Quantification of energy savings from energy conservation measures in buildings using machine learning

Alexander Severinsen
Norwegian University of Life Sciences
Universitetstunet 3
1433 Ås
Norway
as@storekeeper.no

Rob J Hyndman
Department of Econometrics & Business Statistics
Monash University
Clayton, VIC 3800
Australia
rob.hyndman@monash.edu

Keywords

energy savings calculation, IPMVP, energy saving methodology, gradient boosting

Abstract

This paper demonstrates how machine learning is used to measure energy savings from energy conservation measures (ECMs); in particular ECMs with a low expected energy saving. We develop a model that predicts energy consumption in buildings on an hourly level. The model is trained on energy data from the main meter before the ECMs took place. The model is then used to predict energy consumption after the ECMs. The difference between the prediction (the estimated energy consumption in the building given no ECMs) and the actual usage is the estimated savings. According to the International Performance and Verification Protocol (IPMVP) using data from the main meter is a recommended option when the collective savings of several ECMs are analysed, and the savings are expected to be large. For ECMs where the expected savings is less than 10 % the IPMVP recommends system simulation or installation of sub-meters to isolate the ECMs. However, when implementing smaller ECMs (<10 % expected savings) the added cost of installing sub-meters and/or undertaking system simulation could turn a positive cost-benefit analysis into negative due to the increased cost of measurement and verification. For this purpose, we show that recent developments within predictive modelling will enable the building owners to quantify energy savings from ECMs where the expected saving is less than 10 %. The model has a feature set of 32 different variables that can explain energy consumption in buildings. For example, calendar-data, minimum, maximum, and average temperatures in the past 12, 24 and 36 hours. Based on this feature set the model chooses the variables that best explain the energy consumption in each building. Results from analysis in nine Norwegian grocery stores suggests that our methods are able to detect and quantify savings from small ECMs, thus are a cost-efficient and viable alternative to simulation and installing sub-meters.

Introduction

The building segment is one of the largest global consumers of energy; between 30 and 40 % of the global energy consumption occurs in buildings (United Nations Environment Programme, 2007). Accordingly, more energy efficient buildings represent an important opportunity to reduce emissions. In a recent report by the International Energy Agency (IEA) they investigate the global potential for energy savings and find that efficiency gains alone could allow twice as much economics value from the energy it uses compared to today (IEA, 2018).

In order to reduce the environmental impact and costs associated with investing in energy efficient buildings, several energy efficiency programs have been implemented. In Norway, Enova SF (https://www.enova.no/about-enova/), owned by the Norwegian Ministry of Climate and Environment works towards reduced greenhouse gas emissions, and energy and climate technology change. Enova SF has in Norway energy efficiency programs that target both commercial and private building owners. Energy efficiency programs are often carried out through energy service companies (ESCOs) (Satchwell et al. 2010). In the energy efficiency industry, *measurement and verification (M&V)* is the practice of estimating savings from

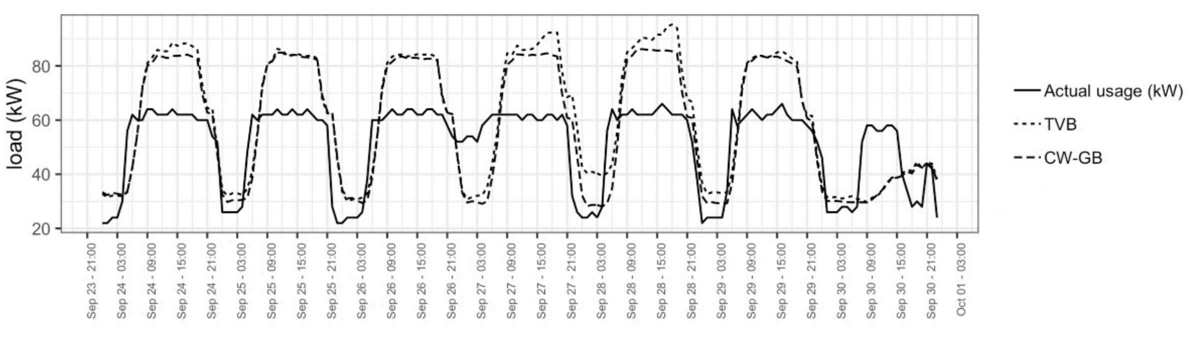

Figure 1. Comparing TVB versus CW-GB against actual energy use after ECM.

different energy conservation measures. The process is crucial for building owners, public funders and the ESCOs.

For this purpose, many ESCOs and building owners follow the Energy Valuation Organization (EVO) methods to estimating energy savings. The EVO (evo-world.org/en/about-en/) is a non-profit organization that has made *The International Performance and Verification Protocol (IPMVP)* (EVO, 2012), which is a framework to measure and verify results from energy efficiency projects. The protocol suggests both the terms and recommended methods to evaluate energy efficiency projects. According to the IPMVP using data from the main meter is a recommended option when the collective savings of several ECMs are analysed, and the savings are expected to be large. For ECMs where the expected savings is less than 10 % the IPMVP recommends system simulation or installation of sub-meters to isolate the ECMs. However, M&V is time-consuming and potentially expensive. For example, Jayaweera and Haeri (2013) finds that M&V expenses can range from 1 to 5 % of the total cost of the energy efficiency project. Further, when performing smaller ECMs (<10 % expected savings) the added cost of installing sub-meters and/or undertaking system simulation could turn a previous positive cost-benefit into negative due to the increased cost of measurement and verification.

In this paper we demonstrate that recent developments within predictive modelling may enable the building owners to quantify energy savings from ECMs where the expected saving is less than 10 % without using sub-meters or system simulation. Two different models will be used to estimate the energy savings. First, gradient boosting with component-wise p-splines (CW-GB). The CW-GB is a non-parametric additive model, with in-built variable selection, established with excellent load forecasting abilities. Second, results will be verified against an acclaimed benchmarking model, the 'Tao Vanilla' benchmark (TVB) model. Further, the CW-GB model will choose the most important variables from a set of 32 different variables that can explain energy consumption in buildings. For example, calendar-data, minimum, maximum, and average temperatures in the past 12, 24 and 36 hours. The models are applied to nine Norwegian grocery stores that completed ECMs during Spring 2018. Results suggests that the methods are able to detect and quantify savings from small ECMs and provide a cost-efficient and reliable alternative to simulation and installing sub-meters.

