Norwegian University of Life Sciences
School of Economics and Business

Philosophiae Doctor (PhD)
Thesis 2022:60

Increased sustainability in the food retail sector through Measurement and Verification of energy conservation measures

Økt bærekraft ved Måling og Verifikasjon av energitiltak i dagligvaresektoren

Alexander Severinsen

Increased sustainability in the food retail sector through Measurement and Verification of energy conservation measures.

Økt bærekraft ved Måling og Verifikasjon av energitiltak i dagligvaresektoren

Philosophiae Doctor (PhD) Thesis

Alexander Severinsen

Norwegian University of Life Sciences

School of Economics and Business

Ås/Adamstuen (2022)

Thesis number 2022:60

ISSN 1894-6402

ISBN 978-82-575-2009-0

Contents

Acknowledgements

I would like to thank Petter Arnestad for brilliant ideas and discussions. After working and running a business together for 8 years I am proud to be your colleague and friend. The thesis would not have been started or finished without you. Off we go - Hardangervidda next!

My supervisors, Øystein Myrland and Atle Guttormsen, thank you for invaluable suggestions and insights. Also, I would like to note that Øystein Myrland was my supervisor when I took my masters degree in Fishery Science in the year of 2000, and later my supervisor when I failed a previous attempt (in another life) at a PhD (in another topic). In more than 20 years I can not remember Øystein ever being impatient. His kind, knowledgeable, and giving person is an important part of where I am today. Thank you! (I do however think it's time I got a taste of some of that *Brinken* brewery stuff!)

In 2019 I presented two of the papers at the ECEEE Summer Study Conference in France. During the conference I had many interesting and thought-provoking suggestions from participants. Some of the most thought-provoking (or was it the interesting ones!?) might have been slightly impacted by those French that insisted on wine for lunch. Anyway - thank you! For future reference, please also note that some of the best Pétanque players are likely to be Norwegians. No pun intended. I would like to thank my co-authors. Øystein Myrland, getting me on track when stuck. Rob Hyndman for ideas and discussions. Helen Marita Sørensen Holst, me and you and an impressive strike of enjoyable meetings, some of them even about research, aka. "The procrastinators" - we did eventually finish that paper! Thank you! Jonathan Fagerström, Josefine Selj, and Kari Aamodt Espegren at the Institute of Energy Technology for working together on the PV and battery paper. Thanks for sharing ideas, and for a friendly and giving collaboration. Without collaborators research is a lonely place and less giving in all the senses of the word.

I would like to thank Trønder Energi (my current employer) for the flexibility and the possibility to finish this project within working hours - and in particular the AI department for the encouragement and support to finish! Further, thank you to the Norwegian University of Life Sciences (NMBU) for accepting me, the office, the flexibility, and the friendly environment.

Additionally, I would like to thank open-source contributors around the world! This entire thesis was written in R-Studio using R-markdown, and all the analysis was conducted on top of other peoples shoulders (read: R-packages). There is no way I would have been able to finish, pretty much anything work related, without the work of those invaluable contributions.

Moreover, the web application that was developed as part of this thesis would not have been born without the help of the excellent programmatic hands of Nichlas Severinsen (formerly part of The Storekeepers) and the genius code from Stian Berger at the northerly Tromsø consulting firm Capia. Not to forget Einar himself (aka. the Trønder who became one of The Storekeepers). Einar, without your overview and architectural and programming skills the data used in this thesis and in the web application would have been very difficult to acquire. Thanks! And good luck with your “new” life as a consultant.

Thank you to the ESCO Ohmia Retail and our customers. The ideas within this thesis has come across as part of a joint collaboration in an environment where sharing knowledge is an integral part of day to day work.

This work was supported by the Norwegian Research Council through the Industrial PhD project (grant number 283002). Without this funding a small company like Storekeeper would have never been able to justify the time and effort that went into this work. Thank you!

My mom. Most likely she will not read this, and I am fine with that! For 21 years my mom kept my master thesis on her night-shelf - between knitting and keeping the family together she occasionally comments: “I think it´s a really interesting topic - I will definitely read it. Soon”. Last but not least, I wish to thank my wonderful wife Linda for being who you are and for endless patience. Thank you for your support, encouragement and love. I promise this will be my only and my last PhD! And the kids! Rest assured, there is no need to worry anymore about your grey haired father at an age of almost 50 still going to school! :-)

1 Abbreviations and definitions

Table 1: Abbreviations and definitions

Abb	Explained
AMI	Advanced metering infrastructure
EVO	Energy Valuation Organization
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
IEA	International Energy Agency
IPMVP	The International Performance and Verification Protocol
TVB	Tao Vanilla Benchmarking model
BL	Broken line model
CW-GB	Component-wise gradient boosting
DEA	Data Envelopment Analysis
CPT	Changing point temperature
ESCO	Energy Service Company
ECM	Energy conservation measure
CV-RMSE	Coefficient of variation root mean square error
GHG	Green house gas
HVAC	Heating, ventilation, and air-conditioning
PV	Photovoltaic panels
EMS	Energy management system
M&V	Measurement and verification
NPC	Net present cost
P&R	Prediction and recommendation
LCOE	Levelized Cost of Electricity
IPCC	Intergovernmental Panel on Climate Change

2 List of papers

Severinsen, Alexander, and Øystein Myrland. 2022. *"Statistical learning to estimate energy savings from retrofitting in the Norwegian food retail market"*. Renewable and Sustainable Energy Reviews, Volume 167, 112691. https://doi.org/10.1016/j.rser.2022.112691.

Severinsen, Alexander, and Rob J. Hyndman. 2019. *"Quantification of Energy Savings from Energy Conservation Measures in Buildings Using Machine Learning"*. In ECEEE Summer Study Proceedings, 757–66, https://www.eceee.org/library/conference_proceedings/eceee_Summer_Studies/2019/4-monitoring-and-evaluation-for-greater-impact/quantification-of-energy-savings-from-energy-conservation-measures-in-buildings-using-machine-learning/

Severinsen, Alexander, and Øystein Myrland. 2022. *"ShinyRBase: Near real-time energy saving models using reactive programming"*. Applied Energy, Volume 325, 119798. https://doi.org/10.1016/j.apenergy.2022.119798

Severinsen, Alexander, and Helen Marita Holst Sørensen. 2022. *"A 3-step framework to benchmark potential and actual energy savings in retrofitting projects"*. Currently under review in Sustainable Cities and Society.

Fagerström, Jonathan, Kari Aamodt Espegren, Josefine Selj, and Alexander Severinsen. 2019. *"Forecasting and Technoeconomic Optimization of PV-Battery Systems for Commercial Buildings."* In ECEEE Summer Study Proceedings, 949–54, https://www.eceee.org/library/conference_proceedings/eceee_Summer_Studies/2019/5-smart-and-sustainable-communities/forecasting-and-technoeconomic-optimization-of-pv-battery-systems-for-commercial-buildings/

3 Abstract

The IPCC Sixth Assessment Report leaves little doubt that we urgently need to respond to be able to reduce human-induced climate change. The report clearly states that human activities is causing alarming and widespread disruption in nature and is affecting billions of people. Floods, heatwaves, and droughts are seen more often than ever, and unfortunately, people who are least able to struggle through are most affected. To avoid ascending loss of life, infrastructure, and biodiversity we have to quickly make major cuts in greenhouse gas emissions (GHG) [1].

Buildings worldwide consume some 40% of all produced energy and are significant contributors to GHG emissions. Hence, energy efficiency retrofitting is a fundamental step in reducing energy consumption. However, one important barrier that hinders renovation projects is uncertainty regarding the expected savings. The main objective of this thesis is to contribute to lower that barrier and to deliver reliable methods to be used to document and monitor energy savings in retrofitting projects. To accomplish this objective, we present 5 different papers.

In the first article, "*Statistical learning to estimate energy savings from retrofitting in the Norwegian food retail market*", we demonstrate two different statistical methods to estimate energy savings. The first method is the Tao Vanilla benchmarking method (TVB). The TVB has previously received a lot of attention within the load forecasting literature, and we argue that with its simple and straightforward specification it should gain more use within the energy and building sector. The TVB model predicts energy consumption on an hourly level. We further suggest using the Broken line (BL) model. The BL model use data on a weekly level, and its underlying logic resembles the approach that many practitioners use when they estimate energy savings in retrofitting projects. The two methods are applied on 5 different food retail stores that undertook a retrofitting project in 2021. The results from the retrofitting projects demonstrates considerable energy savings between 25% and 56%. Furthermore, 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.

While the approach in the first paper was based on a linear regression framework, the second article, "*Quantification of energy savings from energy conservation measures in buildings using machine learning*" takes a somewhat more advanced approach. The paper demonstrates how component-wise gradient boosting (CW-GB) and the TVB model performs to estimate energy savings in ECMs with low expected savings, typically below 10 percent. Often the energy savings ECMs with low expected savings are estimated using either

[1] https://cran.r-project.org/web/packages/segmented/NEWS

simulation or sub-meters, approaches that add costs to the measurement and verification (M&V) process. In the proposed CW-GB and TVB models we use readily available data on an hourly level from the main meter (AMI). The results show that both the TVB and the CW-GB model deliver reliable results. One find that the CW-GB model has a slightly better predictive power measure through CV-RMSE, and that the model gives more detailed insight into what variables are most important to explain energy consumption in the different buildings.

This thesis was written as part of an industrial PhD project in close cooperation with a energy service company (ESCO). The ESCO is specialized in retrofitting food retail stores. Furthermore, often the ESCO's work were for quite large building portfolios (>30 buildings), and within an energy performance contract (EPC) setting. The EPC made it particularly important to deliver reliable (and understandable) baseline models. Also, because the retrofitting was done on many buildings at the same time, the request for baseline models was extensive. The dependency on the energy analyst that undertook the energy savings analysis may even have sparked some frustration from both the ESCO and the analyst himself! Based on work from paper 1 and paper 2 it was decided to operationalize the TVB and the BL model into a web application such that the ESCO could have current and updated baseline models at hand when needed and be self-sufficient in terms of energy analysis. This process motivated the third article "*ShinyRBase: Near real-time energy saving models using reactive programming*". This paper demonstrates how energy savings from retrofitting's in the Norwegian food retail sector is continuously monitored and documented in a web application. The application is built using open-source tools where the baseline model is delivered through a reactive programming framework. The web application framework allows for a fast development cycle without any need-to-know web programming languages like HTML, CSS or JavaScript. The reactive framework delivers several advantages. First, the stakeholders will always have a current and real-time report on the savings. Second, complex methodologies are dynamically used by the end-user. Third, increased involvement by stakeholders and interaction with the analyst related to the methods used in the energy savings analysis lead to collaborative benefits such as faster disseminating of knowledge. These synergy effect leads to a better technical understanding from the end user perspective and enhanced practical understanding for the analyst. Today, the application is used to document and monitor ECMs in several hundred food retail stores. Each day, the results are used to continuously document energy savings, optimize existing ECMs, and to detect errors in the technical infrastructure.

The M&V process is conducted after the ECMs are installed. At the same time, the results from baseline modeling, such as the BL model, may be useful before implementation of the ECMs. In the BL model the changing point temperature (CPT), and the demand for cooling and heating could be useful input when the ESCO performs the audit phase. This phase typically consist of a complete review of all the buildings technical infrastructure. In the fourth article, "*A 3-step framework to benchmark potential and actual energy savings in retrofitting projects.*" we take advantage of the output from the BL models, and conduct a data envelopment analysis. The paper demonstrates a benchmarking framework to document the effect of energy savings and efficiency from retrofitting 34 Norwegian food retail stores. The implemented ECMs consisted of a mix of change of lightning, new coolers, ventilation, freezers, and optimization of the technical control system. The results show that the output from the BL model is useful when benchmarking energy efficiency with DEA (using opening hours, and size of the buildings). Analysis during the audit phase is often refered to as "Prediction and Recommendation" (P&R), as opposed to the M&V that comes after implemetation of the ECMs. The collaboration with the ESCO showed that very little analytically resources went into the audit phase, and only after implementing the ECMs the demand for baseline models occurred. The ESCO had substantial knowledge about the technical infrastructure, however, when benchmarking the energy efficiency the only performance indicator used was energy intensity (kWh/m2). The fourth paper finds that it may be useful to extend this perspective when benchmarking energy efficiency in buildings. Our three-step benchmarking framework offers a tool that the ESCOs can apply to document energy efficiency and energy savings documentation. The results from the proposed framework showcase the advantages of different aggregate levels with the duality of actual savings versus the efficiency; we find the methods to be valuable tools to monitor efficiency and savings throughout the retrofitting project.

In Norway there is a three-part electricity tariff. Electricity cost is divided between a fixed installation cost (EUR/ installation), consumed electricity (EUR/kWh), and demand charges for capacity usage (EUR/kW). To reduced demand charges, industrial customers are looking into supplementing PV installations with batteries to more efficiently reduce peak electricity demand, e.g. peak shaving. The objective of the fifth article, "*Forecasting and technoeconomic optimization of PV-battery systems for commercial buildings*" was to investigate the profitability with peak shaving in Norway for a commercial building. A forecasting algorithm for load prediction was developed, and the economic value of forecasting was determined for a PV-battery system. The load forecasting was developed using component-wise gradient boosting and the results from the model were verified against the TVB model. The economic value of forecasting was determined through simulations with Homer Energy Software that optimizes the net present cost of the systems. The results showed that battery storage was only economically beneficial when forecasting was deployed. Moreover, the cost savings came mainly from reduced demand charges, not from increased self-consumption of PV electricity.

4 Norsk sammendrag

FNs klimapanels siste rapport (AR6)[2] er forstemmende lesning, men også, til forskningsrapport å være, usedvanlig tydelig. I rapporten har 700 eksperter fra 90 forskjellige land bidratt og blitt enige. Kort oppsummert - om vi ønsker å begrense den globale oppvarmingen til 1,5 grad trenger vi kraftige utslippskutt i alle sektorer - umiddelbart. FNs generalsekretær, António Guterres, kaller den siste klimarapporten en «skammens rapport». For å sitere Generalsekretærens relativt umilde beskrivelse av dagens situasjon[3]

> Det er på tide å slutte å brenne planeten, og starte å investere i den rikelige fornybare energien rundt oss

Og om dagens regjeringers og selskapers innsats…

> …en katalog av tomme løfter som setter oss på sporet av en verden der det ikke går an å leve

På global basis forbruker bygninger omtrent 40% av all produsert energi, og er dermed en vesentlig bidragsyter til utslipp av klimagasser. Denne avhandlingen viser at i norsk dagligvare er det mulig å oppnå en energireduksjon på mellom 30 til 55% gitt at det installeres ny teknisk infrastruktur i byggene. I all hovedsak handler denne avhandlingen om hvordan det på en pålitelig måte er mulig å dokumentere og monitorere energireduserende tiltak. Tidligere forskning har vist at usikkerhet rundt oppnådd energibesparelse er en barriere for iverksetting av nye energitiltak. Alle de 5 artiklene i denne avhandlingen omhandler ulike metoder som kan brukes for å redusere denne barrieren og bidra positivt til økt insentiv for mer energieffektive bygg innen norsk dagligvare.

