$ is the mean of the energy consumption in the training data (the reference/baseline year). Y_i is the actual energy use in hour i, $\hat{Y}_i$ is the predicted value of energy use in hour i from the model, estimated on the reference period. Further, n is the sample size, and k is the number of independent variables in the model. This accuracy measure is recommended by the ASHRAE [46] and for reliable baseline models the CV-RMSE is required to be below 20% for the model to be accepted if post retrofit period is less than 1 year, and less then 25% if between 12-16 months after the ECMs.

3 Results

In the following section the total estimated energy savings from the two different modeling approaches is presented, the TVB and the BL model. Furthermore, follows a detailed presentation of the results from the two models.

3.1 Aggregated energy savings

Table 2 sums up the main findings, both the estimated % savings and the CV-RMSE from the two models. First, the ECMs had estimated energy savings ranging from 25% to 56%. Further, note that there are hardly any differences in the percent energy savings if using the BL or TVB models. Store-id 4391 had the largest estimated saving. That store had an actual electricity consumption of 457 000 kWh in 2021, and the models predicted that the consumption without the ECMs would have been 1 040 015 kWh (BL model) and 1 026 125 for the TVB model. Hence, the %-savings was 56% from the BL model, and 55,4% from the TVB model. A substantial saving, nonetheless, this was also the store with the most potential as measured from

Table 2: Aggregate energy savings and CV-RMSE results from TVB and BL models

Store-Id	kWh BL model	Actual kWh	kWh TVB model	CV-RMSE BL	CV-RMSE TVB	% savings BL	% savings TVB
4391	1 040 015	457 500	1 026 125	0,087	0,170	-56,0	-55,4
4396	1 036 211	554 757	1 032 992	0,033	0,075	-46,5	-46,3
4479	204 767	112 911	202 194	0,056	0,085	-44,9	-44,2
4103	417 657	295 295	418 866	0,044	0,134	-29,3	-29,5
4097	529 024	397 062	529 776	0,021	0,087	-24,9	-25,1

the energy intensity (kWh/m^2 pre-ECM, see table 1). The store with the lowest energy savings was store-id 4097 with the estimated savings equal to 24,9% and 25,1%, BL versus TVB, respectively.

Note that the CV-RMSE is less then 25% for all the models, thus well within acceptable limits following the previously discussed ASHRAE guidelines.

3.2 BL Model - Energy Temperature curves

In figure 5 the energy temperature curves (ET - curves) are presented for the 5 different food retail stores. The y-axis represents the weekly energy consumption (kWh) and the x-axis the weekly average outside temperature. The BL model was used to estimate the lines that was fitted to the weekly data for 2018 (blue colored round circles). Note the changing point temperature (CPT), which is the temperature point where the building shifts between heating and cooling needs. It is quite some variation for the CPT between the buildings, ranging from 6.7°C to 15.1°C. The CPT values are valuable in terms of understanding the details about the building envelope, e.g. degree of insulation or the efficiency of the heat recovery system. The black rectangular point in the curves is the energy consumption and average temperature for all the weeks in 2021, thus after the implemented ECMs (note that these were not part of the modeling process / fitted lines). Hence, the distance from the black rectangular points up to the fitted line is the actual energy savings. This is illustrated for store-id 4391 for a winter week and a summer week in 2021. The red arrow that extends from week 3 (x-axis, temperature -10.3 °C and y-axis, actual energy consumption that week of 10 400 kWh) to the fitted line at 28 500 (the kWh given no ECMs). Hence, the saving is 28 500 kWh - 10 400 kWh = 18 100 kWh. The equivalent numbers for the blue line in a summer week is 7 850 kWh actual versus 16 000 kWh predicted, a saving of 8 150 kWh. The point of illustrating this is that the ET - curves gives both a good visual representation of the heating and cooling demands, and is a method used to better understand the seasonal effects of the savings. For instance, since store-id 4 391 had a new refrigeration system including a very efficient heat recovery system it was expected that the ECMs gave more savings in the wintertime, as documented in the figure. Both the CPT values and the visual representation are unique features of the weekly aggregate level.