We start the paper with a figure that illustrates what we are trying to accomplish using load forecasting techniques. In Figure 1 we present the energy savings for a whole week in one of the grocery stores that undertook ECMs. The ECMs were a control-system to optimize energy efficiency through changes in the heat-recovery and ventilation system (controlling fan-speed and heating), cooling of cold drinks in-store, door air locks and heating cables in the entrance ramp. The average cost of the ECMs was €13,000, with €1,500 yearly operating expense. The expected energy savings was estimated to be around a 10 % reduction in energy usage *compared to not implementing the ECMs*. The ESCO made the estimate based on previous experience from other ECM projects where the energy savings were estimated using expensive measurement methods (system simulation and sub-meter energy data). Figure 1 shows the loads after implementation of the ECMs for every hour between September 24th and September 30th, 2018. The solid line shows the actual loads (kW). We can clearly see the pattern of the opening hours; the rising energy use around 07:00, and the reduction around closing hours at 21:00. The dotted lines show the CW-GB and the TVB models developed in this paper. The models were trained using data for 2017, and then the estimated energy usage was 'forecasted' for the period after the ECMs. All the ECMs were completed during Spring 2018. In Figure 1 we see that both the TVB and the CW-GB model follows each other relatively closely. The difference between the actual usage (solid line) and the two models are the estimated savings. The actual usage for the displayed week was 8,586 kWh, and the predicted usage from the TVB model was 10,208 kWh and for the CW-GB 9,851 kWh. Thus, the TVB models predict an energy saving of 16.5 % and the CW-GB of 12.8 %. On September 27th at 18:00 the actual load was 62 kW. The predicted value from the TVB model was 92 kW. This indicates energy savings of 33 % at this particular hour. Differently, the CW-GB predicted a value of 84 at that hour, thus indicating 23 % energy savings. We can see the same pattern September 27th and September 28th. Note that the actual energy consumption during non-operating hours (night-time) is much higher than the model's predictions, thus the ECMs gave an increase in energy this particular night. This could indicate a potential short-term error in the set-up of the ECM.

As we shall see later in this paper, comparing the actual loads with the predicted loads (*the loads given no ECM*) has a number of useful applications. The aggregate savings in the period after the implemented ECMs is one obvious application. However, monitoring the actual loads versus the predicted loads on an hourly level can be useful to optimize the ECMs during the phase-in period. For example, at what hours does the ECM achieve most energy savings, and is there any potential at other hours of the day/night to improve the performance of the ECMs? Also, monitoring the ECMs over time could be important to detect errors in the technical system, and in a longer term (over several years) the predictions may be used to calculate a potential decay rate of the ECMs.

The following sections describe the data that was used to train the models. Further, our modelling strategy is presented together with an overview of the literature. Finally, we provide the corresponding results from the models, some discussion and a conclusion.

Data sources

ELECTRIC LOAD AND WEATHER DATA

Hourly electricity usage is collected from electric meters from the advanced metering infrastructure (AMI) system. The values from the meters are rounded up to the nearest integer. The data is from nine food grocery stores in Norway and consists of two years of hourly data, 2017 and 2018. The training data (the reference period before the ECM took place) is year 2017. Further, all the ECM were implemented between March 4th and April 29th, 2018. The weather data was collected from the Norwegian Meteorological Service (www.met.no). Each stores longitude and latitude was mapped against a 2.5 km × 2.5 km grid of Norway.

FEATURES THAT EXPLAIN ENERGY CONSUMPTION IN BUILDINGS

Buildings energy consumption continuously changes together with differences in opening hours and holidays, and the fluctuating outside temperature. Buildings need heating when the weather is cold, and cooling when it is warm. These variables are important to understand energy consumption in buildings. Figure 2 shows the weekly energy consumption together with the average weekly temperature for one of the stores. There is a strong time-of-year effect, with peak demand during winter and increased demand during warm summer weeks.

Figure 3 shows hourly loads throughout a week. The figure reveals the morning start-up around 04:00, and the morning ramp-up that peaks around 07:00, and the evening setback that starts at off-hours at 21:00. Also, this store has closed on Sundays, where the loads fluctuates around 57 kW.

Weather and calendar data (opening hours, holidays) are crucial data to understand energy consumption in buildings. Table 1 gives a description of the potential features that might impact the energy consumption in buildings. These variables are available for the CW-GB model, and the algorithm will select the set of variables that best predict each building's energy consumption. However, the TVB model will have a fixed set of variables, as described in the next section of the paper.

Models for load forecasting

Load forecasting has several useful applications. First, forecasting may improve the understanding of how energy consumption in a building changes between years. Second, quantification of energy savings from ECMs, and third, detect anomalies. In 1986, the PRInceton Scorekeeping Method (PRISM) was introduced as a standard method to measure ECM savings (Fels and Others 1986). The PRISM is a simple piece-wise linear regression model with monthly electricity consumption and heating degree-days. As energy data became more available, models using daily and hourly data were proposed, both using multiple linear regression (Katipamula, Reddy, and Claridge 1998) and change-point models (Haberl and Thamilseran 1998). Furthermore, Claridge (1998) discusses many of these approaches, such as linear regression, simulations and neural network models. Taylor, Menezes and McSharry (2006) compare seasonal ARIMA, neural networks, double seasonal exponential smoothing, and principal component analysis (PCA) methods, each with their own strengths and weaknesses. Granderson et al. (2009) describe non-linear approaches, such as nearest-neighbor models and locally-weighted regression,