Artikkel 1 *«Statistical learning to estimate energy savings from retrofitting in the Norwegian food retail market"* demonstrerer to ulike metoder for å estimere energibesparelser; broken line (BL) og Tao Vanilla benchmarking metoden (TVB). Hensikten med BL er å videreutvikle det som i Norge er den tradisjonelle tilnærming med bruk av Energi-temperatur kurver (ET-kurver) som dokumentasjon for energibesparelser. Videre, for BL metoden er data typisk på et ukentlig nivå. Dataene finnes derimot på timenivå, og det er interessant å forstå likheter og ulikheter mellom disse nivåene når energibesparelsen skal dokumenteres. Vi anvender derfor TVB modellen på timenivå. For å demonstrere BL og TVB modellene tar vi utgangspunkt i data fra et nylig gjennomført renoveringsprosjekt for 5 dagligvarebutikker. Disse butikkene fikk i løpet av høsten 2020 gjennomført omfattende endringer i teknisk infrastruktur. Våre analyser dokumenterer

[2] https://www.ipcc.ch/report/ar6/wg2/
[3] https://www.nrk.no/klima/fns-klimapanel__-__-det-er-na-eller-aldri-1.15920016

besparelser i 2021 fra 25% til 56%; en betydelig reduksjon i energibruken. Vi finner ingen forskjeller av betydning mellom estimert energibesparelse fra BL og TVB modellen, noe som gir resultatet økt pålitelighet. Videre finner vi at det virker hensiktsmessig å analysere energibesparelsen på både uke- og timenivå. Ukenivå gir et overordnet innblikk i hvordan bygget reagerer på temperaturendringer, noe som er viktig når det planlegges energireduserende tiltak, mens timenivå gir detaljert informasjon om når på døgnet tiltakene virker best. Dette er viktig informasjon for optimalisering av tiltakene under installasjon og innfasing av ny teknisk infrastruktur, samt for å kunne identifisere mulige avvik underveis. Metodene som foreslås forenkler estimering av energibesparelser, og bidrar til økt presisjon og økt innsikt for renoveringsprosjektene.

Artikkel 2 *"Quantification of energy savings from energy conservation measures in buildings using machine learning"* demonstrerer hvordan maskinlæring kan brukes til å estimere energibesparelser som et resultat av energireduserende tiltak i dagligvarebutikker, typisk tiltak som har en relativt lav forventning til reduksjon i energibruk (<10%). Om man følger en internasjonal standard fra IPMVP brukes ofte simulering eller undermålere (f. eks direkte måling av energibruk for belysning eller ventilasjon) for å måle energibesparelsen for denne type tiltak. Både installasjon av undermålere og simulering kan være både tid- og kostnadskrevende. Vi viser i denne artikkelen at et godt alternativ er å gjennomføre analyse basert på timedata levert fra byggenes hovedmåler. Vi bruker data fra 11 dagligvarebutikker som fikk gjennomført mindre energitiltak i 2018. Våre resultater viser at både tradisjonelle lineære metoder og maskinlæring gir pålitelige resultater for estimering av energibesparelser, også når tiltakene gir mindre enn 10% reduksjon i energibruken.

I artikkel 3 *"ShinyRBase: Near real-time energy saving models using reactive programming"* vises det hvordan det er mulig, i nær sanntid, å monitorere og dokumentere energibesparelser som et resultat av energitiltak ved å designe en webapplikasjon basert på open-source og et reaktivt programmeringsparadigme. Designet og produksjonssettingen av denne applikasjonen er en viktig komponent i det resultatet som arbeidsgiver forventet levert som en "gjenytelse" for å la en av sine ansatte forske i en 75% stilling i 4 år. Metodisk brukes TVB og BL modellen, slik som beskrevet i artikkel 1 som et utgangspunkt. Disse metodene programmeres og operasjonaliseres inn i et energiovervåkningssystem (EOS) på en slik måte at energibesparelsene i ulike tiltak alltid vil være oppdaterte. Rammeverket som presenteres er svært godt mottatt blant utviklere og kunder. I dag er denne funksjonaliteten satt i drift og overvåker hver dag energitiltak i flere hundre ulike dagligvarebutikker. Resultatene fra applikasjonen blir brukt til å dokumentere energibesparelser i energisparekontrakter, optimalisere innfasing av ny infrastruktur og til tidlig detektering av potensielle feil i tekniske anlegg.

I artikkel 4 *"A 3-step framework to benchmark potential and actual energy savings in retrofitting projects"* foreslår vi et rammeverk for å benchmarke energibesparelser og energieffektivitet. Det finnes en rekke studier som studerer både besparelser og effektivitet knyttet til implementering av energireduserende tiltak, men eksisterende forskning tar ofte utgangspunkt enten i oppnådd energireduksjon eller energieffektivitet. I vår studie presenterer vi et samlet perspektiv. Vi starter med å gjenbruke TVB og BL modellen fra artikkel 1, og estimerer både temperatur-knekkpunktene hvor byggene skifter mellom et behov for oppvarming og kjøling, og graden av kjøle- og oppvarmingsbehov. Disse estimatene er, sammen med byggenes størrelse og åpningstider, input variabler i en data envelopment analysis (DEA). Vi anvender metoden på 37 dagligvarebygg som gjennomførte en rekke energisparende tiltak høsten 2020. Vi kalkulerer effektivitetscorene både før og etter tiltakene. Videre anvendes DEA multiplier modellen, en metode vi ikke har funnet benyttet i liknende studier, for å studere hvordan inputvariablene påvirker effektivitetsscorene før og etter tiltakene. Resultatene fra studien viser at rammeverket som vi presenterer bidrar til økt innsikt i hvilke energireduserende tiltak som fungerer best, både på et detaljert og et aggregert nivå.

I artikkel 5 *"Forecasting and technoeconomic optimization of PV-battery systems for commercial buildings"* undersøker vi lønnsomheten med å bruke batterier og PV for å redusere energilastene ("peak shaving") i et industribygg. Tariffene, selv om de varierer litt mellom ulike netteiere, er i all hovedsak designet på en slik måte at byggeier betaler mer for den maksimale lasten i en gitt måned. Det er derfor interessant å undersøke potensielle metoder for å kutte lastene; derav bruken av batteri og solcellepaneler (PV). Både TVB og CW-GB modellen (som anvendt i artikkel 2) blir anvendt for å prognostisere maksimale laster. Prognosene er videre brukt som signaler i et simuleringsprogram for å optimalisere bruken av batterier og PV for å avlaste energibruk i perioder med høyest etterspørsel. Resultatene viser at batterilagring kun var økonomisk forsvarlig når det ble anvendt prognosemodeller.

5 Synopsis

Uncertainty about potential energy savings from retrofitting is a known barrier that hinders new renovation projects. The purpose of this thesis is to contribute to lower that barrier, increase the attention for energy conversation measures, and eventually promote action to create more energy efficient buildings - important components to reduce GHG emissions.

The main objectives of this thesis is to

1. Demonstrate different methods that can be used to document energy savings as a result of retrofitting projects in the Norwegian food retail sector
2. Develop and set into production in near real-time a web-application (without any knowledge of web development) that can use the methods from *(1)* to continuously document and monitor energy savings, optimize ECMs, and potentially detect errors in the buildings technical infrastructure
3. Combine the methods from (1) with DEA to present a framework that can be used to study the efficiency in retrofitting projects

5.1 Introduction

The IPCC Sixth Assessment Report leaves little doubt that we urgently need to respond to be able to reduce human-induced climate change. The report clearly states that human activities is causing alarming and widespread disruption in nature and is affecting billions of people. Floods, heatwaves, and droughts are seen more often than ever, and unfortunately, people who are least able to struggle through are most affected. To avoid ascending loss of life, infrastructure, and biodiversity we have to quickly make major cuts in greenhouse gas emissions (GHG) [4].

Buildings worldwide consume some 40% of all produced energy and are significant contributors to GHG emissions. If we investigate the different building categories, we find that 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 (EIA 1999). Furthermore, figure 1 displays the annual temperature corrected specific energy consumption (kWh/m2) for different building categories in Norway. The category *food retail stores* has a consumption of 540 kWh/m2, and is by far the building category with the largest consumption of energy.

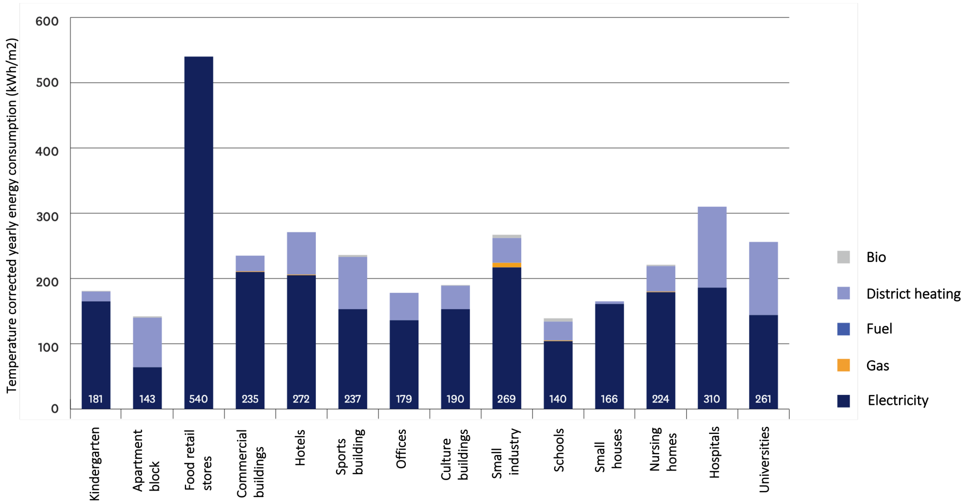


Figure 1: Temperature corrected spesific energy consumption (kWh/m2) in 2017. N=3425.

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 (Galvin 2014).

[4] https://www.ipcc.ch/report/ar6/wg2/

Nonetheless, one important barrier that hinders renovation projects is uncertainty regarding the expected savings (Kontokosta 2016). The main objective of this thesis is to contribute to lower that barrier and to deliver reliable methods to be used to document and monitor energy savings in retrofitting projects. To deliver on that objective we have written 5 different papers that demonstrates and presents different methods and frameworks. To set these 5 papers into context we use the schematic outline in figure 2 recently published by Grillone et al. (2020) where they review data-driven and deterministic methods to quantify energy savings and to predict retrofitting scenarios in buildings.

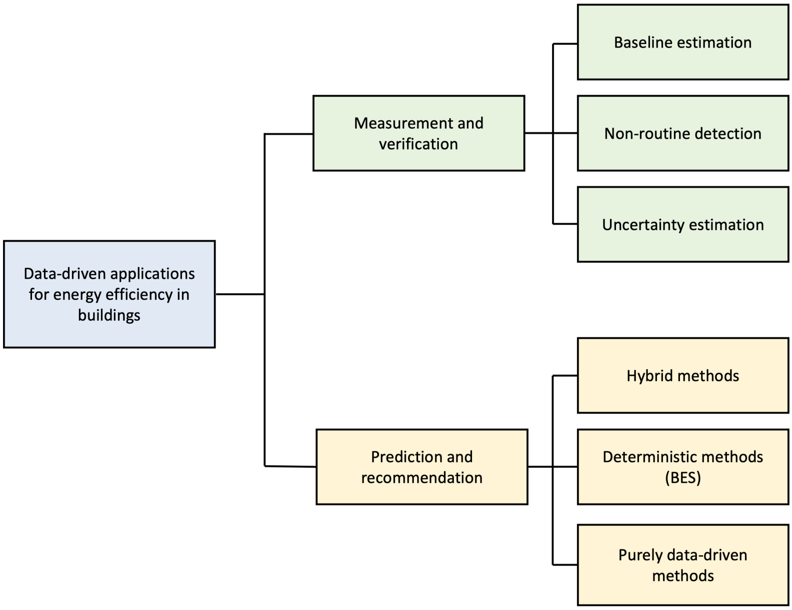


Figure 2: Overview of data-driven applications for energy efficiency in buildings

The two main paths are *Measurement and verification* (M&V) and *Prediction and recommendation* (P&R). The P&R is a process that identifies the most appropriate ECMs for the building under study (Grillone et al. 2020). Hence, the P&R phase is initiated before implementing any ECMs in a retrofitting project. The M&V process use measurements to give accurate estimates of energy savings in a building as a result of implementation of an energy management strategy (Committee 2016). Using data prior to the ECMs a model is trained and later used to predict the energy consumption after the ECMs *given that no ECMs where undertaken.* Comparing the actual energy use with the predicted we can estimate the energy savings (aka. *baseline estimation*), perform *non-routine detection*, and/or estimate uncertainty.

Table 2 positions the 5 papers within this context. Paper 1 and 2 is used to develop and apply baseline estimations. Moreover, while the application developed in paper 3 is primarily used for baseline estimation, it has been used to detect several non-routine events. The near real-time delivery of energy and temperature data was important to enable this, and as such we have chosen to indicate this by the name *M&V 2.0*. Paper 4 is a benchmarking study using hybrid methods, and the proposed 3-step framework is used for both P&R and M&V. Paper 5, the last paper, is within P&R and use purely data-driven methods to enable peak-shaving with optimal use of solar panels and batteries.

Table 2: The papers position in the literature

Papers	Application	Method
Paper 1	M&V	Baseline
Paper 2	M&V	Baseline
Paper 3	M&V 2.0	Baseline, non-routine detection
Paper 4	M&V/P&R	Baseline, hybrid
Paper 5	P&R	Purely data-driven

In the next section a closer look at the data used for all the papers is given. Furthermore, the methods are presented. Next, a summary presentation of the results from the papers are given, together with the discussion. At last we suggest some relevant future research opportunities and offer a conclusion.

5.2 Data

In Figure 3 an overview of the data pipeline that was used in this PhD project is presented. The energy data used throughout the 5 papers were collected from Elhub (https://elhub.no/en/). 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. Statnett, the system operator of the Norwegian power system, owns and runs, Elhub AS. Each day the grid owners has to provide updated energy data from the advanced metering infrastructure system (AMI) to Elhub. This system gives easy access to high-frequency data, daily updates, and allow for energy savings being estimated close to real-time (Grillone et al. 2020). The Elhub service was launched in February 2019. All the energy data from February 2019 and onwards for the 5 papers in this thesis stems from Elhub[5]. 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. Furthermore, the temperature data gathered is modeled weather

[5] https://cran.r-project.org/web/packages/segmented/NEWS

data that use several of the closest weather stations to set the temperature.

Data about the buildings, such as size and opening hours, was collected through an API with the building owners. All the buildings that are analyzed throughout the 5 papers are Norwegian food retail stores, and the building owners wanted the store names to be anonymized.

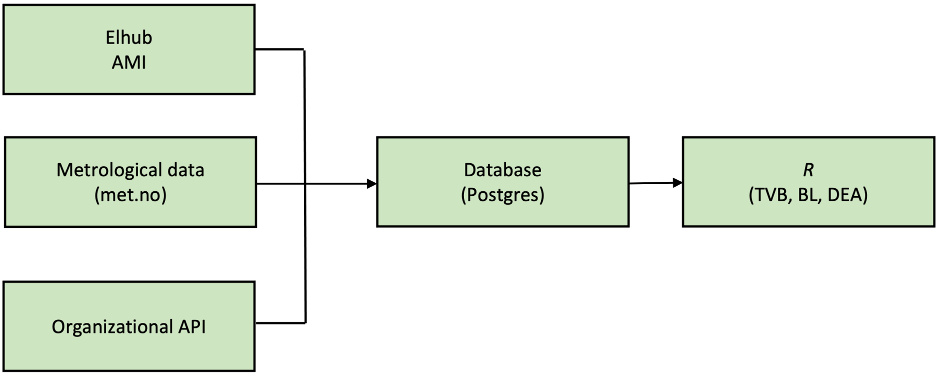


Figure 3: Data pipeline for the PhD project

Figure 4 illustrates two typical weekdays of hourly electricity consumption for one of the food retail stores from paper 5. The energy consumption follows the same pattern both days. During the night the consumption waver around 150 kW, and when the store opens the load shift to around 200 kW. Also, note the extra peak ("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 "stand-by" when the store is not open).