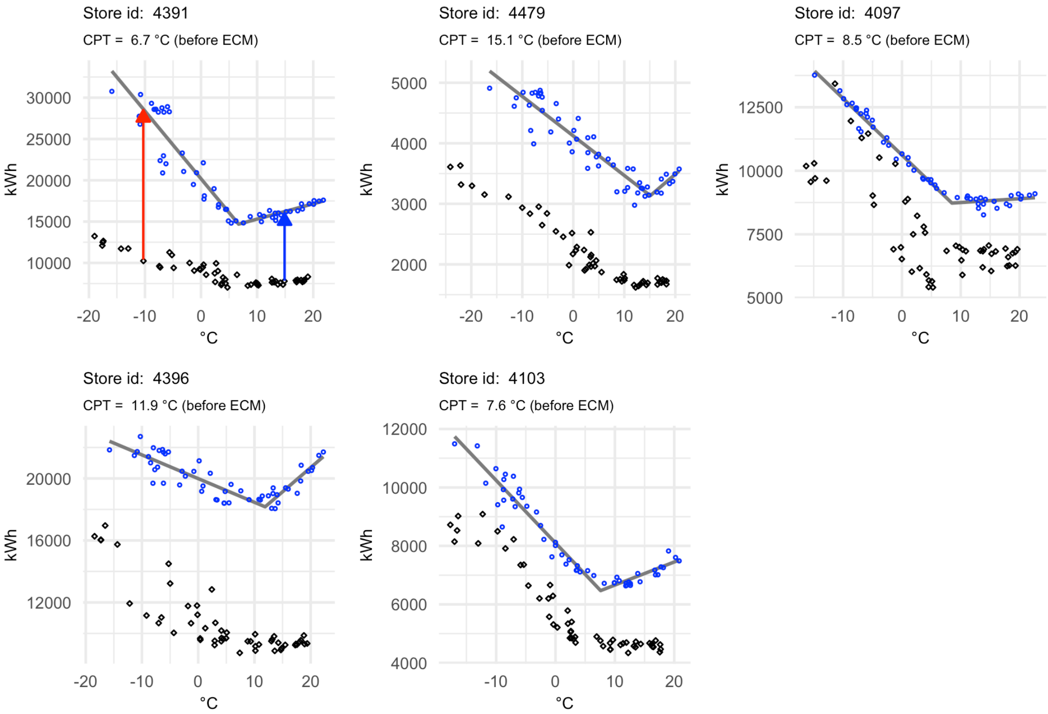

Figure 5: ET curves for the reference periode estimated with the broken line model, and the weeks for 2021 displayed

3.3 TVB model - savings on the hour

In figure 6 the energy savings results is presented for week 2 in 2021. The solid line is the actual load, and the dotted line the predicted load given that the building performed as before the implemented ECMs. Note that the difference between the lines represent the actual savings. There is a substantial savings for all the weekdays. Also, the peak in the morning is not present anymore, or at least much less pronounced. The TVB model gives a much more detailed understanding of the actual savings compared to the BL model. For example, plots like the one illustrated in figure 6 can even be used to detect and narrow down the cause of errors in the technical system. The predictions can spark questions like "is the reduced savings between 22:00 and 01:00 due to a slower night-mode shift in the ventilation, or could it be that the automatic lighting switch is not working properly?" These are questions that are much easier to investigate when the savings are presented on an hourly level.