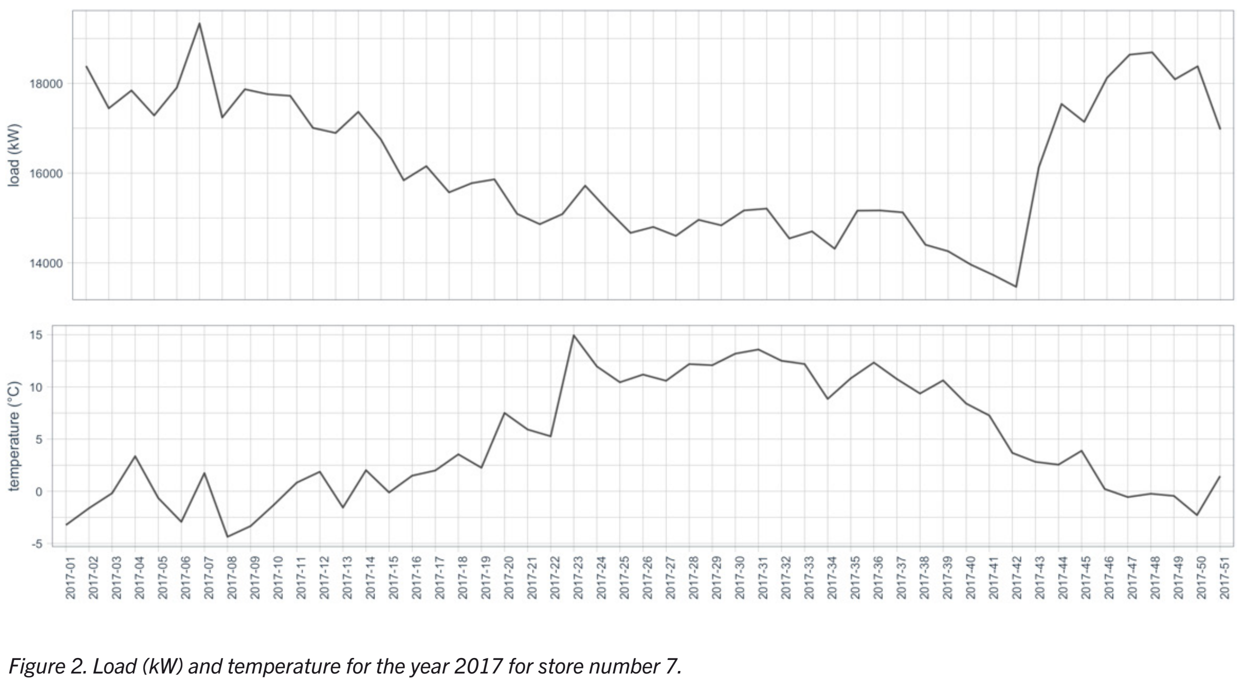

Figure 2. Load (kW) and temperature for the year 2017 for store number 7.

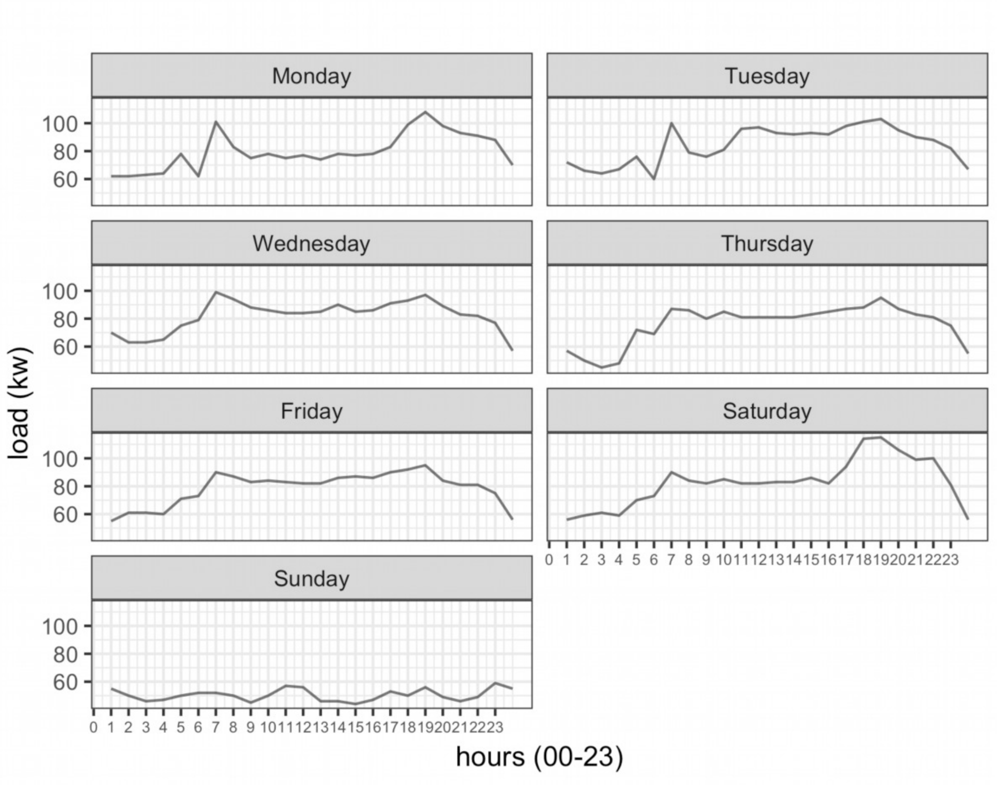

Figure 3. Hourly loads (kW) throughout a week.

Table 1. Variables used in the CW-GB model to train/learn the energy consumption in buildings (before any ECMs are implemented).

Name of predictor	Description
weekday	Weekday (Monday–Sunday)
hour	Current hour
month	Current month
weekday_x_hour	Interaction between hour and weekday
holiday	Holiday (=1 if holiday)
temps20	Outdoor temperature above 20 °C = 1, else 0
hourTemps	Interaction between hour and temperature
monthTemps	Interaction between month and temperature
temps1-temp12	12 variables, temperature lagged 1–12 hours
temp24	Temperature lagged 24 hours
temp48	Temperature lagged 48 hours
temp.avg.12h	Average temperature past 12 hours
temp.avg.1d	Average temperature past 24 hours
temp.avg.2d	Average temperature past 48 hours
temp.avg.3d	Average temperature past 72 hours
temp.avg.7d	Average temperature past 7 days
temp.24.previous	Average temperature past 24 hours, lagged 24 hours
temp.min.1d	Lowest temperature past 24 hours
temp.min.2d	Lowest temperature past 48 hours
temp.min.7d	Lowest temperature past 7 days
temp.max.1d	Highest temperature past 24 hours
temp.max.2d	Highest temperature past 48 hours
temp.max.7d	Highest temperature past 7 days

Mathieu et al. (2011) use multiple regression with a time-of-week indicator variable (similar to what we use in the TVB model described in the next section) and a piece-wise linear and continuous outdoor air-temperature dependence, while recently Touzani, Granderson, and Fernandes (2018) use gradient boosting based on decision trees. In this paper, we propose to estimate energy savings with models that were previously established to perform well in competition with other models, namely TVB and CW-GB. Both of these methods have been rigorously tested to perform well in competition with more than a 100 other models. None of these methods have, as far as we have been able to ascertain, previously been used to estimate energy savings. A more detailed introduction of the two models follows in the next two sections.