5.3 Methods

In the following section the methods applied in the 5 papers are presented. Since the research primarily is centered on measurement and verification (M&V), and in particular baseline estimation (paper 1, 2, 3, partly 4) the methods sections start with a description of the methods used to estimate energy savings, in particular the TVB model. Further, non-routine event (NRE) detection is presented. While energy saving baseline models was the main objective of paper 3 (the web application), we have throughout the project seen that the web application in many instances was used to detect errors in the technical infrastructure (e.g. a ventilation running in day-mode during the night). Hence, more details are given within this part of the thesis.

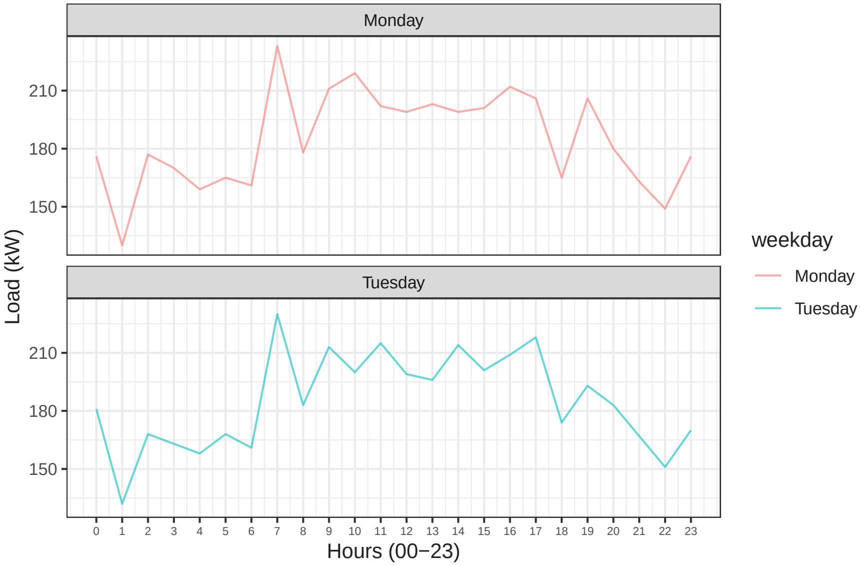


Figure 4: Hourly loads (kW) throughout a week

5.3.1 Measurement and Verification

The Efficiency Valuation Organization (EVO) is a non-profit organization whose services is designed to increase investments in energy efficiency projects worldwide. The EVO is politically, geographically, commercially and technologically neutral. The work they do is primarily directed to practitioners. In 1997, the EVO published the International Performance and Verification Protocol (IPMVP) (Committee 2016). The protocol was originally developed to promote investments in renewable energy project around the world, particularly within energy and water efficiency.

Energy savings cannot be measured directly because savings represent the absence of energy consumption. Instead, savings are established by comparing energy consumption before and after implementation of energy conservation measures (eg. change to LED lighting and/or other measures to reduce energy consumption). To be able to relate the reduction in energy consumption to the ECM it is important to adjust for changes in conditions between the periods that are compared. This is illustrated in figure 5.

Furthermore, the IPMVP defines best practices to quantify energy savings. Four different options are available to establish energy efficiency savings:

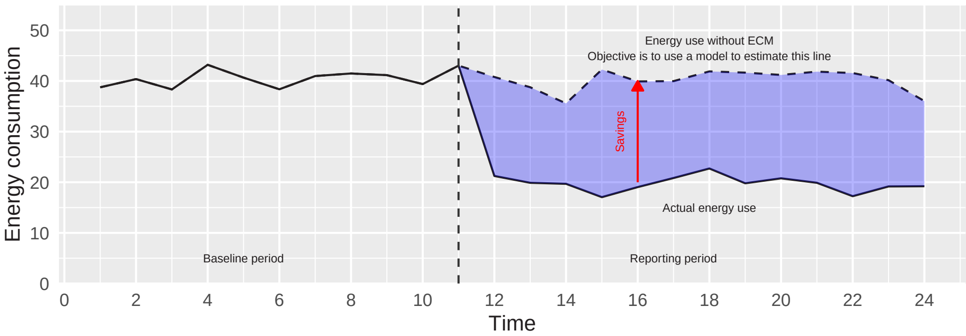


Figure 5: Illustration of measurement and verification. Baseline = before the retrofitting, Reporting period= after the ECM

- Option A: Partially Measured Retrofit Isolation. This option involves use sensors to monitor the consumption of the equipment affected by the installed ECMs. The consumption is isolated from the energy use of the rest of the building. The option use partial measurement, which means that that some parameter(s) are estimated rather than measured.
- Option B: Retrofit isolation. This option is partly equivalent to option A, however no estimations are allowed ad full measurement of all the relevant parameters is required.
- Option C: Whole building. In this options data from utility meters are used to evaluate the energy performance of the whole building. Note that this option establish the total savings of all implemented ECMs and is only applicable in retrofitting projects where savings are expected to have a large impact, making them distinguishable from energy variations unrelated to the applied measures
- Option D: Calibrated simulation. This option involves using energy modeling software that allows prediction of energy consumption. Typically, the models used for this option are calibrated to match the buildings real (metered) data.

In the present thesis *option C* is used for paper 1-4 (the scope of paper 5 was not to estimate energy savings). Note that *option C* is recommended when the expected savings are large, to be able to distinguish the savings from unrelated energy variation. Notwithstanding, in paper 1 we argue that, at least for retail food stores, this option may also be used to document energy savings for smaller installed ECMs. This will be discussed in more detail when discussing the results. Also, the IPMVP *option C* is the same as option C from the ASHRAE guideline 14 for measurement of Energy, Demand and water savings (Guideline and Others 2014), published by the American Society of Heating, Refrigeration, and Air-Conditioning Engineers (ASHRAE). The ASHRAE include metrics that can be used to evaluate the reliability of baseline models. Accordingly, to measure the accuracy of the models from paper 1-5 the *coefficient of variation root mean square error* (CV-RMSE) is calculated. The CV-RMSE is computed in the following way,

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

where $\bar{Y}$ is the mean of the energy consumption in the training data (the reference 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. The ASHRAE guideline requires the CV-RMSE 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.

5.3.1.1 Baseline estimation Throughout this thesis 3 different methods are used for baseline estimation. First, the machine learning approach, component-wise gradient boosting with penalized splines (CW-GB), and the details about the estimation is outlined in paper 2. The rationale behind the use of CW-GB was the model performance in the Kaggle global energy forecasting competition 2012, where the CW-GB ranked fourth out of 105 participating teams (Taieb and Hyndman 2014). The CW-GB was implemented using the `mboost` `R` package with 5-fold cross-validation (T. Hothorn and Hofner 2018). Furthermore, CW-GB was tested in the baseline module in the web application (as developed in paper 3). Nonetheless, the users of the application expected fast estimation of the models, and the CW-GB often took more than 60 second to compute. At the same time the TVB model performed excellent in terms of the CV-RMSE values, and only took a few second to compute. Hence, only the TVB model is currently possible to use in the web application. Anyhow, the `mboost` package is still actively maintained. For instance, in the latest `mboost` version convenience functions for hyper-parameter selection, faster computation of predictions and improved visual model diagnostics are available. Code base and more details can be found at https://github.com/boost-R/mboost. Future work testing out new features will be initiated and given that some initial (unpublished) analysis show that the TVB models is somewhat more challenging to apply on building categories outside of food retail stores, the `mboost` may be a promising challenger. These issues are given more details in the thesis section 'Identified gaps for future study'.

Second, the broken line model and its details are outlined in paper 1. The main advantages of choosing this modeling approach were its resemblance with the approach used in the ESCO industry in Norway, combined with the ease of computation thorough the `R` package `segmented` (Muggeo 2022). While not discussed in paper 3, the `segmented` package is operationalized into the baseline web application, and is actively used by the ESCO, often to check if the results are the same as in the web application TVB model (for a screenshot from the web application, see figure 12). More details about the `segmented` package can be found at https://cran.r-project.org/web/packages/segmented/segmented.pdf. As with the `mboost` package used in paper 1, the `segmented` package is also continuously updated with new features. For example, just recently (2022-05-30) the package maintainer introduced a function to fit segmented mixed models, i.e. segmented models with random effects changepoints[6].

Third, the Tao Vanilla benchmarking model was used in all 5 papers. The TVB model has previously proven easy to implement and produce accurate results (Hong 2010). The model has been introduced several times in the papers. However, due to space limitation in the published papers the exposition has been brief. In the following section we provide a more detailed introduction.

5.3.1.1.1 The Tao Vanilla benchmarking model The Tao Vanilla benchmarking (TVB) model originates from Tao Hong's PhD thesis "Short term electric load forecasting" (STLF) (Hong 2010). In his dissertation Hong disassembles the major methods that have been used for STLF in the literature and reassemble the key elements into a multiple linear regression framework that can be applied to STLF. In particular the TVB is formulated with the use of qualitative variables, polynomial regression, and interaction regression. The model is a linear regression model with some well specified features,

$$\begin{aligned} Y_t = \beta_0 + \beta_1 M_t + \beta_2 W_t + \beta_3 H_t + \beta_4 W_t H_t + \beta_5 T_t + \beta_6 T_t^2 + \\ \beta_7 T_t^3 + \beta_8 T_t M_t + \beta_9 T_t^2 M_t + \beta_{10} T_t^3 M_t + \beta_{11} T_t H_t + \beta_{12} T_t^2 H_t + \beta_{13} T_t^3 H_t + \beta_{13} Trend + \beta_{14} Load_{t-1} \end{aligned} \tag{1}$$

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, $Load_{t-1}$ is the load the previous hour, and $Trend$ is a trend variable. Further, T_t is the outside temperature for time t.

To come up with this particular specification of the model Hong tested seven different linear regression models. The test was demonstrated using the case study of one week ahead hourly forecast for a medium US utility, and the mean absolute percentage error (MAPE) was used as a performance measure. The model from (1) was the best performer.

Energy consumption in a retail food store varies based on month, weekday and hour. The energy consumption is larger during the winter, stores are closed in the weekend, and the consumption is likely to vary between

[6] https://cran.r-project.org/web/packages/segmented/NEWS

hours, hence, month, weekday and hour are included in the model. Furthermore, another important variable is the interaction between *Hour* and *Day*. A food retail stores typically opens at 07:00 and closes at 22:00. The *Hour x Day* interaction is included to incorporate this feature into the model. This can also be seen in figure 4 where we see the increase and the decrease in energy consumption during opening and closing hours.

Because the load often increases both when the temperature drops and increases, it is necessary to take this into account. This could be incorporated with linear piecewise functions. However, that would require cut-off temperatures which may be different across different buildings. Thus, this is included in the model using 3rd ordered polynomials of the temperature. Also, the model include interaction effects between the polynomials of the temperature and the calendar variables *Hour* and *Month*, respectively. The rationale behind this is observed in figure 6 and 7 with scatter plot of load temperatures across month and across hours. There are observed differences for load versus temperature for both month and hour (while only sightly) which justifies the inclusion of the interaction.

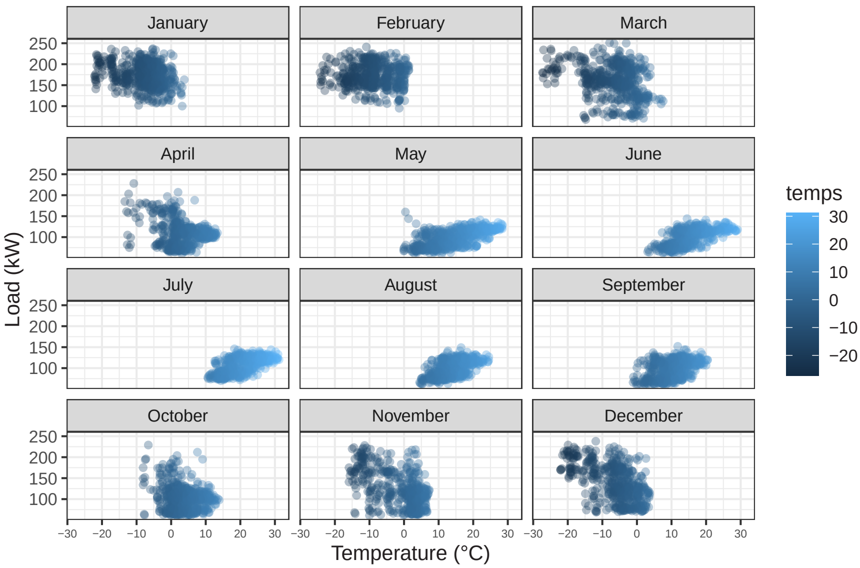


Figure 6: Hourly loads (kW) and temperatures during a year for a food retail store

Note that in (1) both *trend* and the lagged load ($Load_{t-1}$) of the dependent variable is part of the explanatory variables. Since baseline models for energy savings should represent only features from the training data the previous hour load and trend are excluded from any of the TVB models in this thesis.

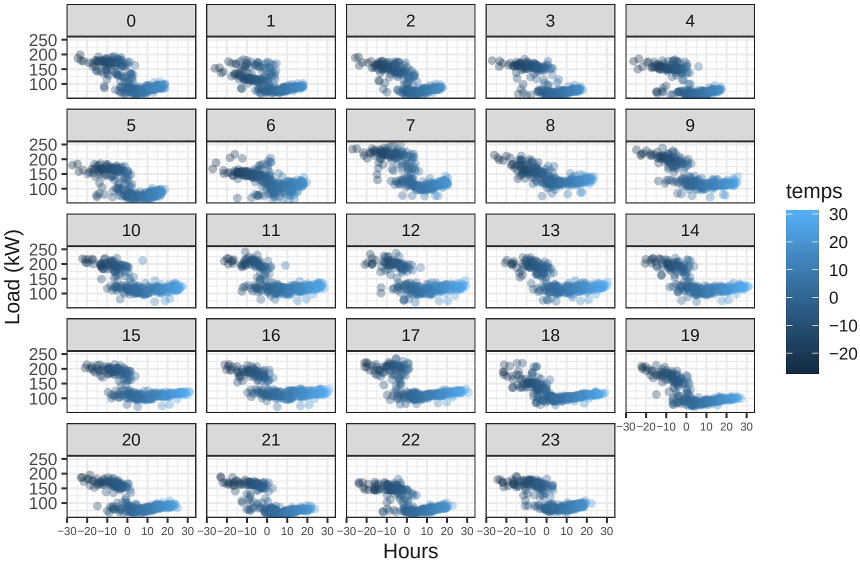


Figure 7: Hourly loads (kW) and temperatures between 00:00 and 23:00

In Hong's thesis the TVB load forecasting performance was tested against possibilistic linear models (PLM) (Hong 2010, 117) and artificial neural networks (ANN) (Hong 2010, 136). The TVB outperformed both of these modeling alternatives. However, regarding ANN there has been substantial improvements over the last decade, additionally, the structural setup of the network has impact on the performance. In the section "Identified gaps for future study" some further ideas will be promoted.

The TVB was used to produce benchmark scores for GEFCom2012 (Hong, Pinson, and Fan 2014) and was eventually ranked in the top 25% among over 100 teams. It has also been implemented as a base model in the commercial software package SAS Energy Forecasting. The same model was used in the weather station selection framework proposed by Hong, Wang, and White (2015), and in Wang, Liu, and Hong (2016) the TVB is extended with recency effects.