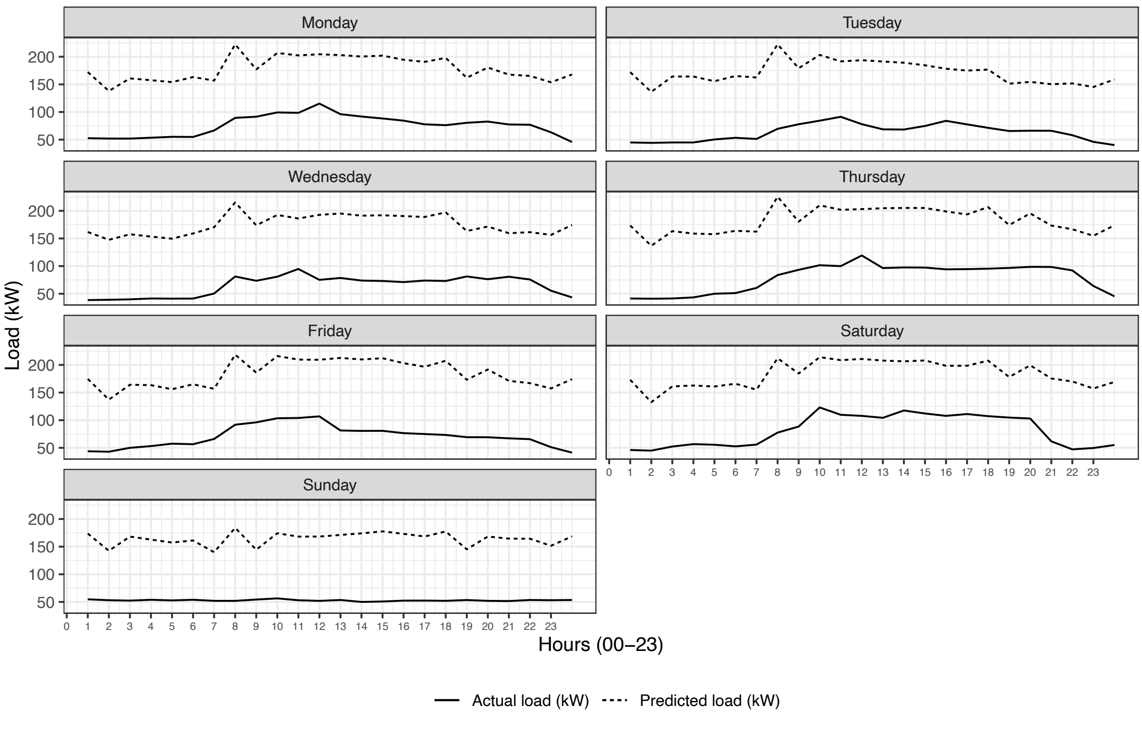

Figure 6: Actual and predicted loads (given no ECM) from the TVB model

3.4 TVB model - peak shaving

Peak shaving refers to measures taken to reduce the electricity loads, often at specific hours. There may be several reasons to take this into account. For the owners of the food retail stores the grid rent is often designed in such a way that the building owners pay for the maximum load each month. For instance, if the building had a flat load at 220 kW through December 2021, but then suddenly the hour on 16th December 07:00 was 431 kW, then 431 kW would be what the grid rent was based on.

In figure 7 the actual and the predicted (given no implemented ECM) average, minimum and maximum loads for each month of 2021 for one of the stores using the estimates from the TVB model is plotted. The maximum predicted load for November 2021 was 235 kW and the actual was 117 kW. This amounts to a load reduction of 235-117 = 118 kW. To relate this to the grid rent this store has *Vevig AS* as their grid operator. The effect price in the winter months is NOK 55.9. Thus, the ECMs gave a savings in November 2021 of 118 kW * 55.9 = NOK 6 596. As can be seen in the figure the load reduction was substantial across all months of the year. This is an important perspective that is not possible to study on a weekly level.