COMPONENT-WISE GRADIENT BOOSTING WITH PENALISED SPLINES

Boosting has a history of excellent prediction performance within statistics and machine learning (Schapire and Freund 2012). Further, Bühlmann and Yu (2003) developed *component-wise gradient boosting* to handle models with a large set of independent variables. In this paper we use *component-wise gradient boosting with penalised splines* (P-splines) (Bühlmann and Hothorn 2007). Boosting yields data-driven variable selection, implicit penalization and shrinkage of effect estimate. Boosting is also robust against multicollinearity and flexible in terms of modelling different types of effects (Mayr and Hofner 2018). Ben Taieb and Hyndman (2014) used CW-GB in the Kaggle global energy forecasting competition 2012, where the CW-GB ranked fourth out of 105 participating teams. The following is a more detailed overview of the applied procedure:

We label the outcome variable, energy consumption, y and the predictors (temperature variables and calendar data) $x_1, \ldots, x_p$. The objective is to model the relation between y and $X := (x_1, \ldots, x_p)^T$, and to estimate the "optimal" prediction of y given x. To achieve this objective, we minimize the loss function $\rho(y,f) \in \mathbb{R}$ over a prediction function f depending on x. Since we use a generalised additive models (GAM) the loss function is the negative log-likelihood function of the outcome distribution. In the gradient boosting the objective is to estimate the optimal prediction function f^*, defined by

$$f^* := argmin_f \mathbb{E}_{y,x}[\rho(y, f(x^T))], \tag{1}$$

where it is assumed that ρ, the loss function, is differentiable with respect to f.

1. Start the function estimate $\hat{f}^{[0]}$.
2. Determine the set of *base-learners*. Each of the base-learners act as a modeling alternative for the predictive model. We set the number of base-learners equal to P and $m = 0$.
3. Increase m by 1
   a. Compute the negative gradient $-\frac{\partial \rho}{\partial f}$ of the loss function and evaluate it at $\hat{f}^{[m-1]}(x_i^T), i = 1, \ldots, n$. This gives the negative gradient vector

   $$\mathbf{u}^m = (u_i^{[m]})_{i=1,\ldots,n} := (\frac{\partial \rho}{\partial f}(y_i, \hat{f}^{[m-1]}(x_i^T)))_{,i=1,\ldots,n} \cdot$$

   d. Fit each of the base learners individually to the negative gradient vector. Estimate the negative gradient $\mathbf{u}^m$ for all the vectors of the predicted values P.
   e. This step selects the base-learner that fits $\mathbf{u}^m$.
   f. The current estimate is updated by setting $\hat{f}^{[m]} = \hat{f}^{[m-1]} + v\hat{u}^{[m]}$ where $0 < v \leq 1$.
4. Steps 3 and 4 are iterated until m_{stop} is reached.

In step 3c) and 3d) the algorithm performs variable and model selection. There are two hyper parameters that needs to be estimated, M, the number of steps, and v, a step length factor. However, Friedman (2001) shows that a small v can prevent overfitting. We set $v = 0{,}15$ and $m = 500$.

THE TAO VANILLA BENCHMARK MODEL

The results from the CW-GB model is compared against the TVB model. This model was first published in Hong (2010) and was later used as a benchmark model in the GEFCom2012 load forecasting competition (Hong, Pinson, and Fan 2014). The model performed among the best 25 of 100 teams. Also, TVB is integrated as a standard load-forecasting model in the commercial software package SAS Energy Forecasting. The model is a multiple linear regression model

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

where Y_t is the load forecast 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 year, day of the week and hour of the day. Further, T_t is the temperature corresponding to time t. Note that the original TVB model includes trend and past loads. In this study the model will reflect how a particular building perform based on a reference period, thus trend and lagged predictors are not included.

The *Coefficient of Variation Root Mean Square Error* CV(RMSE) is used as a measure of the variability between the actual and predicted values and will be used to rank TVB versus CW-GB. CV(RMSE) is computed in the following way:

$$CV(RMSE) = \frac{\frac{\sum(\hat{Y}_i - Y_i)^2}{n-p-1}}{\bar{Y}} \tag{3}$$

where $\bar{Y}$ is the mean of the number of measured energy values in the training data, Y_i is the actual energy usage in hour i, $\hat{Y}_i$ is the predicted value of energy in hour i, n is the sample size, and p is is the number of features in the model. The models are implemented using the 'mboost' R package with 5-fold cross-validation (T. Hothorn and Hofner 2018).

Results

RELIABILITY OF THE MODELS

Table 2 shows the CV(RMSE) for both the CW-GB and the TVB, in addition to the percentage difference between the two modeling alternatives. For store number 2 both models have the same CV(RMSE) with 0.112, other than that all the CW-GB models perform better than the TVB. The average percentage improvement is 2 %, and the maximum percentage improvement is 5 %.

The American 'ASHRAE' guidelines specifies that the CV(RMSE) calculated on the training period should be less than 0.25 if 12 months of post-measure data are used (American Society of Heating, Refrigeration and Air Conditioning Engineers 2014). The results from both CW-GB and TVB are well below for all the nine stores, thus both the modelling approach performs well in terms of estimating energy savings.