5.3.1.2 Non-routine event detection Non-routine event detection is often a challenge in the M&V industry and is prevalent to all the previously introduced baseline estimation methods. Non-routine events (NREs) are defined as change in the energy use of a building that are not caused by any variation of the features in the baseline model (Grillone et al. 2020). In a food retail store this may be faults in the meter readings, unplanned downtime in the technical system or promotional events in the store. These non-routine issues are important to detect and adjust as the estimated savings will not be correct without necessary adjustments. Granderson et al. (2017) point out that these adjustments are often done manually and may require engineering expertise. Within food retail these events are quite frequent, and the adjustment are often done in collaboration with the building owners. Nonetheless, while it is not very difficult to do the actual adjustment, it can be challenging to detect that a NRE has occurred. Often an ESCO has ongoing projects on several different buildings, and it is not unlikely that NREs bypass unnoticed. Also, as the time span between the NRE and the energy saving analysis increases this adds to the complexity. It is often challenging to re-construct events inside a store months later. Services such as Elhub that provides updated energy data on a daily basis enables current updates of baseline models and simplifies detection of NREs.

In Touzani et al. (2019) they suggest a NRE detection algorithm based on a statistical change point detection method and a dissimilarity metric. The dissimilarity metric measures the proximity between the actual time series of the post-retrofit energy consumption and the estimated baseline model. This is novel way of detecting errors that separates itself from current methods that use individual time series. The authors also propose that the methodology may be used to detect errors in the technical system. A dissimilarity metric as proposed by Touzani et al. (2019) would improve the above-mentioned EPC monitor, and will be tested in future versions of the application.

However, in Touzani et al. (2019), historic energy consumption (before the ECM) is used to predict what the energy consumption would have been without the installed ECMs. It is the difference between the predictions and the actual data that is used to detect potential NREs. Yet, an error in the technical infrastructure could be present in the training data. For instance, suction pressure in a component in the refrigeration system may have been non-optimal in certain time intervals, and as a consequence the energy consumption larger than necessary. If such an error (or rather, non-optimal setting) was present, and not detected, during the training of the model, the above methods will not work. To detect such instances a more general approach to time-series outlier detection may be applied.

In the R forecast package there is a function that can detect unusual values in a time series[7]. The function decomposes the time series into three components: trend, seasonal and remainder:

$$Y_t = T_t + S_t + R_t$$

The idea is to remove seasonality and trend to ease detection of outliers in the remainder. Since we are working with hourly data that have multiple seasonal patterns we first apply Multiple Seasonal-Trend de-

[7] https://robjhyndman.com/hyndsight/tsoutliers/

composition using Loess (MSTL) (Bandara, Hyndman, and Bergmeir 2021). Compared to other methods, MSTL gives reliable results with low computational cost. The strength of the seasonality is measured through

$$F_s = 1 - \frac{\mathrm{Var}(y_t - \hat{T}_t - \hat{S}_t)}{\mathrm{Var}(y_t - \hat{T}_t)}.$$

Given $F_s > 0.6$, a seasonally adjusted series is calculated:

$$y_t^* = y_t - \hat{S}_t.$$

Further, Hyndman points out that the seasonal strength threshold is necessary because $\hat{S}_t$ may be noisy and overfitted if there is no underlying seasonality (or if it is weak). Hence, outliers may "disappear" in the seasonal component. If $F_s \leq 0.6$, then $y_t^* = y_t$. In the next step the trend component is re-estimated with Friedman's super smoother to the y_t^* data. The outlier is then potentially detected in the estimated remainder series

$$\hat{R}_t = y_t^* - \hat{T}_t.$$

Interquartile ranges (IQR) are used to single out the outliers, e.g., if less $Q1 - 3 \times IQR$. This perspective is not present if we use the baseline models to detect outlier as in Touzani et al. (2019). Presently, this approach is not applied within this thesis, but will appear in future versions of the web application that developed as part of paper 3. A reliable schema for detection NRE is vital to deliver correct estimates of energy savings.

5.3.1.3 Data envelopment analysis - benchmarking energy efficiency The M&V and the accompanying methods have been introduced in the last few pages. Often, in our experience, the M&V phase is initiated at a stage when the ESCO has already installed the ECMs, and the deadline for delivering the energy savings analysis is approaching (often in near future…). Hence, the energy analyst is contacted! Nevertheless, there are several good reasons that it may be worthwhile to train the baseline models during the audit phase when the ESCO is evaluating what ECMs are likely to provide the best energy savings results. First, training the models "up-front" gives the stakeholders time and opportunity to evaluate the quality of the energy data. Not rarely, there will be several issues that need to be handled. Missing data, errors in AMI equipment, information about the stores, and all sort of things that may happen, but should not happen, inside the stores during the reference period. Thus, data cleaning is vital, and much more well served if done during the audit when "everyone" is working on collecting information. Second, the results from training the models may be useful information during the audit phase.

For example, in the benchmarking paper (paper) 4 the BL model provided the demand for energy consumption for cooling and heating (and the CPT value) through all the buildings in the retrofitting portfolio. This information could be useful when evaluating the technical infrastructure. This is what motivated paper 4 which is an attempt at using the output from the baseline models throughout a retrofitting project. To tie things together for benchmarking the energy efficiency data envelopment analysis (DEA) is chosen. The methods is outlined in detail in paper 4. The `R` package `benchmarking` (Bogetoft and Otto 2020) was used to conduct the DEA.

5.4 Results and discussion

In paper 1 the main objective was to compare and demonstrate the TVB and the BL model, two different baseline models to estimate energy savings from ECMs. The TVB model predict energy consumption in buildings on an hourly level, while the BL model use weekly data. The two aggregate levels complement each other as the results give insights into different aspects of how the ECMs work. This is illustrated in figure 8 for store-id 4391. On the left-hand side of the plot the energy temperature curve estimated with the BL model is presented. The red arrow illustrates the energy savings in the winter, and the blue arrow the savings in summer. We see that the savings are larger in the winter. Also, the changing point temperature (CPT) is 6.7 °C. On the other hand, the right-hand side of the figure shows the estimated savings on an hourly level from the TVB model. The dotted line is the predicted load given that the building performs as before the installed ECMs. While the BL model gave an overview of the energy savings on a level that can be visualized for a whole year (using weekly data), the TVB model give more specific details. For instance, store-id 4391 used to have a morning peak around 08:00 (> 200 kW), and as a consequence of the implemented ECMs this peak has been reduced. Thus, the TVB model gives the opportunity to study the savings from a different perspective. Which hour gave the largest saving? Are there any hours where the ECMs does not work as expected? In some cases, information on an hourly level may also be used to optimize the ECMs. For instance, during our collaboration with the ESCO several stores that had ventilation running on day-mode during the night was detected using the TVB model.

To conclude, since both modeling approaches and the aggregate levels give useful and complementary insight, the practical solution is for the ESCOs to use both the BL and the TVB model in ongoing retrofitting projects. Additionally, if the results from two the models support each other that gives more reliability to the results.

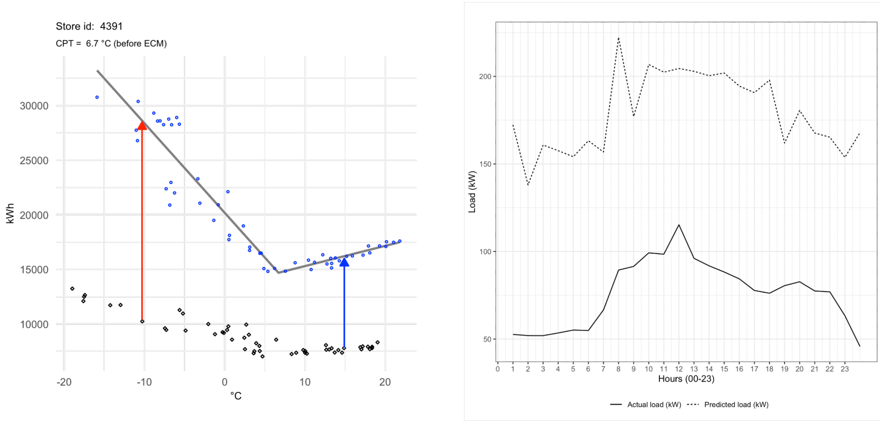


Figure 8: The BL model versus the TVB model for store-id 4391

On contrary, if the approaches do not support each other, then that again is useful information for further investigation.

While paper 1 takes a relatively simple approach to estimate the energy savings, the objective of paper 2 is to investigate how a more modern machine learning approach will perform, and if the additional cost in terms of complexity and computation can be a worthwhile investment. The paper use component-wise gradient boosting (CW-GB) to design a baseline model, and the results are compared with the same model that was applied in paper 1, the TVB model. The demonstration is done on ECMs that was implemented in nine different retail stores during spring 2017, and the effect of the ECMs were followed through 2018.

In figure 9 the relative variable importance for the 9 stores are presented (as in the paper). For all the stores the opening hours (hour, temps) are among the most important features to explain the stores energy consumption. Also, different variants of temperature are important. Naturally, this finding did not come as a big surprise for any of the stakeholders in the project. However, with a richer feature set available for the CW-GB model this result could potentially be relevant in future projects. For example, there is ongoing work to get information about how many customers are in the stores at any given time, revenue (as a proxy of visitors), technical infrastructure, building envelop, and more detailed climatic data (wind, sun irradiance). Hence, having these variables as part of the modeling process it might be more useful to fully understand how the individual variables impact each buildings energy consumption.

The main objective of paper 2 was to compare the TVB model with the CW-GB model, a “simple” model versus a more modern machine learning approach. The paper demonstrated that both models could reliably be used to estimate savings from ECMs with expected savings less than 10 %. One advantage of the CW-GB

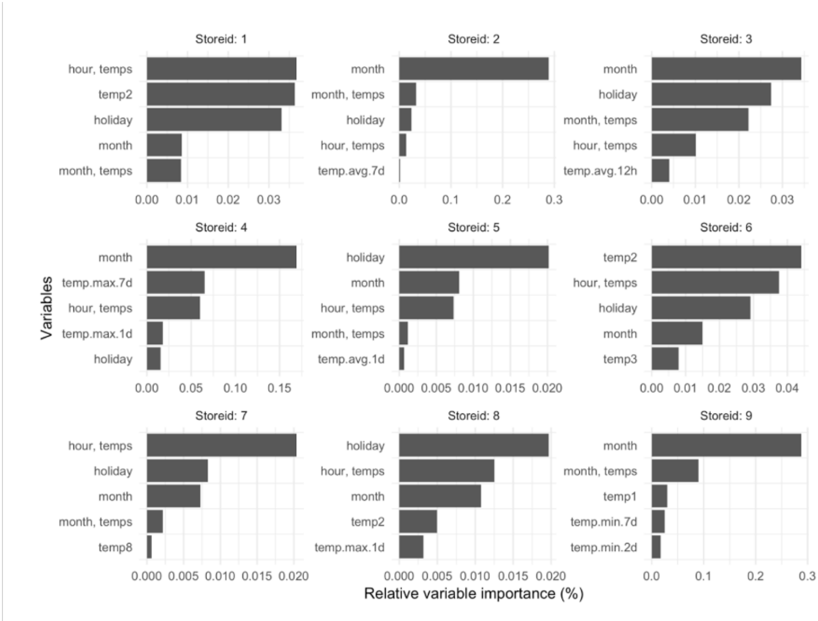


Figure 9: Relative variable importance (%) as estimated from the CW-GB model

model was that it can take into account each building's unique set of energy consumption predictors. Also, the precision as measured through the CV-RMSE values was slightly better in the CW-GB model for all the stores. We found that the TVB model was less computationally expensive, while the CW-GB model, given its iterative nature (finding the set of variables that best explain energy use), took longer to compute.

At last, paper 2 had strict space limitation in terms of the intended audience, hence, the profitability of the ECMs was not discussed. Nonetheless, table 3 in the paper "Estimated savings - the difference between actual energy consumption and predicted" presented the kWh in 2018 and the predicted kWh (given no ECM). For example, Store 3, used kWh 534 160, while the predicted kWh usage was kWh 669 130, thus the savings was kWh 134 970 for 2018. Further, the monthly spot prices (in NOK) from 2018 up until May 2022 for the price area NO1 (where the stores are located) is displayed in figure 10. As can be seen from the time series the spot prices have gone through a substantial increase since the start of 2021. The average price was 0.523 NOK in 2018, and in the last twelve months the average has increased to 1.515 NOK. Thus, the savings in 2018 is kWh 134 970 * NOK 0.523 = NOK 70 598. Using the spot price average for the last 12 months this saving would have been NOK 204 479. Also, the average installation cost of the ECMs was NOK 124 894, with NOK 14 410 yearly operating expense. Note that the savings are only estimated for

2018, but the effect are expected to last for several years. These issues are also raised in the next section of the thesis were identified gaps for future study are presented as it would be useful to follow the savings from the ECMs for a longer time period than a year to investigate if the savings maintain or if any diminishing effects occur.

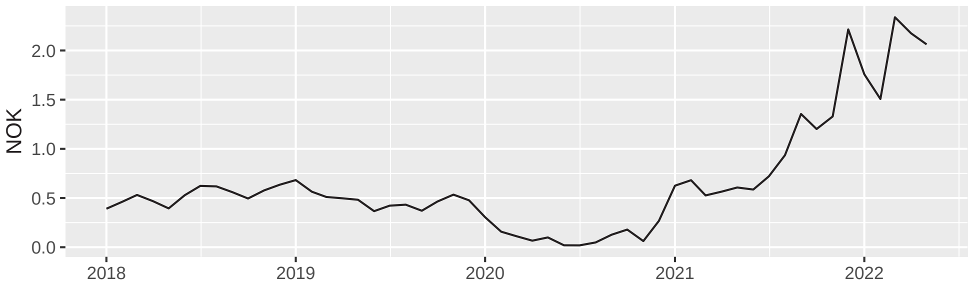


Figure 10: Spot prices (Nordpool) 2018-2022 for area NO1

In paper 3 “ShinyRBase: Near real-time energy saving models using reactive programming” the TVB model is operationalized into a web application using open-source tools and a reactive programming framework. The web framework allows for a fast development cycle without any need-to-know web programming languages like HTML, CSS or JavaScript. While the most important objective with the web application was to make the ESCO and other stakeholders self-sufficient in terms of setting up energy saving baseline models, another important point is the flexibility the framework offers. For instance, the `R` library `tidypredict` (Kuhn 2020) was used to run predictions from the TVB model inside the database `PostgreSQL`. Accordingly, it is straightforward to test other modeling approaches such as random forest, XGBoost or Tree models. [8]

Paper 3 focused the attention on how the TVB model was used to automate baseline models in the web application. However, the BL model that was presented in paper 1 is today operationalized into the web application that the ESCO use daily. Figure 12 displays a screenshot from the web applications user interface for baseline models with the BL model. The text in the screenshot is in Norwegian, but in this example the energy-temperature curve (ET-curve) as estimated from the BL model is for the year 2018 (the user can choose this in the menu under ‘Referanseperiode’, and the ECM period is configured in ‘Tiltaksperiode’). The y-axis is the weekly energy consumption, and the x-axis is the average weekly temperature. The red dots show the energy consumption and the corresponding temperature in the ECM period, while the yellow dots show the same, but for the last 5 weeks. The distance between the dots and the estimated ET-curve is

[8] a full list of supported models can be found at: https://github.com/tidymodels/tidypredict

the energy savings.

The ESCO often use the BL functionality to check if the results are the same as in the baseline models estimated using the TVB model. Indeed, as found in paper 3, the results from the BL and the TVB model should coincide.

Also, after setting the baseline module into production in a web application several NREs has occurred. These were often detected by one of the project owners working on the project, without any structured approach to investigating the events. To alleviate this process and enable a more automated method to detect NRE we developed an EPC monitor web dashboard using the `R Shiny` library described in paper 3 (Chang et al. 2021).