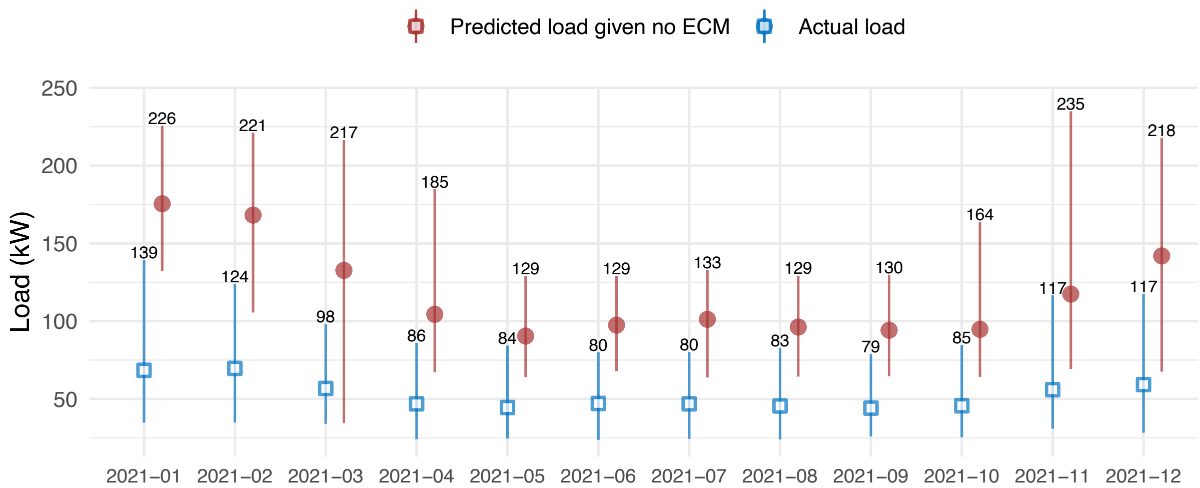

Figure 7: Average, minimum and maximum load (kW) for each month. Results from TVB model, showing the peak shaving that occured on a monthly basis for the store with the largest energy savings

4 Discussion

Previous research documents that the energy efficiency in the existing building stock has a considerable potential, and that the lack of reliable methodologies to evaluate the effect of energy efficiency measures may have impaired progress [6]. Incidentally, in the review of the baseline models several reliable data-driven methods, both statistical, machine learning and from the Bayesian point-of-view are presented. Nonetheless, several of these approaches are relatively complex, and research has established that interpretability of models may keep the clients from accepting black-box models [5]. In other words, it is a bit of a paradox. The industry calls for reliability, and when given turns it down as too complex! The present paper tries to balance complexity versus usability, keeping a steady eye on the reliability of the models in terms of predictive power. Hence, a demonstration of two different baseline models that clearly demonstrates the potential energy savings that retrofitting's may lead to. Table 1 displays that the difference in energy intensity across the stores was substantial, and store-id 4391 was identified as the food store with the largest potential for energy savings with energy use of 753 kWh/m^2. Also, according to the ESCO this store was, during the energy audit before the implemented ECMs, identified with several technical issues in the control system that lead to a higher energy consumption. Hence, in the ECM period the technical system was optimized, and the store was renovated with new LED-lights, refrigeration and HVAC system. The energy savings for 2021 was estimated to be 56%, which amounts to almost kWh 550 000. Furthermore, the store with the lowest energy intensity was store-Id 4103 with 303 kWh/m^2. The estimated savings for this store was a kWh reduction of 30%.

The average food retail store in Norway consumes 500 000 kWh. Thus, based on the average estimated savings in this paper the potential is a reduction of 35%; annually kWh 175 000. Indeed, apply this to the 4000 food retail stores in Norway the aggregate yearly savings is 700 000 000 kWh, 7000 GWh (1 GWh = 1000 000 kWh). In comparison, the average Norwegian household has an electricity consumption of 16 079 kWh[2]. Hence, the potential energy savings in the food retail stores equals the same amount of energy that 41 893 households consume: or rather, a medium sized Norwegian city. Off course, such a direct extrapolation may not be entirely correct as there are many confounding factors, still it serves as an example to illustrate the scope in the industry. The untapped potential is substantial.

However, in a study by [18] they find that renovations have a low impact on property prices, and lack of reliable information is cited as main barriers that hinders renovation projects for residential buildings. Hence, research that document the savings in different building categories may contribute to better understand the

[2] https://www.ssb.no/energi-og-industri/artikler-og-publikasjoner/vi-bruker-mindre-strom-hjemme

potential and make the results known to the actors within the industry. For instance, for food retail stores the owners often has long-term tenancy agreement, and the savings from renovations of the technical systems is expected to last for up to 10 years. The ESCO we collaborate with has implemented both the BL and the TVB approach, and the ECMs will be followed up yearly. It will be interesting to follow the effect of the ECMs over the years and gain more knowledge about any possible diminishing effects.