Variable importance

Each store has its own "optimal" set of features chosen by the CW-GB variable selection procedure. Figure 4 plots the relative variable importance from the fitted CW-GB model for each of the stores. The 5 most important variables, after excluding *(weekday_x_hour)*, for each store are shown. The interaction variable between weekday and hour, *(weekday_x_hour)*, is excluded from the plot because it 'hides' the effect of the other variables. It is by far the most important variable to explain energy consumption because it models the stores opening hours. Investigating the other variables there is some variation in terms of what variable is most important. For four of the stores the variable '*month*' is second most important (after *weekday_x_hour*). This would likely be due to changing temperatures in the different seasons. In the other stores *temp2* (temperature lagged 2 hours, reflecting some thermal inertia in the building envelope), *holiday*, and the interaction variable '*hour, temps*' is second most important. It is somewhat surprising that '*hour, temps*' is such an important variable. However, the ventilation system is set up to run in reduced performance mode when the stores are closed. During opening hours, the system will consume more power to heat or cool air depending on the outside temperature. Interestingly, store number four has the maximum temperature the past 7 days (*temp.max.7d*) as third most important, while store number 9 has the minimum temperature past 7 and past 2 days among the most important variables.

Table 2. CV(RMSE) for VTB and CW-GB.

Store number:	CV(RMSE) CW-GB	CV(RMSE) TVB	% improvement
Store 1	0.125	0.129	3.10
Store 2	0.112	0.112	0.00
Store 3	0.086	0.087	1.15
Store 4	0.090	0.093	3.23
Store 5	0.097	0.098	1.02
Store 6	0.088	0.089	1.12
Store 7	0.133	0.134	0.74
Store 8	0.116	0.119	2.52
Store 9	0.132	0.139	5.04

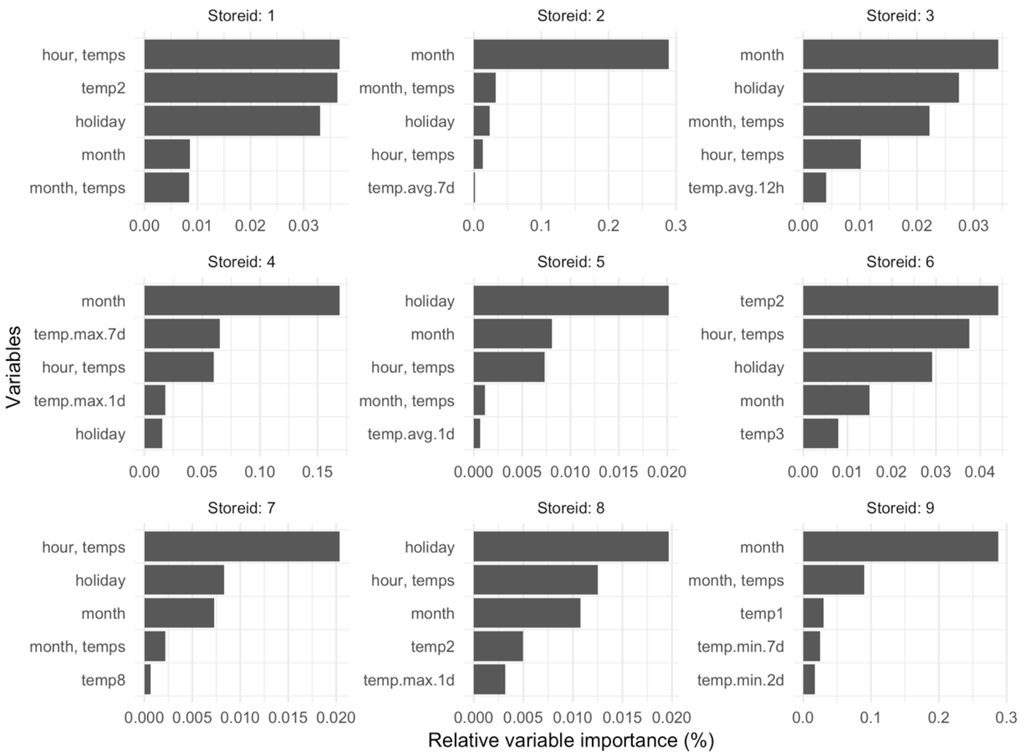

Figure 4. Relative importance of the different variables.

Table 3. Estimated savings – the difference between actual energy consumption and predicted.

Store number:	kWh 2018	CW-GB Predicted kWh 2018	TVB Predicted kWh 2018	% CW-GB Savings	% TVB Savings
Store 1	374,388	410,224	406,851	8.7	8.0
Store 2	331,275	384,743	388,220	13.8	14.7
Store 3	534,160	669,130	668,539	20.1	20.1
Store 4	489,173	548,803	551,569	10.8	11.3
Store 5	311,820	340,389	339,905	8.4	8.2
Store 6	300,065	355,427	355,618	15.6	15.6
Store 7	345,161	405,694	405,246	14.9	14.8
Store 8	280,668	331,796	330,487	15.4	15.0
Store 9	190,195	195,195	195,917	2.6	2.9

AGGREGATED ENERGY SAVINGS

In Table 3, the aggregated energy savings results for the ECMs for each of the nine stores are presented, along with actual demand (*kWh 2018*) and predicted demand in the period after the ECMs. The calculations are based on the difference between the actual demands and the predicted demand for each hour, aggregated over the entire ECMs period (from Spring 2018, up until December 2018). The average percentage reduction for the 9 stores from CW-GB is 12.28 %, while 12.3 % from TVB. There is little difference in the aggregated savings from the CW-GB and the TVB models. In store number 9 the percentage reduction in energy as a result of the ECMs is 2.6 %, while the same store had a 2.9 % reduction according to TVB. On the other hand, store number 3 had an estimated percentage reduction of 20.1 % (from both models).

Weekly energy savings

For building owners (and ESCOs) it is important to continuously monitor the impact of the ECM. Figure 5 shows a weekly savings report for the 9 stores. The figure shows the weekly percent energy reduction (based on the difference between the actual and the predicted energy consumption that week. For example, in several of the stores the % savings trend is trending upward (store number: 2, 6, 7, 8). To phase in and optimize a new control-system for heating, cooling and air locks is a continuous learning process that may take several months after launch. That may explain the improved performance over time as learning enhances the ECM project. Furthermore, both store number 9 and 4 (since week 40) is for many of the weeks performing worse than what the store would have used *if the ECM had not been implemented*. The estimated weekly savings from TVB and CW-GB follows each other closely (correlation = 0.94), but the CW-GB seems to be more 'stable' – look at the discrepancy in week 19 in store number 7, and week 31 for store number 2 and 3.