In figure 11 we display a snapshot from the EPC monitor for three of the food retail stores that the ESCO monitors after implementation of ECMs. The column “% savings” is the aggregated savings for 2022, kWh baseline is the actual energy consumption, and kWh baseline is the predicted consumption (models trained using data from 2019). The ESCO had some dismal connotations when we at first used “Beta-coefficient” and “P-values” as column names, hence, they were reborn as “Trend score” and “Security”, respectively, and “Security” even as the inverse of the P-value (1-p). The sparklines gives a visual representation of the weekly savings. Figure 11 clearly shows that two of the stores (Store 10022 and 10023) had a weekly negative savings trend, with p-values of 0.02 and 0.05, and coefficients of 0.99 and 1.42. On the other hand, store 10020 is doing fine and the savings has increased each week (p-value = 0.00). Although do note that this was a store where only an optimization of the technical infrastructure was implemented, hence the low average energy savings of -7,4%.

Store-Id	↑ Trend Score	Security	kWh ECM	kWh baseline	% savings
Store 10020	-0.86	1,00	128223	138464	−7,4
Store 10021	0.9	0,57	128642	257135	−50,0
Store 10022	0.99	0,98	180574	181105	−0,3
Store 10023	1.42	0,95	115228	322316	−64,3

Figure 11: NRE detection (from 2022-021-021) within the Shiny web application

In our experience this approach has been able to detect several NREs during the last year. For instance, a ventilation system that was running in day-mode during the night, and AMI meters that stopped working. As Grillone et al. (2020) points out, if we fail to adjust for NREs the estimated savings may be too high or too low. On the other hand, the potential adjustment depend on the type of NRE and the retrofitting contract that the ESCO has with the building owners. Off course, if the AMI meters stops working, it will obviously look like the savings are larger than actual. However, a non-routine event may also be an error in the technical infrastructure, and the ESCO may be responsible to repair.

In summary, the reactive framework delivers three distinct advantages.

1) The stakeholders will always have a current and real-time report on the savings.

2) Complex methodologies are dynamically used by the end-user.

3) Increased involvement by stakeholders and interaction with the analyst related to the methods used in the energy savings analysis leads to collaborative benefits such as faster disseminating of knowledge

At last, ECMs control-systems, like presented in paper 3 does not operate themselves. It is crucial to educate persons internally to use and understand the baseline models, and to be able to act on the information. Often, as Johansson and Thollander (2018) points out, the focus is solely on energy-efficient implementation of technology, while human behavior is overlooked. This finding is important to take into consideration and to be able to fully use the potential in a baseline web application.

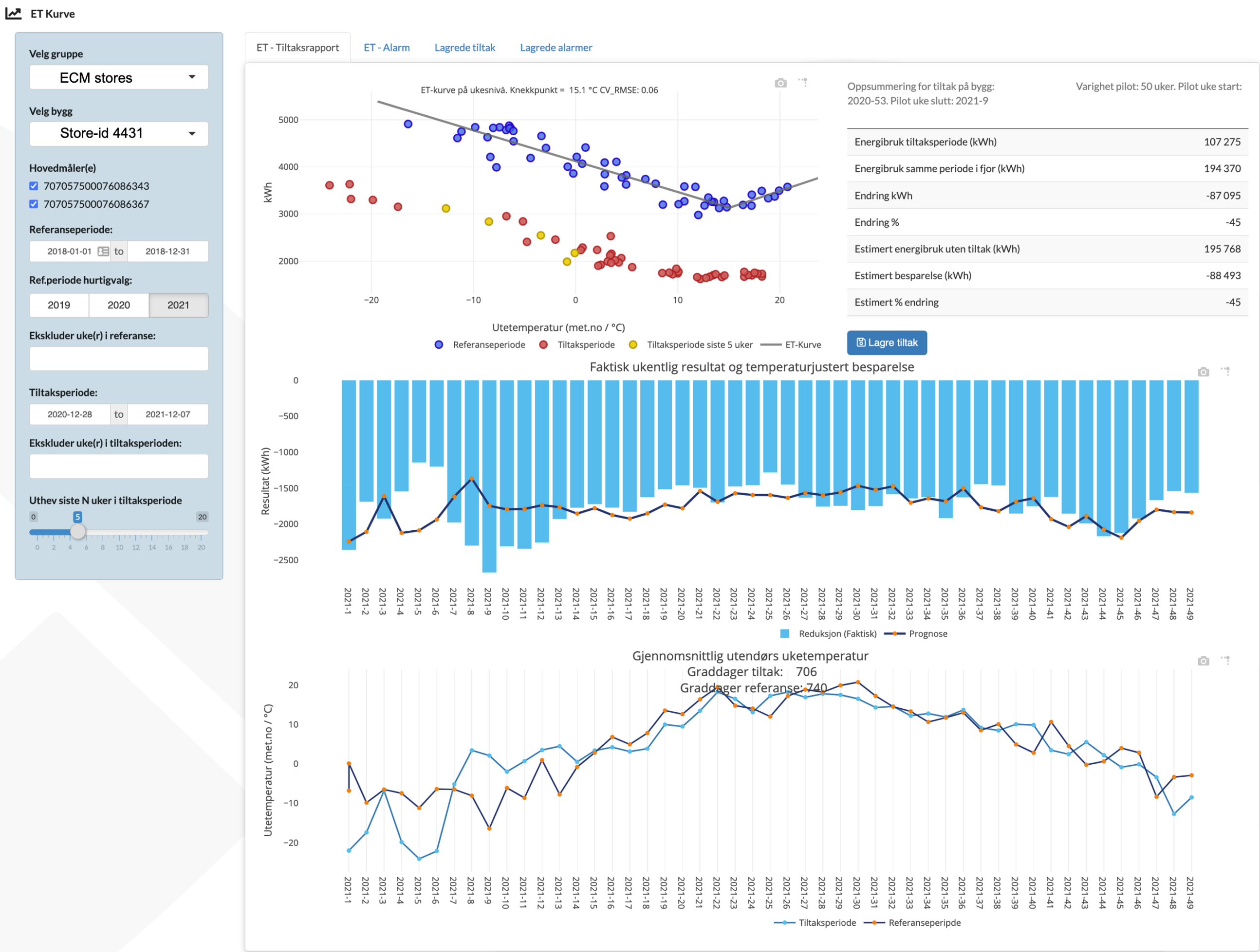


Figure 12: Screenshot from the web applications userinterface for the BL model

The fourth paper *"A 3-step framework to benchmark potential and actual energy savings in retrofitting projects"* is an extension of paper 1 and 2, where we analyze not only the energy savings, but also energy efficiency with the use of data envelopment analysis (DEA). The paper propose a framework for benchmarking energy efficiency and savings.

During the energy audit the ESCO often visit and collects information about the building(s) that are part of the retrofitting project. A detailed examination of the technical infrastructure is an important part of this process. Nevertheless, in our experience, not much analysis is conducted on the energy consumption of the building during the audit. Often, the only number that is part of the audit phase is the energy intensity (Yearly kWh / m2) which is used to rank the different buildings. However, it may be of interest to analyze both current load profiles and efficiency in more details. In paper 4 we argue that it is worthwhile to train the baseline models during the audit phase, and not postpone this step until the M&V phase. There are two prevalent advantages to this. First, during the audit phase the ESCO has full attention to the buildings under contract, and at this stage it is likely to be easier to understand any issues with the data that will be used to train the baseline models. Often the energy data has issues that needs to be considered (technical issues with the meters, building/technical infrastructure non-operative for periods, promotional events). Getting attention to these issues is easier at a time when the ESCO is already working on related issues. Second, the information that training the baseline models gives may be useful input into the audit phase. For instance, the ESCO use, among other things, this initial stage to understand the heating and cooling demand throughout the building portfolio. Accordingly, the BL model and the TVB can give detailed information about the cooling and heating demand in the buildings.

The above describes the initial stage of the framework proposed in paper 4. Next, a data envelopment analysis (DEA) is conducted to benchmark the energy efficiency. The efficiency scores from the DEA reflect not only the size of the buildings, but also opening hours and heating and cooling demand (from the BL model). The benchmark scores leads to a more correct ranking of the buildings as compared to the standard way of using energy intensity. Furthermore, the DEA is re-done after the ECMs are installed and have been in operation for a year. This enables an investigation of the potential change in efficiency scores, and how these may relate to the actual savings. This may give valuable insights that can be used to improve the effect of current and future retrofitting projects. For instance, there might be an inefficient heating or cooling system in some of the stores. Additionally, the accomplished energy savings might not have been realized in terms of the potential as identified by the efficiency scores. Such issues may require further investigation.

To sum up. The objective of this framework is to have a more holistic approach to the retrofitting project. Training the baseline models during the audit may give more reliable (and faster reporting) during M&V.

Further, take advantage of the output from the baseline models in the audit and implementation phase, and at last, extend the typical energy intensity ranking with more information about the buildings.

In paper 5 *"Forecasting and technoeconomic optimization of PV-battery systems for commercial buildings"* the objective was to investigate the profitability with peak shaving (reducing the maximum loads) in Norway for a commercial building. To optimize profitability it is crucial to have information about when the peak will occur. Hence, a forecasting algorithm for load prediction was developed, and the economic value of forecasting was determined for a PV-battery system. The previous TVB and the CW-GB model was applied. The economic value of forecasting was determined through simulations with *Homer Energy Software* that optimizes the net present cost of the systems. The results showed that battery storage was only economically beneficial when forecasting was deployed.

Finally, the research in this thesis has been conducted in close cooperation with the ESCO, and today the web application that was developed as part of this work is in daily use to follow the energy savings across several hundred food retail stores. Additionally, over the last years energy performance contracts (EPC) has a larger share of the work conducted by the ESCO. Because EPC contracting deliver promised energy savings to building owners it is critical to have reliable tools to document the savings. For example, Lee, Lam, and Lee (2015) conducts a survey of 34 ESCOs and 168 retrofitting customers regarding the inherit risks in EPC contracts. The study finds that ESCOs worry about possible payment default after installation, uncertainty of baseline measurement, and increase in installation costs in EPC projects. On the other hand, the customers primary concerns are long payback period, repayment ability and project complexities. By the same token, all the stakeholders agrees that promotion of successful retrofitting projects is important to enhance adoption of future EPC contracts.

Documenting energy savings is not the only factor that need to be established to advocate new retrofitting projects. For instance, Minetto et al. (2018) studied factors within the HVAC market in the food retail market that could contribute to a faster implementation of energy efficient technology. They find that increased knowledge about the opportunity to receive public funding and increased knowledge about new technology are factors that may contribute to this respect. Furthermore, recent research shows that cost savings is the most important motivation to invest in energy efficient technology in supermarkets in the UK. Low operating cost is a requirement to keep the food retail prices low. Yet, sustainability and corporate branding is also an important aspect that drives willingness to invest in new and energy efficient technical infrastructure (Ochieng et al. 2014). As such, it will be important to bundle these factors into powerful advertisements to stimulate new projects.

5.5 Identified gaps for future study

While this thesis is currently at full length, that does not imply that the project is complete. For the duration of the project several research opportunities have been identified. For example, the energy saving analysis conducted in the different papers accounted for the first year of the installed ECMs. Notwithstanding, the lifespan of the ECMs are expected to last for at least 10 years, which is the length of many EPC contracts. Hence, investigating possible diminishing return of the ECMs is an important topic for future research. Additionally, over the years maintenance of the technical infrastructure will be more involved. In that respect statistical models to deliver predictive maintenance is a promising research area. A recent review can be found in Burak Gunay, Shen, and Newsham (2019). For instance, the reactive framework presented in paper 3 may be further developed to connect with data from sub-meters (e.g., ventilation, refrigeration) and to deliver models that can predict possible breakdowns. Such a tool will not only have value from a research perspective but can potentially be very important to reduce service costs in the food retail sector.

Furthermore, we have seen that the ESCOs energy performance contracts, through several different ECMs, have delivered considerable energy savings. However, it may be useful to investigate the building owners' drivers and barriers to further ease adoption of efficient energy conservation measures in new projects. For instance, Brunke, Johansson, and Thollander (2014) investigates conditions that are associated with the adoption of energy cost-effective ECMs in the Swedish iron and steel industry. They find that only four of 23 companies invest in ECMs with a payback time of more than three years. While this finding might not be directly transferable to other industries or countries, it may still be important to consider that the technical infrastructure in the food retail sector should last for more than 10 years, and that it is important to develop a strategy to convince the building owners to adapt a longer perspective in their investments. Also, half the firms that participated in the Brunke, Johansson, and Thollander (2014) study reported insufficient top management support as important barriers for new ECM projects. Hence, it is important to raise awareness of energy efficiency within the companies. As such, it would be interesting to study how and if tools like the *ShinyRBase* web application (paper 3) can raise awareness for energy efficiency within the company.

Also, as sensors get more prevalent it will likely be less expensive and less complicated to analyze the actual energy savings using the IPMVP Option B, which is to isolate the installed ECMs and calculate the actual savings without any estimation. To be sure, technological advances will require new and innovative research to take advantage of these new and promising data sources.

It is also important to highlight that all the energy savings results throughout the paper was presented as point values, despite the fact that the results were estimated, and that uncertainty exists. Yet, by providing a range of uncertainty it is easier to evaluate the risk associated with investments in the ECMs (Tian et al. 2018). Given that the savings exists as an estimate with a lower and a upper bound a retrofitting business case may fell short in the lower bound, and be profitable in the upper bound. Subsequent research should further investigate issues surrounding this estimation uncertainty.

Note also that the TVB, the CW-GB and the BL models applicability have been demonstrated. Nevertheless, the scope of the demonstration was narrow; only buildings from the food retail sector were involved. It seems very unlikely that the presented models will work for every building category. Still, the reactive framework we developed in paper 3 is possible to adapt for different building categories. For instance, several of the machine learning models that were applied in the ASHRAE Kaggle competition can be implemented and tested within the same framework as described. Progress within this area may be useful research to develop baseline models across different building categories.

At last, Franconi et al. (2017) points to three considerations for how to choose the methods for calculating energy savings. First, regulatory requirements that may well differ between countries and states. Second, the methods validity in terms of over- and underestimating energy savings. Third, how to strike a balance towards the methods rigor and reliability versus the costs and the value of the energy savings. There is little information in the literature about optimal approaches, and contributions would be useful and appreciated.

5.6 Conclusion

The IPCC Sixth Assessment Report urgently stress the importance of major cuts in greenhouse gas emissions. There is no doubt that human activities cause climate change, and that these changes have led to widespread disruption in nature and affect billions of people. Globally, buildings consume 40% of all produced energy and are major contributors to GHG emissions. To that end, energy efficiency retrofitting is an important step in reducing energy consumption. Still, one important barrier that hinders renovation projects is uncertainty regarding the expected savings. The main objective of this thesis was to deliver reliable methods to be used to document and monitor energy savings in retrofitting projects. Through 5 different papers this thesis has demonstrated different methods to benchmark, document and monitor efficiency and energy savings as a result of energy conservation measures (ECMs) within the Norwegian food retail sector. The papers demonstrates that the potential for energy savings is substantial - savings up until 56% is documented. The average food retail store in Norway consumes 500 000 kWh. Hence, based on the average estimated savings in this thesis the potential energy reduction is 35%; annually kWh 175 000. Apply this to the 4000 food retail stores in Norway the energy saving potential is 700 GWh, the same amount of energy that 41 893

Norwegian households consume.

Through the delivered web application, the ESCO has access to near real-time energy saving models. Result and reports from the application has been actively promoted in new business cases, and the feedback from the stakeholders is that the application is an integral part of day-to-day business. The baseline web application module is today in operation as a tool to document and monitor energy savings for several hundred food retail stores.

At the start of this project, it was recognized that an important barrier that hinders renovation projects is uncertainty regarding the expected savings, and that the main objective of this thesis was to contribute to lower that barrier and to deliver reliable methods to be used to document and monitor energy savings in retrofitting projects. Through the presented papers it has been demonstrated that statistical models can be an important component to promote a sustainable and energy efficient food retail sector.