In practice many of the ESCOs that we have worked with have used a basic linear regression model to fit a line to the weekly data, not taking into consideration the cooling needs during the summer months. Some other ESCO divided the data into a summer dataset and a winter dataset and fitted individual lines to these. The broken line model implemented in this paper automate this and fits a line that takes into consideration both heating and cooling needs, and at the same time finds the CPT value. The methods are easy to implement using the R package 'segmented' [17]. A review of relevant literature of energy saving models did not find this approach published elsewhere, hence, the method is a useful contribution to further automate the ESCOs workflow. Furthermore, the TVB also proved a useful and reliable method to estimate the energy savings on an hourly level. The TVB model has seen several applications within the load forecasting literature [7–10]. However, within measurement and verification (M&V) in the ESCO industry, the TVB was applied in [12], but in a different context documenting energy savings from ECMs with expected small effects. In this paper the method is applied for food retail stores that has undergone extensive retrofitting. Further, since the TVB baseline model is estimated on an hourly level, this feature allows the analysis of the impact of the energy savings in ways that are not possible on other aggregate levels. Specifically, a comparison between how the ECMs performs on weekdays versus weekends, or when the store is open versus closed. That again may be used to benchmark top-performers, and maybe relate that to optimal settings within the HVAC steering units.

4.1 Limitations and suggestions for future research

In 2019 The American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) hosted the Kaggle competition "Great Energy Predictor III. How much energy will a building consume?"[3] The competition attracted 4,370 participants from 94 countries. The prize money for the winning team was $25,000. A detailed overview of the machine learning workflows and the winning teams is presented in [47]. The top 5 solutions were reproduced by [47] and the accompanying code can be found on github[4]. The winning solutions are presented in table 3. As can be seen 4 out of 5 used multiple methods and post-

[3] https://www.kaggle.com/c/ashrae-energy-prediction
[4] https://github.com/buds-lab/ashrae-great-energy-predictor-3-solution-analysis

processing of data with ensembling and weighting. All the winning solutions used Light GBM, three of the winning teams used Catboost, and two used XGBoost.

Table 3: Kaggle top 5 performing teams - modeling solutions

Rank	Team	Features	Modeling	Post-processing
1	Matthew Motoki and Isamu Yamashita (Isamu and Matt)	28 features	CatBoost, LightGBM, and multi-layer perceptron	Ensembled the model predictions using weighted generalized mean.
2	Rohan Rao, Anton Isakin, Yangguang Zang, and Oleg Knaub (cHa0s)	Temporal features, building metadata, statistical features of weather data	Catboost, XGBoost, LightGBM, and Feed-forward Neural Network	Weighted average
3	Xavier Capdepon (eagle4)	21 features including raw weather and meta data	Catboost, Keras CNN, LightGBM	Weighted average
4	Jun Yang (不用 leakage 上分太难了)	23 features iweather lag features and aggregates	XGBoost and Light GBM	Ensembles. Weights were determined using the leaked data.
5	Tatsuya Sano, Minoru Tomioka, and Yuta Kobayashi (mma)	Target encoding using percentile and proportion and the weather data temporal features	LightGBM	Weighted average

Some of these methods are quite technical and involving and requires a thorough understanding of tuning machine learning models. However, it would be very useful to study the relationship between the predictive power of these methods and those presented in this paper. Furthermore, are there particular methods that are more suitable to specific building types? And maybe most important, what is the practical value in terms of the estimated energy savings when you compare the winning solutions with simpler methods? Since this competition was recent there is a lack of research papers that apply and review these issues. Nonetheless, that would be welcomed and useful research for applied analysts within the field.