Figure 6 shows how comparing actual versus predicted loads can be used to both estimate the average energy savings across different days and hours, and to optimize the ECM. For example, the actual loads after ECM implementations has a steeper morning-ramp start up than the predicted values (05:00–08:00), indicating a further energy savings potential. Further, on Thursdays the actual loads are higher than the predicted during nighttime. Also, on Sunday afternoon the actual loads are larger than the predicted, something that might indicate an inefficient setup of the ECM these hours. Thereupon, monitoring the ECMs on an hourly level can indicate which hours that savings are greatest. For example, does the ECM provide greater impact at off-hours or operating hours, or weekdays versus weekends?

Discussion

In this study a general approach to model energy savings in buildings is developed. We demonstrate that CW-GB is a method that can reliably be used to estimate savings from ECMs with expected savings less than 10 %. The model takes into account each building's unique set of energy consumption predictors. Moreover, the approach delivers a better performance than the TVB model for all 9 stores. Nevertheless, both the TVB and the CW-GB model fulfil the requirements from the ASHRAE guideline (*CV(RMSE) < 25 %*). We find that the TVB model is less computationally expensive, while the CW-GB model, given its iterative nature (finding the set of variables that best explain energy use), takes somewhat more computation time. Still, the CW-GB for one store, with a year of training data, only takes about a minute to run on a modern computer. Thus, both models are feasible as part of the M&V process. One disadvantage with the suggested approach is that it is not possible to isolate the individual ECMs that took place. For example, was the tuning of the ventilation system a better energy efficiency measure than controlling the heating cables in the entrance ramp? To answer that question sub-meters should be installed. It may also be possible to schedule the system to turn on and off the different ECMs such that they operate individually at different days. In that way it may be possible to use the same data and approach as in this paper to analyse each ECMs separately.

As Figure 4 displayed, different buildings have different sets of features that explain the energy consumption. Given that the CW-GB model performed somewhat better than the TVB implies that it is useful to allow the modelling process to be able to choose among different features when the model is trained. Also, it could be useful to carefully investigate and compare what variables are important across different stores. For example, if it turns out that the interaction variable '*hour, temps*' has no effect; it might be worth investigating if reduced performance mode in the ventilation system is actually working properly.

Over recent years energy data from the main meters have become readily available, and many sources of meteorological weather data have become freely available. Furthermore, in Feb-

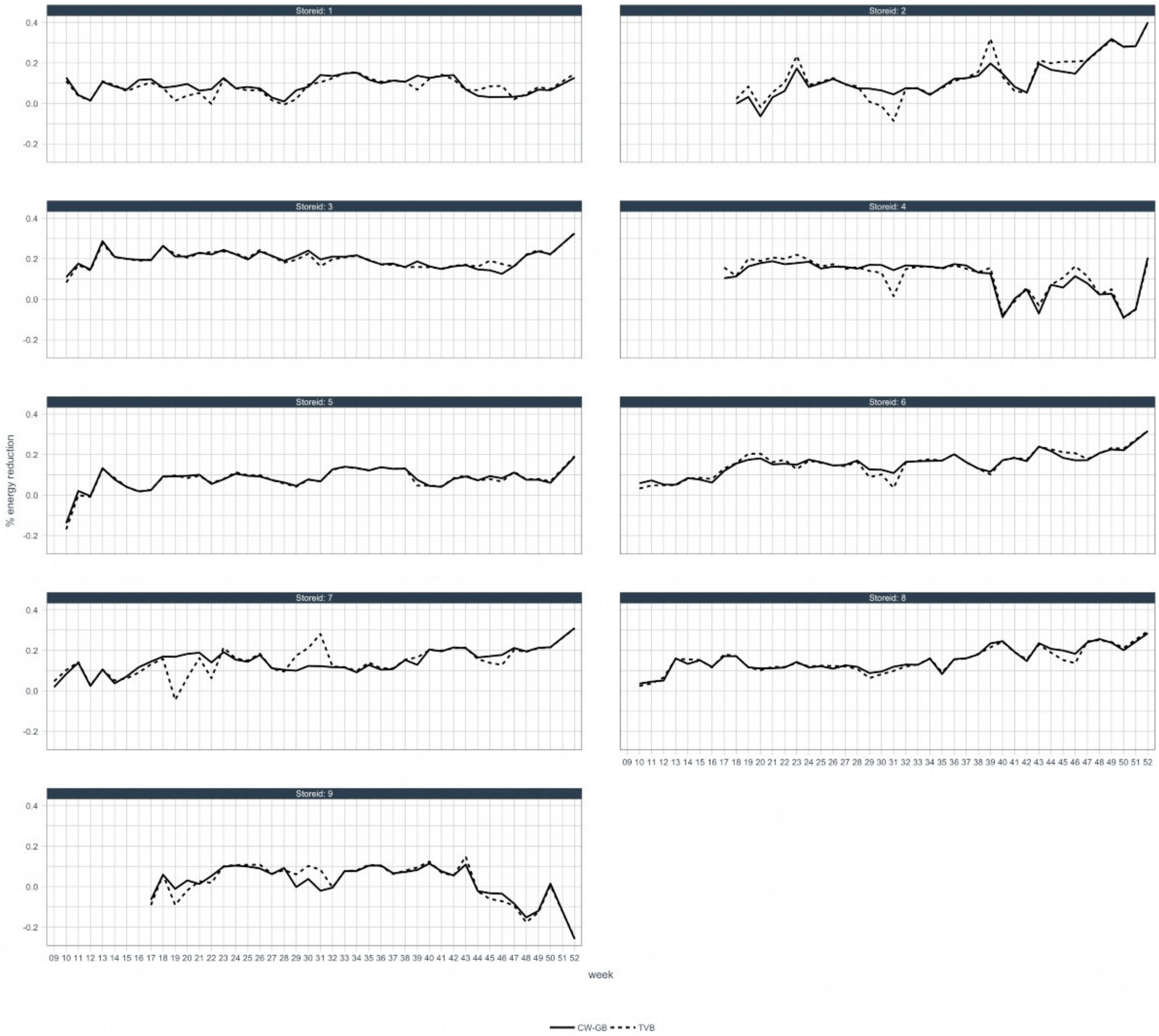

Figure 5. Estimated weekly savings for the 9 stores.