6 References

Bandara, Kasun, Rob J Hyndman, and Christoph Bergmeir. 2021. "MSTL: A Seasonal-Trend Decomposition Algorithm for Time Series with Multiple Seasonal Patterns," July. https://arxiv.org/abs/2107.13462.

Bogetoft, Peter, and Lars Otto. 2020. *Benchmarking with DEA and SFA*. https://cran.r-project.org/web/packages/Benchmarking/.

Brunke, J C, M Johansson, and P Thollander. 2014. "Empirical Investigation of Barriers and Drivers to the Adoption of Energy Conservation Measures, Energy Management Practices and Energy Services in the Swedish Iron and Steel Industry." *Journal of Cleaner Production*.

Burak Gunay, H, Weiming Shen, and Guy Newsham. 2019. "Data Analytics to Improve Building Performance: A Critical Review." *Autom. Constr.* 97 (January): 96–109. https://doi.org/10.1016/j.autcon.2018.10.020.

Chang, Winston, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert, and Barbara Borges. 2021. *Shiny: Web Application Framework for r*. https://CRAN.R-project.org/package=shiny.

Committee, IPMVP. 2016. *International Performance Measurement and Verification Protocol (IPMVP)*. https://evo-world.org/en/products-services-mainmenu-en/protocols/ipmvp.

EIA. 1999. "The Commercial Buildings Energy Consumption Survey." 1999. https://www.eia.gov/consumption/commercial/data/1999/index.php?view=methodology.

Franconi, Ellen, Matt Gee, Miriam Goldberg, Jessica Granderson, Tim Guiterman, Michael Li, and Brian Arthur Smith. 2017. "The Status and Promise of Advanced M&V: An Overview of "M&V 2.0" Methods, Tools, and Applications." https://doi.org/10.2172/1350974.

Galvin, Ray. 2014. "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'." *Energy and Buildings*. https://doi.org/10.1016/j.enbuild.2013.11.004.

Granderson, Jessica, Samir Touzani, Samuel Fernandes, and Cody Taylor. 2017. "Application of Automated Measurement and Verification to Utility Energy Efficiency Program Data." *Energy Build.* 142 (May): 191–99. https://doi.org/10.1016/j.enbuild.2017.02.040.

Grillone, Benedetto, Stoyan Danov, Andreas Sumper, Jordi Cipriano, and Gerard Mor. 2020. "A Review of Deterministic and Data-Driven Methods to Quantify Energy Efficiency Savings and to Predict Retrofitting Scenarios in Buildings." *Renewable Sustainable Energy Rev.* 131 (October): 110027. https://doi.org/10.1016/j.rser.2020.110027.

Guideline, Ashrae, and Others. 2014. "Measurement of Energy, Demand, and Water Savings." In *ASHRAE*

Guidel 14-2014.

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. https://doi.org/10.1016/j.ijforecast.2013.07.001.

Hong, Tao, Pu Wang, and Laura White. 2015. "Weather Station Selection for Electric Load Forecasting." *Int. J. Forecast.* 31 (2): 286–95. https://doi.org/10.1016/j.ijforecast.2014.07.001.

Johansson, Maria T., and Patrik Thollander. 2018. "A Review of Barriers to and Driving Forces for Improved Energy Efficiency in Swedish Industry–Recommendations for Successful in-House Energy Management." *Renewable Sustainable Energy Rev.*

Kontokosta, Constantine E. 2016. "Modeling the Energy Retrofit Decision in Commercial Office Buildings." *Energy Build.* 131 (November): 1–20. https://doi.org/10.1016/j.enbuild.2016.08.062.

Kuhn, Max. 2020. *Tidypredict: Run Predictions Inside the Database*. https://CRAN.R-project.org/package=tidypredict.

Lee, P, P T I Lam, and W L Lee. 2015. "Risks in Energy Performance Contracting (EPC) Projects." *Energy Build.* 92 (April): 116–27. https://doi.org/10.1016/j.enbuild.2015.01.054.

Minetto, Silvia, Sergio Marinetti, Pietro Saglia, Nina Masson, and Antonio Rossetti. 2018. "Non-Technological Barriers to the Diffusion of Energy-Efficient HVAC&R Solutions in the Food Retail Sector." *International Journal of Refrigeration* 86 (February): 422–34. https://doi.org/10.1016/j.ijrefrig.2017.11.022.

Muggeo, Vito M. R. 2022. *Segmented: Regression Models with Break-Points / Change-Points (with Possibly Random Effects) Estimation*. https://CRAN.R-project.org/package=segmented.

Ochieng, E G, N Jones, A D F Price, X Ruan, C O Egbu, and T Zuofa. 2014. "Integration of Energy Efficient Technologies in UK Supermarkets." *Energy Policy*. https://doi.org/10.1016/j.enpol.2013.12.002.

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.

Taieb, Souhaib Ben, and Rob J Hyndman. 2014. "A Gradient Boosting Approach to the Kaggle Load Forecasting Competition." *International Journal of Forecasting* 30 (2): 382–94. https://doi.org/10.1016/j.ijforecast.2013.07.005.

Tian, Wei, Yeonsook Heo, Pieter de Wilde, Zhanyong Li, Da Yan, Cheol Soo Park, Xiaohang Feng, and Godfried Augenbroe. 2018. "A Review of Uncertainty Analysis in Building Energy Assessment." *Renewable Sustainable Energy Rev.* 93 (October): 285–301. https://doi.org/10.1016/j.rser.2018.05.029.

Touzani, Samir, Baptiste Ravache, Eliot Crowe, and Jessica Granderson. 2019. "Statistical Change Detection of Building Energy Consumption: Applications to Savings Estimation." *Energy Build.* 185 (February): 123–36. https://doi.org/10.1016/j.enbuild.2018.12.020.

Wang, Pu, Bidong Liu, and Tao Hong. 2016. "Electric Load Forecasting with Recency Effect: A Big Data Approach." *Int. J. Forecast.* 32 (3): 585–97. https://doi.org/10.1016/j.ijforecast.2015.09.006.

Papers

6.1 Paper 1: Statistical learning to estimate energy savings from retrofitting in the Norwegian food retail market

Renewable and Sustainable Energy Reviews 167 (2022) 112691

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

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

A. Severinsen [a,*], Ø. Myrland [b]

[a] Norwegian University of Life Sciences, School of Economics and Business, Campus Ås, Universitetstunet 3, 1433 Ås, Norway
[b] UiT The Arctic University of Norway, School of Business and Economics, Breivangveien 23, 9010, Tromsø, Norway

ARTICLE INFO

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

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 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.

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

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.

* Corresponding author.
E-mail addresses: alexander.severinsen@nmbu.no (A. Severinsen), oystein.myrland@uit.no (Ø. Myrland).

https://doi.org/10.1016/j.rser.2022.112691
Received 15 March 2022; Received in revised form 18 May 2022; Accepted 5 June 2022
Available online 16 June 2022

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 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 4391 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.

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 41	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

1.1. Novelty of the paper

First, the TVB-model, published in Ref. [7], has been used frequently 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 Ref. [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 2.5 km × 2.5 km grid of Norway. Further, the temperature data gathered is modeled weather data that use several of the closest weather stations to set the temperature.

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

Furthermore, Fig. 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.

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.

Fig. 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.

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. Fig. 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 Ref. [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

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

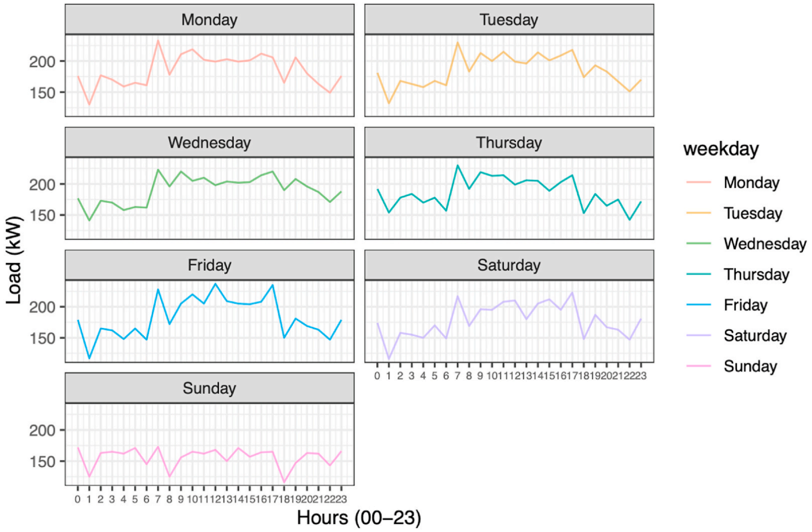


Fig. 1. Hourly loads (kW) throughout a week.

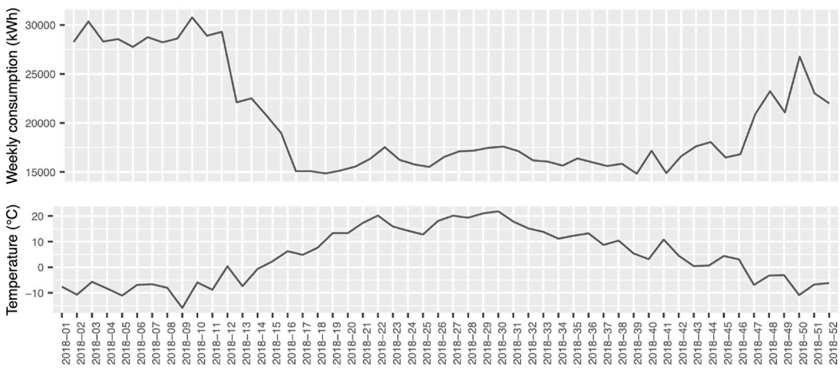


Fig. 2. Weekly energy consumption (kWh) and site specific weekly average temperature (°C) for a food retail store.

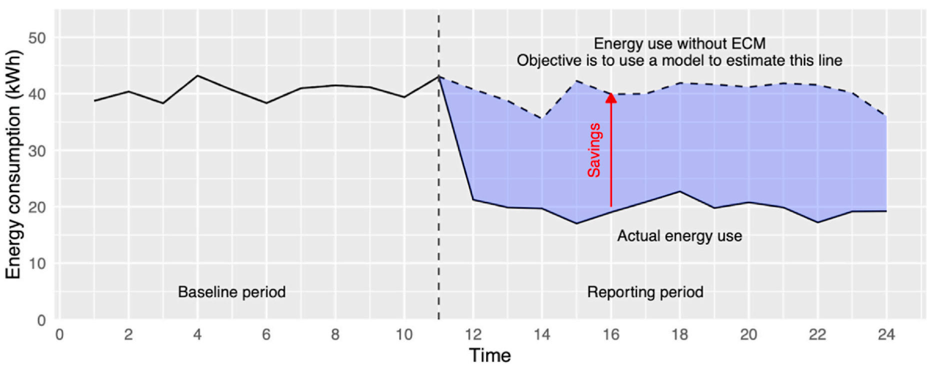


Fig. 3. Illustration of measurement and verification. Baseline = before the retrofitting, Reporting period = after the ECM.

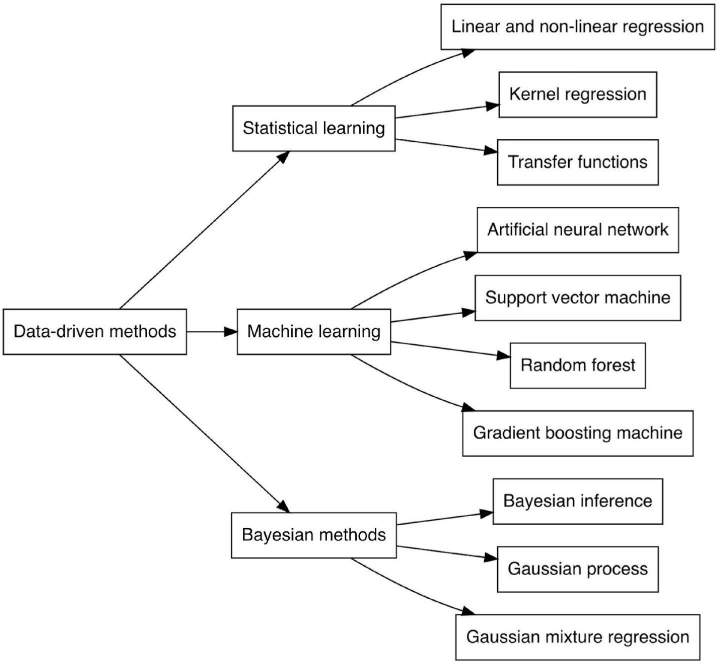


Fig. 4. Overview of baseline modeling approaches to estimate energy savings.

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 Fig. 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 Ref. [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 Ref. [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 Ref. [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.

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 $\mu = E[Y]$ 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 Ref. [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

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

where $I(\cdot) = -I(\cdot)$ and γ is the parameter to be interpreted as a reparameterization 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 $\psi = \psi + \widetilde{\gamma}/\widetilde{\beta}_2$.

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 Ref. [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,

$$CV - RMSE = \frac{\frac{\sum (\widehat{Y}_i - Y_i)^2}{n-k-1}}{\overline{Y}} \tag{4}$$

where $\overline{Y}$ 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*, $\widehat{Y}_i$ 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 and 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 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 Fig. 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

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	0087	0,170	−56,0	−55,4
4396	1 036 211	554 757	1 032 992	0033	0,075	−46,5	−46,3
479	204 767	112 911	202 194	0,056	0085	−44,9	−44,2
4103	417 657	295 295	418 866	0,044	0134	−29,3	−29,5
4097	529 024	397 062	529 776	0,021	0087	−24,9	−25,1

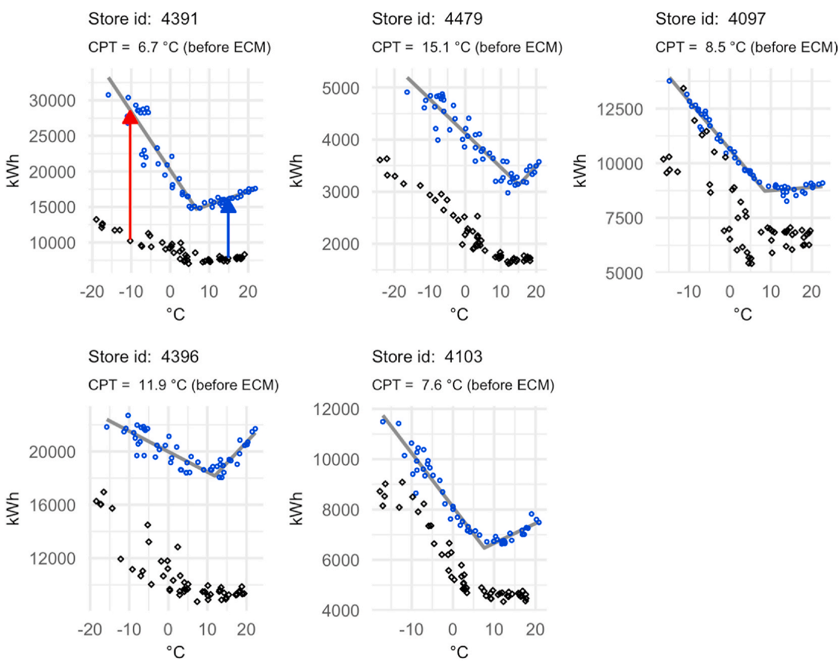


Fig. 5. ET curves for the reference period estimated with the broken line model, and the weeks for 2021 displayed.