There is probably no single modeling solution that fits all building categories. As such, it seems sensible to approach the modeling process with different tools. For food retail stores this paper finds that simple, but well specified, linear approaches work well. As [6] points out, the advantages of linear and nonlinear regression are that the models are easy to interpret and explain, but the models have limitations and may be too simple to capture complex relationships. For the ESCO we worked together with in this paper interpretation and simplicity was important features, and one have demonstrated that both the BL and the TVB model gave reliable results. At the same time, there are several other variables that may have contributed to more precision in the models. For example, in addition to temperature, it would be sensible to test other meteorological data such as wind speed, and solar irradiance. Also, the number of customers in the stores may impact the energy consumption (e.g., increased use of ventilation, door air locks, opening/closing of coolers and freezers).

Furthermore, future research may benefit from incorporating uncertainty measures into the estimated energy savings. As, [48] points outs, such measures can help the stakeholders make more informed decisions. In this paper the difference between the TVB and the BL model using the point estimates gave little practical difference. However, it would be interesting to investigate if the same finding applies when looking at modeling uncertainty. For instance, [49] demonstrates that some methods that use both daily and hourly data underestimates the uncertainty, and that finding applied somewhat more for the hourly models.

At last, machine learning is often associated with the drawbacks that interpretability is demanding. However, the field has made several advances that seem very promising to be able to explain the inner workings of machine learning models. For instance, [50] has proposed the popular framework LIME[5]. The Local Interpretable Model-agnostic Explanations (LIME) has received a lot of attention in recent years, the aforementioned paper has since its publications in 2016 received almost 8000 citations. LIME increase interpretability of the model through local interpretations, as opposed to global interpretations, the standard way to interpret data-driven models. Applied work that implement LIME for energy baseline "black-box models" is scarce, and future contributions may augment the M&V literature.

5 Conclusion

This paper demonstrates two different methods, the BL and the TVB model, to estimate the energy savings from retrofitting in 5 different Norwegian food retail stores. The technical systems in the stores were upgraded with new refrigeration, HVAC and LED lighting. The aggregated energy savings ranged from 25% to 56%, hence, substantial savings was documented. The two models used to document the savings was trained on the same data, energy consumption and outside temperature, but differed in terms of aggregation level. The TVB models was estimated on an hourly level and the BL model on a weekly level. There was practically no difference between the aggregated savings from the two different model approaches, and the precision measured from the CV-RMSE was acceptable for both models for all the buildings. The advantage of the weekly BL model is that it is easy to compare and visualize how changes in outside temperature effect the energy consumption. For instance, both the CPT values and the change coefficients can be studied across the buildings to gain a better understanding of the energy consumption in the buildings. Nonetheless, the hourly TVB model has some unique features specific for the hourly level. The savings can be studied on a detailed level: which days have the highest savings? What about night versus day? Any specific hours that perform worse? These are questions that can be answered through models on an hourly level. Hence, since

[5]Readily available in open-source software such as R, for more detail see: https://cran.r-project.org/web/packages/lime/vignettes/Understanding_lime.html

both aggregate levels give useful insight, the practical solution is for the ESCOs to use both the BL and the TVB model in ongoing retrofitting projects. Also, the literature does not seem to offer specific advice about which models is preferred to estimate energy savings, it seems a worthwhile effort to use both the BL and the TVB model on different aggregate levels. If the results from two the models support each other that gives more reliability to the results, also the aggregate levels complement each other in terms of enhanced understanding. However, if the approaches do not support each other, then that again is useful information for further investigation, most likely some data anomaly that was not foreseen.

6 Acknowledgements

This work was supported by the Norwegian Research Council (NRC) through the Industrial PhD project [grant number 283002]. Opinions expressed and conclusions arrived at are those of the authors, and are not necessarily to be attributed to the NRC. Also, thank you to the ESCO Ohmia Retail AS for access to energy data, and many helpful discussions throughout the project.

References

[1] Allan RP. Climate change 2021: The physical science basis : Working group I contribution to the sixth assessment report of the intergovernmental panel on climate change. WMO, IPCC Secretariat; 2021.