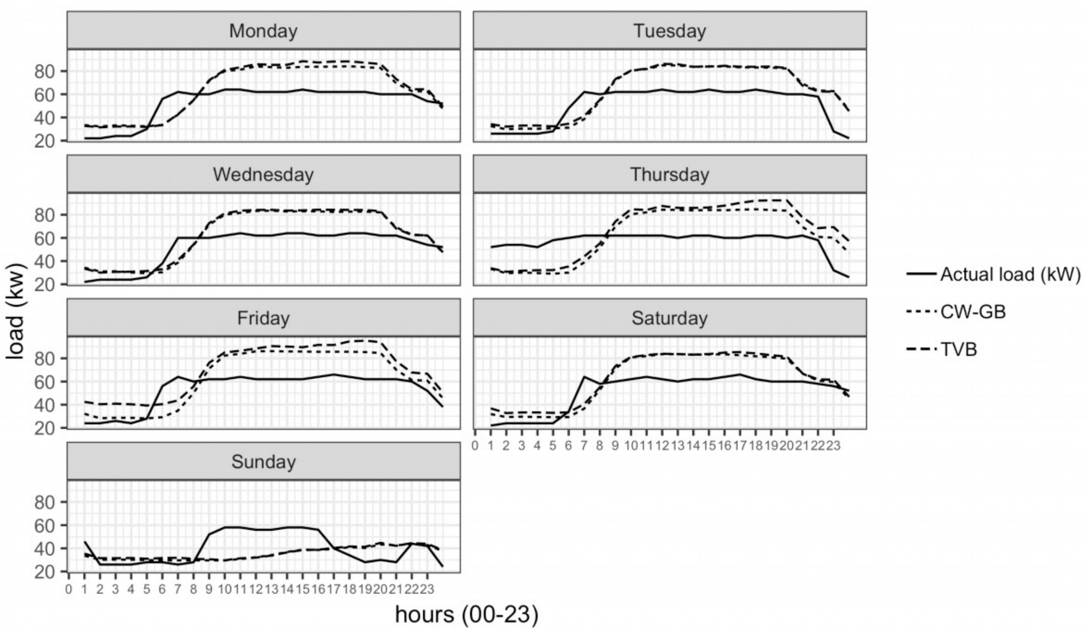

Figure 6. Actual loads and predicted loads for store number 4 in week 39.

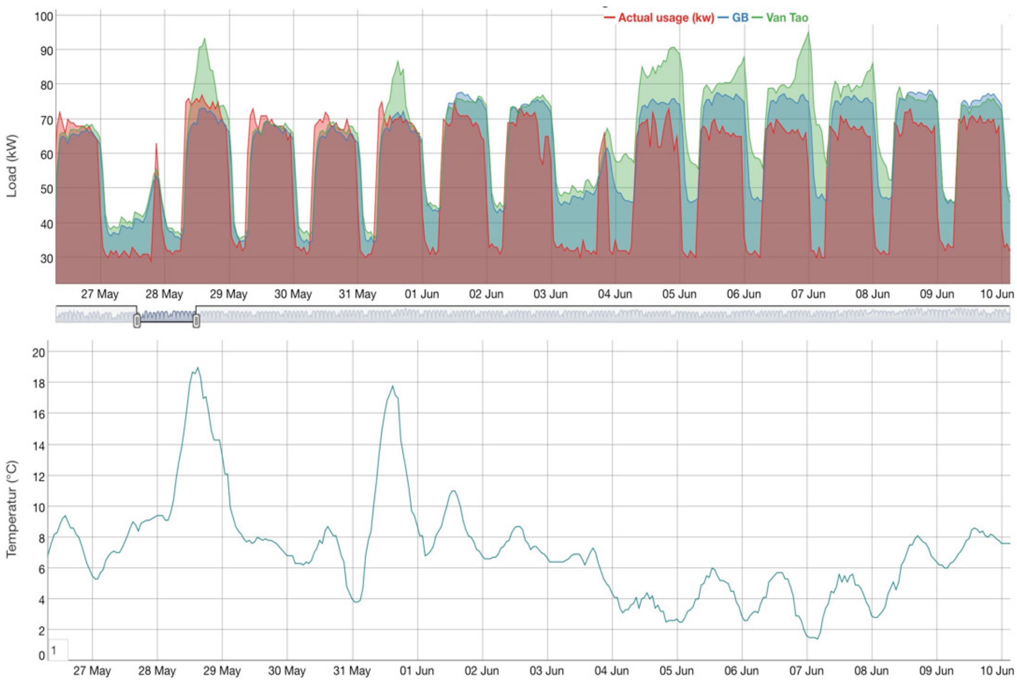

Figure 7. Example web-application. The predicted loads and corresponding temperatures.

ruary 2019, the Norwegian Elhub (elhub.no) will be launched. It is a central repository for data from all electricity meters in Norway, and all the Norwegian grid operators are required to send data to the repository on a daily basis (hourly loads). The Elhub is owned by Statnett, the system operator of the Norwegian energy system, which again is owned by the Norwegian state. The increasing availability of data from smart meters, and meteorological data such as outside temperatures, paired with new development within predictive modelling, has given new approaches to reliably estimate energy savings from ECMs. Table 2 in the result section shows that it is possible to detect energy savings using both the TVB and the CW-GB model. The CW-GB turned out slightly more precise than the TVB from a CV(RMSE) perspective, but both models performed well enough to reliably estimate savings (according to the ASHRAE guidelines). Further, the availability of data (updated every day) has also made it possible to continuously 'score' the actual loads with predicted loads. This again can be used to set up a web-based monitoring system to estimate energy savings, optimize ECMs and detect anomalies. Figure 7 shows an example of a web-based application where the user can 'zoom' in on the data and systematically get an overview of the performance of the ECMs. Future research will explore more of these opportunities, including automatic error detection.