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 7850 kWh actual versus 16 000 kWh predicted, a saving of 8150 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 4391 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.

3.3. TVB model - savings on the hour

In Fig. 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 Fig. 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.

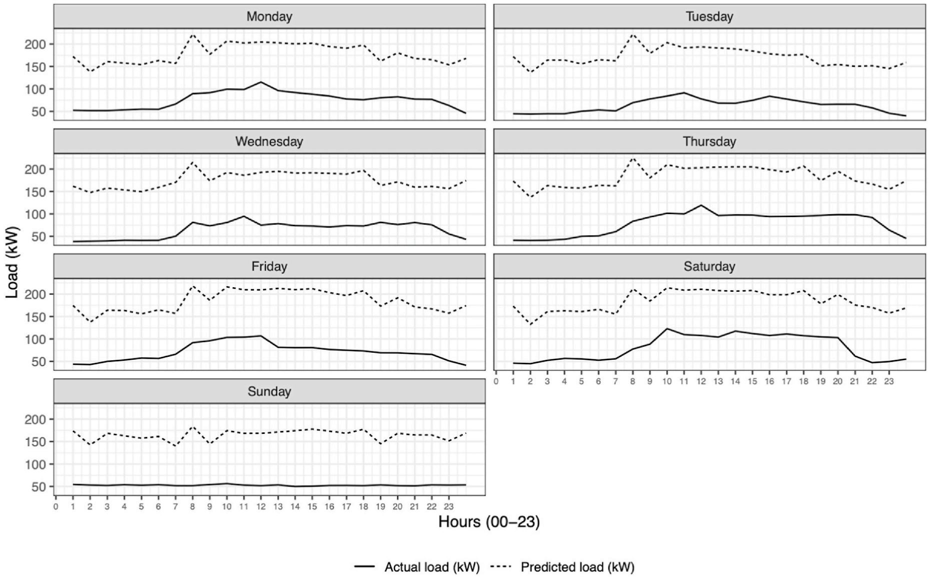


Fig. 6. Actual and predicted loads (given no ECM) from the TVB model.

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 Fig. 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* 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 6596. 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.

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

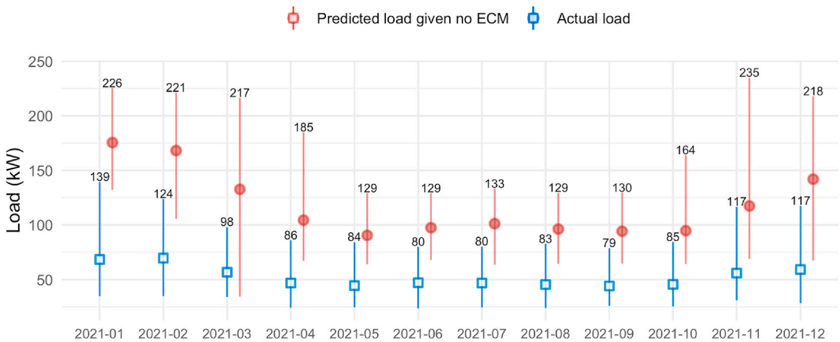


Fig. 7. Average, minimum and maximum load (kW) for each month. Results from TVB model, showing the peak shaving that occurred on a monthly basis for the store with the largest energy savings.

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 Ref. [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 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 Ref. [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 4370 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 Ref. [47]. The top 5 solutions were reproduced by Ref. [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-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.

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

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 weather 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

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

[3] https://www.kaggle.com/c/ashrae-energy-prediction.

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

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 Ref. [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 Ref. [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 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.

[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.

Authorship contributions

Category 1
Conception and design of study: A. Severinsen, Ø. Myrland.
Acquisition of data: A. Severinsen.
Analysis and/or interpretation of data: A. Severinsen, Ø. Myrland.
Category 2
Drafting the manuscript: A. Severinsen.
Revising the manuscript critically for important intellectual content: A. Severinsen, Ø. Myrland.
Category 3
Approval of the version of the manuscript to be published (the names of all authors must be listed): A. Severinsen, Ø. Myrland.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Alexander Severinsen reports financial support was provided by Trønder Energi AS. Alexander Severinsen reports a relationship with Trønder Energi that includes: employment and equity or stocks. The author, Alexander Severinsen, have the last 4 years worked on an Industrial PhD.

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. Available from: https://www.eia.gov/consumption/commercial/data/1999/index.php?view=methodology. [Accessed 14 January 2022].
[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. Energy Build 2014;69:515–24.
[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. 2021. https://leanpub.com.
[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. Renew Sustain 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. Comput Stat 1999;12:249–64.
[16] Feder PI. The log likelihood ratio in segmented regression. AOS (Acta Odontol Scand) 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, vol. 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. Comput Sci Appl – ICCSA 2020;2020. 201–13.
[22] Fallahi Z, Smith AD. A comparison of commercial building retrofits using EnergyPlus for energy and emissions savings, 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. Renew Sustain Energy Rev 2013 Jul;23:272–88.
[26] Zhao H-X, Magoulès F. A review on the prediction of building energy consumption. Renew Sustain 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. Renew Sustain 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. Renew Sustain Energy Rev 2015 Oct;50:1352–72.
[29] Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. Renew Sustain 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. Renew Sustain 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 Eng 2017 Jan;205:2370–7.
[36] Diaz JA, Ramos JS, Delgado MCG, Garcia DH, Montoya FG, Dominguez SA. 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. J Build Perform Simul 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] Lindelof 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 TJTR. 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 PROCEEDCGS. 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. Renew Sustain 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?. In: Explaining the predictions of any classifier; 2016 Feb. Available from: https://arxiv.org/abs/1602.04938.

6.2 Paper 2: Quantification of Energy Savings from Energy Conservation Measures in Buildings Using Machine Learning

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

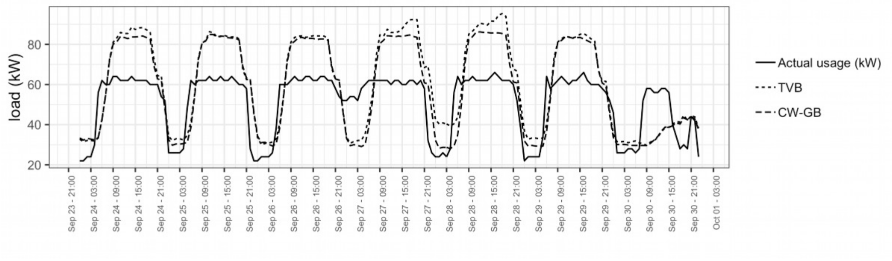


Figure 1. Comparing TVB versus CW-GB against actual energy use after ECM.

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,

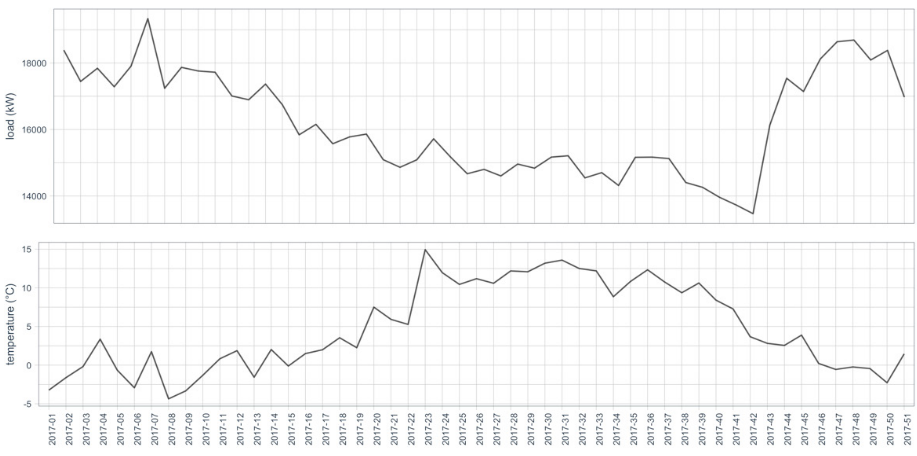


Figure 2. Load (kW) and temperature for the year 2017 for store number 7.

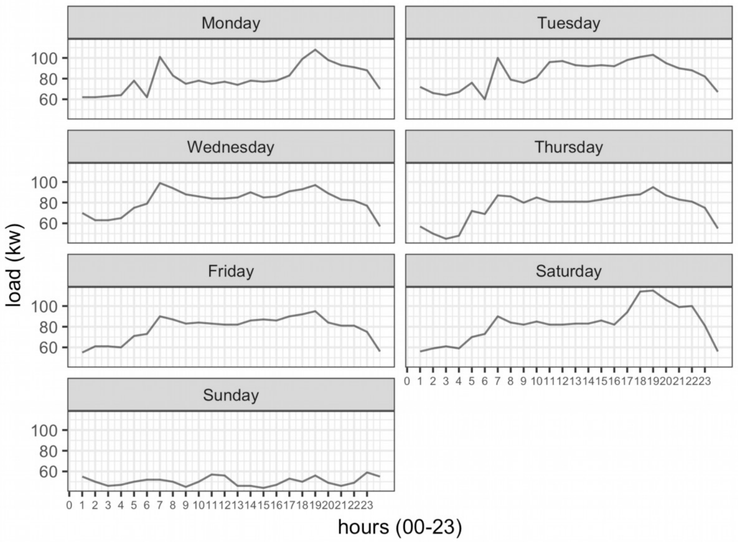


Figure 3. Hourly loads (kW) throughout a week.

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:

$$CV(RMSE) = \frac{\frac{\sum(\widehat{Y}_t - Y_i)^2}{n-p-1}}{\overline{Y}} \tag{3}$$

where $\overline{Y}$ is the mean of the number of measured energy values in the training data, Y_i is the actual energy usage in hour i, $\widehat{Y}_t$ 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

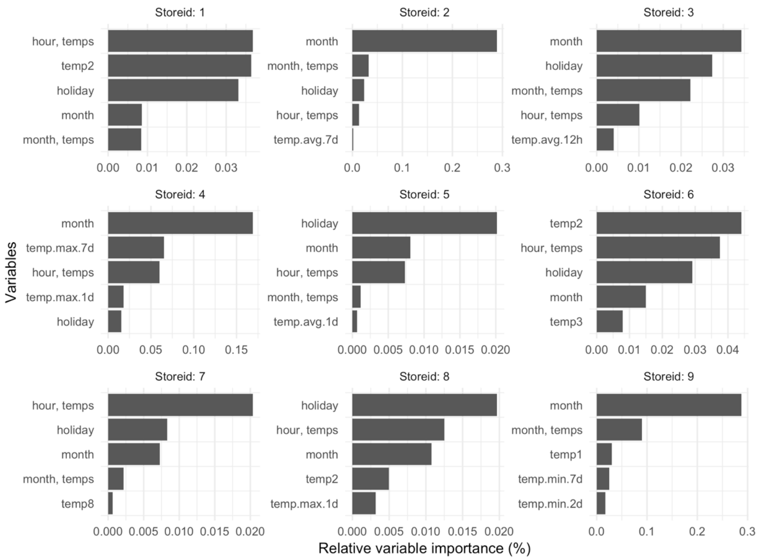


Figure 4. Relative importance of the different variables.

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-

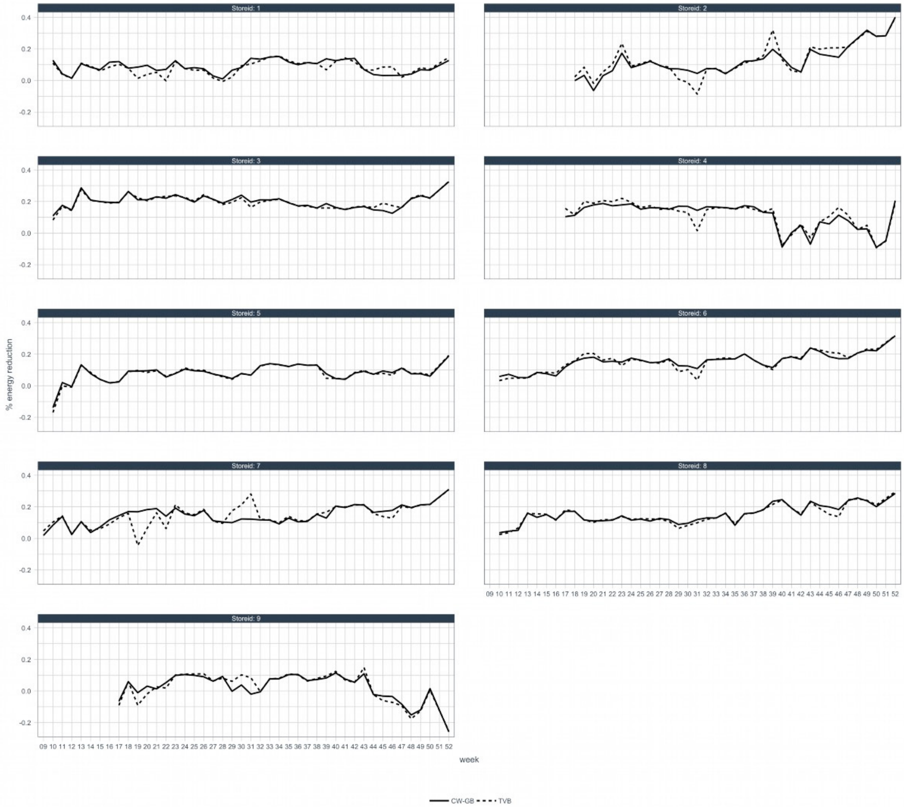


Figure 5. Estimated weekly savings for the 9 stores.

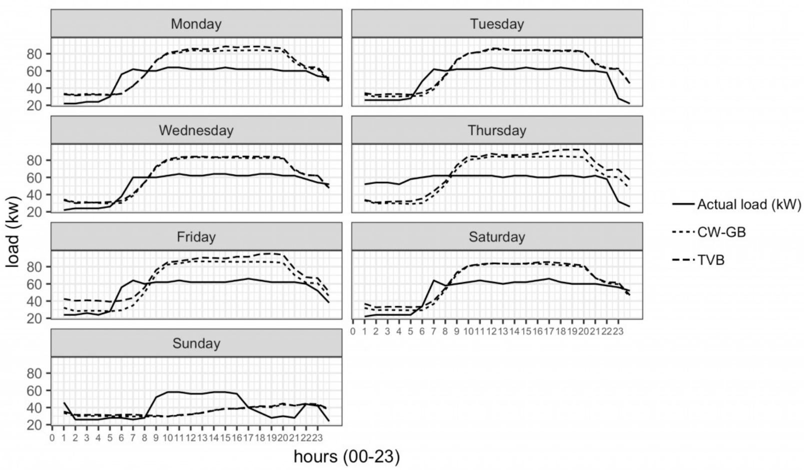


Figure 6. Actual loads and predicted loads for store number 4 in week 39.

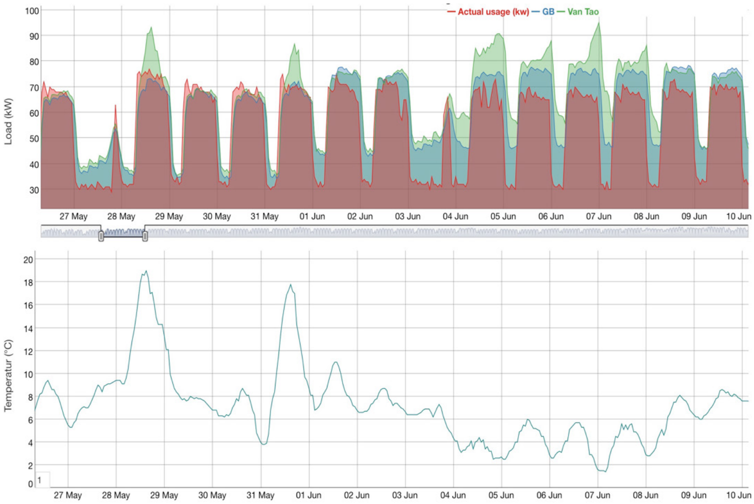


Figure 7. Example web-application. The predicted loads and corresponding temperatures.