[2] EIA. The commercial buildings energy consumption survey [Internet]. 1999 [cited 2022 Jan 14]. Available from: https://www.eia.gov/consumption/commercial/data/1999/index.php?view=methodology

[3] Galvin R. Making the 'rebound effect' more useful for performance evaluation of thermal retrofits of existing homes: Defining the 'energy savings deficit' and the 'energy performance gap.' Vol. 69, Energy and Buildings. 2014. p. 515–24.

[4] Kontokosta CE. Modeling the energy retrofit decision in commercial office buildings. Energy Build. 2016 Nov;131:1–20.

[5] Molnar C. Interpretable machine learning a guide for making black box models explainable. https://leanpub.com; 2021.

[6] Grillone B, Danov S, Sumper A, Cipriano J, Mor G. A review of deterministic and data-driven methods to quantify energy efficiency savings and to predict retrofitting scenarios in buildings. Renewable Sustainable Energy Rev. 2020 Oct;131:110027.

[7] Hong T. Short term electric load forecasting [Internet] [PhD thesis]. North Carolina State University; 2010. Available from: https://repository.lib.ncsu.edu/handle/1840.16/6457

[8] Wang P, Liu B, Hong T. Electric load forecasting with recency effect: A big data approach. Int J Forecast. 2016 Jul;32(3):585–97.

[9] Hong T, Pinson P, Fan S, Zareipour H, Troccoli A, Hyndman RJ. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. Elsevier; 2016.

[10] Hong T, Pinson P, Fan S. Global energy forecasting competition 2012. Int J Forecast. 2014 Apr;30(2):357–63.

[11] Hong T, Pinson P, Fan S. Global energy forecasting competition 2012. Elsevier; 2014.

[12] Severinsen A, Hyndman RJ. Quantification of energy savings from energy conservation measures in buildings using machine learning. In: ECEEE summer study proceedings. 2019. p. 757–66.

[13] Hastie TJ, Tibshirani RJ. Generalized additive models. CRC Press; 1990.

[14] Seber GAF WCJ. Generalized additive models. New York: Wiley; 1989.

[15] Kuchenhoff H, Ulm K. Comparison of statistical methods for assessing threshold limiting values in occupational epidemiology. Computational Statistics. 1999;12:249–64.

[16] Feder PI. The log likelihood ratio in segmented regression. aos. 1975 Jan;3(1):84–97.

[17] Muggeo VMR. Segmented: An R package to fit regression models with Broken-Line relationships. R News. 1991;8:20–5.

[18] Tuominen P, Klobut K, Tolman A, Adjei A, Best-Waldhober M de. Energy savings potential in buildings and overcoming market barriers in member states of the european union. Energy Build. 2012 Aug;51:48–55.

[19] Committee I, IPMVP Committee. International performance measurement and verification protocol: Concepts and options for determining energy and water savings, volume I. 2001.

[20] Guideline A, Others. Measurement of energy, demand, and water savings. In: ASHRAE guidel 14-2014. 2014.

[21] Silva J, Brás J, Noversa R, Rodrigues N, Martins L, Teixeira J, et al. Energy performance of a service building: Comparison between EnergyPlus and revit. Computational Science and Its Applications – ICCSA 2020. 2020. p. 201–13.

[22] Fallahi Z, Smith AD. A comparison of commercial building retrofits using EnergyPlus for energy and emissions savings. Volume 6B: Energy. 2016.

[23] Haberl JS, Thamilseran S. The great energy predictor shootout II. ASHRAE J. 1998;40(1):49.

[24] Wang Z, Srinivasan RS. A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. Vol. 75, Renewable and Sustainable Energy Reviews. 2017. p. 796–808.

[25] Foucquier A, Robert S, Suard F, Stéphan L, Jay A. State of the art in building modelling and energy performances prediction: A review. Renewable Sustainable Energy Rev. 2013 Jul;23:272–88.

[26] Zhao H-X, Magoulès F. A review on the prediction of building energy consumption. Renewable Sustainable Energy Rev. 2012 Aug;16(6):3586–92.

[27] Ahmad AS, Hassan MY, Abdullah MP, Rahman HA, Hussin F, Abdullah H, et al. A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renewable Sustainable Energy Rev. 2014 May;33:102–9.