Note, that we have approached the modelling task a bit differently from the classical train-test development of models. For example, when Touzani, Granderson, and Fernandes (2018) models the energy savings using a gradient boosting model with a decision tree, the model was developed on two years of data prior to the ECM. The model was trained on one year of data two years prior to ECM and tested on data one year before the ECM. The model was further used to predict the loads. This approach is very sensible for a modelling perspective. To prevent overfitting the model is trained, then tested, and at last applied in production for prediction purposes for time-series that was not part of the model development. However, in the grocery sector, and in particular in our study, many of the stores had other energy consuming activities (in-store promotion, and smaller ECMs) that varied between 2016 and 2017, thus making the model testing with 2017 data infeasible. Nevertheless, the gradient boosting model in this paper uses 5-fold cross-validation within the training data and the TVB model uses a fixed set of variables. This reduces the chances of overfitting.

Conclusions

A trustworthy process of M&V is important to understand and improve energy efficiency measures. This paper has demonstrated that both the TVB and the CW-GB can be used to estimate energy savings from ECMs with expected energy savings around 10 %. In many ECM projects methods such as system simulation and installing sub-meters have been used to estimate energy savings. However, these methods are potentially expensive and time consuming. The methods demonstrated in this paper have practical value for ESCOs and buildings owners to provide proof of energy savings achieved and contribute with information that can be used to optimize the ECMs on an hourly level. The methods are based on readily available data from smart meters and freely available meteorological data. Also, from a Norwegian perspective, the launch of Elhub (a central repository for all smart meters in Norway) could contribute to better access, data quality and increased use of the data to analyze ECMs.

References

American Society of Heating, Refrigeration and Air Conditioning Engineers. 2014. *ASHRAE Guideline 14, Ashrae Guideline 14–2014 for Measurement of Energy and Demand Savings.*

Ben Taieb, Souhaib, and Rob J Hyndman. 2014. "A Gradient Boosting Approach to the Kaggle Load Forecasting Competition." *International Journal of Forecasting* 30 (2). Elsevier: 382–94.

Bühlmann, Peter, and Torsten Hothorn. 2007. "Boosting Algorithms: Regularization, Prediction and Model Fitting." *Stat. Sci.* 22 (4). Institute of Mathematical Statistics: 477–505.

Bühlmann, Peter, and Bin Yu. 2003. "Boosting with the L2 Loss: Regression and Classification." *J. Am. Stat. Assoc.* 98 (462). Taylor & Francis: 324–39.

Claridge, D E. 1998. "A Perspective on Methods for Analysis of Measured Energy Data from Commercial Buildings." *J. Sol. Energy Eng.* 120 (3). American Society of Mechanical Engineers: 150–55.

EVO. 2012. *International Performance Measurement and Verification Protocol*.

Fels, Margaret F, and Others. 1986. "PRISM: An Introduction." *Energy Build.* 9 (1–2). Elsevier Science: 5–18.

Friedman, Jerome H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." *Ann. Stat.* 29 (5). Institute of Mathematical Statistics: 1189–1232.

Granderson, Jessica, Mary Ann Piette, Girish Ghatikar, and Phillip Price. 2009. "Building Energy Information Systems: State of the Technology and User Case Studies." *Handbook of Web Based Energy Information and Control Systems*.

Haberl, J S, and Sabaratnan Thamilseran. 1998. "The Great Energy Predictor Shootout II." *ASHRAE J.* 40 (1). American Society of Heating, Refrigeration; Air Conditioning Engineers, Inc.: 49.

Hong, Tao. 2010. "Short Term Electric Load Forecasting." PhD thesis, North Carolina State University. https://repository.lib.ncsu.edu/handle/1840.16/6457

Hong, Tao, Pierre Pinson, and Shu Fan. 2014. "Global Energy Forecasting Competition 2012." *Int. J. Forecast.* 30 (2): 357–63.

IEA. 2018. "Energy Efficiency 2018. Analysis and Outlooks to 2040." https://webstore.iea.org/download/direct/2369?filename=market_report_series_energy_efficiency_2018.pdf

Jayaweera, Tina, and Hossein Haeri. 2013. "The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures." *Contract* 303. forum.cee1.org: 275–3000.

Katipamula, S, T A Reddy, and D E Claridge. 1998. "Multivariate Regression Modeling." *J. Sol. Energy Eng.* 120 (3). American Society of Mechanical Engineers: 177–84.

Mathieu, Johanna L, Phillip N Price, Sila Kiliccote, and Mary Ann Piette. 2011. "Quantifying Changes in Building Electricity Use, with Application to Demand Response." *IEEE Trans. Smart Grid* 2 (3). IEEE: 507–18.

Mayr, Andreas, and Benjamin Hofner. 2018. "Boosting for Statistical Modelling-a Non-Technical Introduction." *Stat. Modelling* 18 (3–4). SAGE Publications India: 365–84.

Satchwell, A, C Goldman, P Larsen, D Gilligan, and T Singer. 2010. "A Survey of the US ESCO Industry: Market Growth and Development from 2008 to 2011." pubarchive.lbl.gov.

Schapire, Robert E, and Yoav Freund. 2012. "Boosting: Foundations and Algorithms (Adaptive Computation and Machine Learning Series)." The MIT Press.

Taylor, James W, Lilian M de Menezes, and Patrick E McSharry. 2006. "A Comparison of Univariate Methods for Forecasting Electricity Demand up to a Day Ahead." *Int. J. Forecast.* 22 (1): 1–16.

T. Hothorn, T. Kneib, P. Buehlmann, and B. Hofner. 2018. *Mboost: Model-Based Boosting, R Package Version 2.9–1*. https://CRAN.R-project.org/package=mboost.

Touzani, Samir, Jessica Granderson, and Samuel Fernandes. 2018. "Gradient Boosting Machine for Modeling the Energy Consumption of Commercial Buildings." *Energy Build.* 158 (January). Elsevier: 1533–43.

United Nations Environment Programme. (2007) Sustainable buildings and construction initiative buildings and climate change: status challenges and opportunities. Paris: UNEP.