[28] Raza MQ, Khosravi A. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renewable Sustainable Energy Rev. 2015 Oct;50:1352–72.

[29] Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. Renewable Sustainable Energy Rev. 2018 Jan;81:1192–205.

[30] Yildiz B, Bilbao JI, Sproul AB. A review and analysis of regression and machine learning models on commercial building electricity load forecasting. Renewable Sustainable Energy Rev. 2017 Jun;73:1104–22.

[31] Fels MF, Others. PRISM: An introduction. Energy Build. 1986;9(1-2):5–18.

[32] Katipamula S, Reddy TA, Claridge DE. Multivariate regression modeling. J Sol Energy Eng. 1998 Aug;120(3):177–84.

[33] Mathieu JL, Price PN, Kiliccote S, Piette MA. Quantifying changes in building electricity use, with application to demand response. IEEE Trans Smart Grid. 2011;2(3):507–18.

[34] Aris SM, Dahlan NY, Nawi MNM, Nizam TA, Tahir MZ. Quantifying energy savings for retrofit centralized hvac systems at selangor state secretary complex. Jurnal Teknologi. 2015;77(5).

[35] Wang H, Xue Y, Mu Y. Assessment of energy savings by mechanical system retrofit of existing buildings. Procedia Engineering. 2017 Jan;205:2370–7.

[36] Díaz JA, Ramos JS, Delgado MCG, García DH, Montoya FG, Domínguez SÁ. A daily baseline model based on transfer functions for the verification of energy saving. A case study of the administration room at the palacio de la madraza, granada. Appl Energy. 2018 Aug;224:538–49.

[37] Brown M, Barrington-Leigh C, Brown † Z. Kernel regression for real-time building energy analysis. Journal of Building Performance Simulation. 2012 Jul;5(4):263–76.

[38] Yalcintas M. Energy-savings predictions for building-equipment retrofits. Energy Build. 2008 Jan;40(12):2111–20.

[39] Lindelöf D, Alisafaee M, Borsò P, Grigis C, Viaene J. Bayesian verification of an energy conservation measure. Energy Build. 2018 Jul;171:1–0.

[40] Betts MG, Forbes GJ, Diamond AW. Thresholds in songbird occurrence in relation to landscape structure. Conserv Biol. 2007 Aug;21(4):1046–58.

[41] Ulm K. A statistical method for assessing a threshold in epidemiological studies. Stat Med. 1991 Mar;10(3):341–9.

[42] Hastie TJ TR. Generalized additive models. London: Chapman & Hall; 1990.

[43] Hinkley DV. Inference in Two-Phase regression. Vol. 66, Journal of the American Statistical Association. 1971. p. 736–43.

[44] Muggeo VMR. Estimating regression models with unknown break-points. Stat Med. 2003;22(19):3055–71.

[45] Severinsen A, Holst HMS. Using machine learning and mathematical programming to benchmark energy efficiency of buildings. In: ECEEE SUMMER STUDY PROCEEDINGS. Hyeres, Nice: ECEEE; 2017. p. 1083–9.

[46] American Society of Heating, Refrigeration and Air Conditioning Engineers. ASHRAE guideline 14, ASHRAE guideline 14–2014 for measurement of energy and demand savings. 2014.

[47] Miller C, Arjunan P, Kathirgamanathan A, Fu C, Roth J, Park JY, et al. The ASHRAE great energy predictor III competition: Overview and results. Vol. 26, Science and Technology for the Built Environment. 2020. p. 1427–47.

[48] Tian W, Heo Y, Wilde P de, Li Z, Yan D, Park CS, et al. A review of uncertainty analysis in building energy assessment. Renewable Sustainable Energy Rev. 2018 Oct;93:285–301.

[49] Touzani S, Granderson J, Jump D, Rebello D. Evaluation of methods to assess the uncertainty in estimated energy savings. Energy Build. 2019 Jun;193:216–25.

[50] Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": Explaining the predictions of any classifier. 2016 Feb; Available from: https://arxiv.org/abs/1602.04938