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.

6.3 Paper 3: ShinyRBase: Near real-time energy saving models using reactive programming

Applied Energy 325 (2022) 119798

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

ShinyRBase: Near real-time energy saving models using reactive programming

A. Severinsen[a,*], Ø. Myrland[b]

[a] Norwegian University of Life Sciences, School of Economics and Business, Campus Ås, Universitetstunet 3, 1433 Ås, Norway
[b] UiT The Arctic University of Norway, School of Business and Economics, Breivangveien 23, 9010 Tromsø, Norway

HIGHLIGHTS

- Demonstration of an open-source tool using a reactive programming framework for measurement and verification and energy saving models.
- Fast development cycle without any need-to-know web programming languages like HTML, CSS or JavaScript.
- A use case documents energy savings in 40 different Norwegian food retail stores.

ARTICLE INFO

Keywords:
Real-time energy savings evaluation
Building energy retrofitting
Measurement and verification 2.0
Data driven models
Tao Vanilla Benchmarking model

ABSTRACT

To document energy savings from retrofitting a building, a reliable baseline model is needed. The development and implementation of the baseline model is an important step in the measurement and verification (M&V) process. Usually, an energy analyst enters the stage, collects data, do the estimation and delivers the baseline model. The modeling work of the energy analyst is done on either a proprietary or open-source statistical software, often using a coding script. If stakeholders want an updated report on energy savings, the analyst must re-do the whole process, for example on a monthly basis. This workflow is based on an *imperative* programming paradigm. The analyst holds on to the code that performs the analysis and re-run the code when agreed upon. The consequence of this workflow is that stakeholders are dependent on the energy analyst and that updated energy savings results must be planned and scheduled. However, emerging M&V 2.0 technologies enables automation of the energy saving reports. This paper demonstrates how energy savings from retrofitting's in the Norwegian food retail sector is continuously monitored and documented in a web application. The application is built using open-source tools where the baseline model is delivered through a *reactive* programming framework. As an energy savings baseline model, the Tao Vanilla benchmarking model (TVB) was set into production in the web application. The TVB is a linear regression model with well specified features, easy to interpret and has a history of excellent prediction performance. The proposed web application framework allows for a fast development cycle without any need-to-know web programming languages like HTML, CSS or JavaScript. The reactive framework delivers several advantages. First, the stakeholders will always have a current and real-time report on the savings. Second, complex methodologies are dynamically used by the end-user. Third, increased involvement by stakeholders and interaction with the analyst related to the methods used in the energy savings analysis leads to collaborative benefits such as faster disseminating of knowledge. These synergy effect leads to a better technical understanding from the end user perspective and enhanced practical understanding for the analyst. Finally, the paper presents an integrated look at the energy kWh savings versus the cost of the retrofitting's.

1. Introduction

The Intergovernmental Panel on Climate Change (IPCC) objective is to provide governments with scientific information to develop climate policies. The IPCC has a 195 member countries and thousands of contributors. The IPCC scientists assess thousands of published scientific papers each year, and the most recent report is summarized with the following quote,

* Corresponding author.
E-mail addresses: Alexander.severinsen@nmbu.no, alexander.severinsen@nmbu.no (A. Severinsen), oystein.myrland@uit.no (Ø. Myrland).

https://doi.org/10.1016/j.apenergy.2022.119798
Received 11 April 2022; Received in revised form 26 July 2022; Accepted 3 August 2022

It is unequivocal that human influence has warmed the atmosphere, ocean and land. Widespread and rapid changes in the atmosphere, ocean, cryosphere and biosphere have occurred [1].

The building sector worldwide uses about 32 percent of all generated energy and 19 percent of energy-related greenhouse emissions [2]. Research show that the potential for increased efficiency in terms of energy cuts are some 30 to 40 percent [3]. Green buildings receive a lot of attention, especially in new design construction. And many new building standards have been developed internationally. For example, the German Passiv House Institute (PHI), Net Zero from the International Living Futures Institutes, Leadership in Energy and Environmental Design (LEED), and R-2000 by Natural Resources Canada. In Norway the standards NS 3700 for residential buildings and the NS 3701 for non-residential buildings apply. These standards are important to make sure that the environmental impact of new buildings stay as low as possible. Still, only about 1 to 3 percent of old buildings are replaced per year [3], thus the existing buildings will still be around in the foreseeable future. According to Statistics Norway the existing building stock in Norway is 4.23 million, and more than 2.6 million of these buildings are non-residential [4]. The number of buildings globally and the potential for energy reductions makes retrofitting an important addition to cut greenhouse gas emission and support the Paris climate agreement. Also, according to analysis by the Rockefeller Foundation and Deutsche Bank's climate change shop the business opportunity in retrofitting is substantial. An investment in the United States of $279 billion in retrofitting buildings could yield more than $1 trillion in energy savings over ten years. This equals 30 percent of the countries annual spending on electricity, and represents a created potential of more than 3.3 million cumulative job years of employment. Nonetheless, the existing upgrade rate is only 2.2 percent each year [3,5].

In this paper we collaborate with a medium sized Norwegian energy service company (ESCO) that has specialized in retrofitting within the retail food sector. The ESCO has recently agreed upon an energy performance contract (EPC) for 40 Norwegian food retail stores. The contract includes a yearly guaranteed energy savings target, and the ESCO also handles all the energy related issues within the contract period, including service and maintenance. The energy conservation measures (ECM) were implemented in the autumn of 2020 and consisted of a mix of the following: change to more efficient LED lightning, new coolers and/or freezers, new heating, ventilation and air conditioning (HVAC) and/or optimization of the stores control systems. The contract has a 10-year duration, and the first year of measuring the energy savings is 2021.

The main contribution of this paper is to demonstrate how to document and monitor energy savings using a dynamic web application based on a open-source reactive programming framework. The development of the application has been done in collaboration with the ESCO and other stakeholders in the EPC project over a two year period. The implementation of the application as a means of setting up baseline models and monitoring the energy savings for the 40 food retail stores has led to several synergy benefits; both collaborative and practical. The energy savings results from the modelling process was seen in real time in the application, on an hourly basis, and automatically updated each day. As a results of this there was no dependency on a energy analyst. The reports was always up-to-date, and the knowledge sharing between participating parties with different skill sets resulted in better baseline models. There was quick detection of unwanted energy increases and follow up of potential errors in the buildings technical system. In short; the web application gave the users a more coherent and reliable process of documenting the energy savings. This is in line with both [6] and [7] who find that web applications facilitate efficient collaboration between scientists and stakeholders and that the cross collaboration between researchers and easy dissemination of results is important for external validity [8].

Today, there are several web-based systems available that can monitor the energy consumption in buildings, but they do not offer any advanced modeling solutions, and are often based on proprietary software [9]. Open-source solutions are not common, and previous solutions focus on visualisations and reporting of electricity consumption, and does not offer baseline energy saving models [10,11]. Up until recently there has been little available research or case studies that document the use of M&V 2.0 energy saving estimation [12]. However, in 2018 [13] presented a residential energy management system, *reEMpy*, that is based on Python. This system is also aimed to energy service companies (ESCOs) to provide a solution that may be used to assess the energy needs of real life use cases by evaluating different algorithmic models, including load forecasting. Nonetheless, our solution differs in terms of the approach. Our proposed framework does not offer code for a final ready-made applications, but a reactive coding framework to allow quickly prototyping new functionality and baseline energy saving models in close collaboration with end-users. Furthermore, [14] proposes a platform for real-time M&V of energy performance. The platform computing tier is developed using the Java development toolkit within the Eclipse software. Conversely, the reactive framework presented in this paper requires no prior development skills. Also, in [14] the platform is demonstrated on one utility, while the usecase for the presented framework, *ShinyRBase*, is given for 40 food retail stores.

The web application *ShinyRBase*, was developed and implemented using R, a free software environment for statistical computing and graphics [15]. R is one of the most popular programming languages for statistics. Furthermore, the R library Shiny [16] was used to develop the reactive programming framework. Shiny makes it very easy to build interactive web applications straight from R without any need to know HTML, CSS or JavaScript. Additionally, Shiny makes it straightforward to use more than 18.000 available packages for a wide range of applications[1] [17]. For example, the popular R packages ggplot [18] and dygraphs [19] is used as tools for interactive visualization and to enhance user interaction with the baseline energy savings model. Also, tidypredict [20] is implemented to save and run predictions for the relevant models inside a database. As will be shown the reactive application gives offers a number of advantages to promote M&V 2.0. The Shiny library, which is the main component in the application, has been in active development since 2016 and has more than 5000 unique peer reviewed works to promote user interaction with scientific research [21].

1.1. Novelty of the paper

The novel contributions of the present papers are threefold. First, the application is developed in collaboration with implementers and utilities. The reactive programming framework, as opposed to the imperative paradigm, offers a flexible and dynamic application that easily can be adapted to different building types. For instance, the vast number of R libraries can be integrated to deliver on changing M&V environments. To be able to convince investors in energy efficiency projects it is critical to provide current and trustworthy energy savings calculations. In that respect, the proposed reactive framework in this paper contributes to the current literature.

Second, the application is demonstrated on a use case continuously monitoring and documenting the energy savings of a large retrofitting project for 40 Norwegian food retail stores. This is the first paper that document the energy savings of a renovation project of this scale within the food retail sector in Norway. Previous research has established that uncertainty regarding the expected savings is a major obstacle that hinder renovations projects [46]. As such, the reactive framework and the documented energy savings in this paper is an example that may motivate new renovation projects, and ultimately produce cost savings

[1] https://cran.r-project.org/web/packages/#:~:text=Currently%2C%20the%20CRAN%20package%20repository%20features%2018316%20available%20packages.

and a reduction of GHG emissions.

Third, the baseline model used to estimate energy savings in the web application is a linear regression model, the Tao Vanilla Benchmarking model (TVB). The model was first published in [23], and was later used in the GEFCom2012 load forecasting competition as one of the top 25 performing contributions [24]. The TVB model is evaluated through the guidelines outlined in The International Performance Measurement and Verification Protocol (IPMVP), developed by the Efficiency Valuation Organization (EVO) [25]. In particular, the IPMVP measure "option C: whole building" is applied. Hence, data from utility meters are used to evaluate the energy performance of the whole building. Note that this option establishes the total savings of all implemented ECMs. In previous research the TVB was used in [26] to document energy savings from retrofitting's with small expected effects. Further, [27] use the TVB model to document energy savings in 5 different Norwegian food retail stores that undertook major retrofitting's. However, in the present paper the TVB is implemented to document energy savings for food retail stores within a reactive framework, near-real time for 40 different stores. 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 as a benchmark model, and adds to the already established data-driven tools within the M&V industry.

The paper is split into 6 different sections. In the first section a presentation of the ESCO and details about the EPC contract and the food retail stores is given. Second, an overview of the data and the features that will be used for the TVB models. Third, the TVB model and the measures used for model performance is outlined. Furthermore, the fourth section gives a detailed presentation on how the Shiny library's reactive programming framework is used to implement the TVB model. Fifth, the energy savings results from the EPC project is presented. The results are shown as they appear in the web application. Finally, the findings linked to both individual and synergy effects of this project based on user interaction with the application is given. Also, future development improvements that would have given the application more value is presented, both from a scientific and a practitioners point-of-view.

1.2. *The ESCO and the energy performance contract*

The energy service company (ESCO), Ohmia Retail AS, is a medium sized Norwegian company. The company has developed a product that is marketed towards the food retail sector in Norway as '*Energy as a service.*' The customer pays a fixed monthly fee and need not to worry about necessary equipment investments, insurance and maintenance. Thus, Ohmia Retail takes full responsibility for all the stores technical infrastructure; freezers, coolers, lighting, and the HVAC system. For the food retailer this is a great asset as they can maintain their primary activity, food retail. The ESCO has signed an Energy performance contract (EPC) that includes a guaranteed energy savings for a building portfolio of 40 food retail stores in Norway. The contract has a 10 year duration, and started January 1st 2021. Hence, to document and monitor the guaranteed energy savings it is important for both the ESCO and the customer to utilize reliable methods. Historically the ESCO has used very basic methods to document savings, based on degree day normalization and often using quarterly reports. Thus, the ESCO had great interests in methods that could improve their monitoring workflow, which again motivated this research project.

1.3. *Building portfolio, electric load and weather data*

Norway has a central repository, Elhub (elhub.no), that daily collects energy use on an hourly level for almost all commercial and household buildings using the advanced metering (AMI) system. It is mandatory for all the Norwegian grid operators to update the central repository every day. This service was launched in February 2019. All the energy data from February 2019 and onward for the 40 buildings in this paper originates from Elhub. The energy data from 2018 up until January 2019 is collected from the building energy management system that was in operation before the launch of Elhub. Outside temperature data is collected from the Norwegian Meteorological Service (https://www.met.no). Each stores position (longitude and latitude) is mapped against a 2.5 km × 2.5 km grid of Norway. The temperature data used in the model stems from the closest weather stations. All energy and weather data is downloaded automatically on an hourly level on a daily basis into a PostgreSQL database. The PostgreSQL is an open-source object-relational database with more than 30 years of active development [28]. This database has a central role in this project. The results from the baseline model is stored, documented and used for daily predictions, all within the database. The next sections present the model and the framework used to set the model into production.

2. Methods

2.1. *Estimating energy savings - the baseline model*

To estimate the energy savings this paper follows the International Performance Measurement and Verification Protocol (IPMVP) option C: whole building". Thus, data from utility meters are used to evaluate the energy performance of the whole building [25]. In particular hourly energy and temperature data is used to train a baseline energy saving model. This process involves choosing representative energy data for one whole year before any retrofitting is conducted, and then train a model to predict (after the retrofitting) what the energy consumption would have been without the ECMs. Furthermore, the Tao Vanilla benchmark (TVB) model [23] is used to estimate energy savings. Some energy conservation measures, for example changing to LED lights, may have an expected savings target below 10 %. In these cases the EVO recommends using sub meters as a means of documenting the savings. However, in this project there was no data available from sub-meters. Nonetheless, [26] finds that the TVB model is a good candidate to estimate savings that are below 10 %. Moreover, The TVB model is a well specified regression model and easy to estimate and understand. The following model specification is used:

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

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. Furthermore, T_t is the outside temperature corresponding to time t. Note that the original TVB model includes trend and past loads. In this study the TVB model will reflect how a specific building perform based on a reference period, thus trend and lagged predictors are not included. The simplicity of model choice has several advantages for implementation in the web application. Also, as seen in the next section all the models were estimated within the web application.

Because the load often increases when the temperature drops and when the temperature increases, it is necessary to take this into account. This could be incorporated with linear piecewise functions. However, that would require cut-off temperatures which may be different across different buildings. Thus, this is included in the model using 3rd ordered polynomials of the temperature. Also, the model includes interaction effects between the polynomials of the temperature and the calendar variables *Hour* and *Month*, respectively. The rationale behind is that energy loads may be different when the temperature varies, and there might be differences across different months and hours.

Furthermore, to come up with the model specification, in his thesis [23], Hong tested seven different linear regression models. The testing was demonstrated using the case study of one week ahead hourly forecast for a medium US utility, and the TVB model was found to have best performance. Furthermore, the TVB load forecasting performance was also tested against possibilistic linear models (PLM) [23], p. 117] and artificial neural networks (ANN) [23], p. 136]. The TVB outperformed both modeling alternatives.

The performance of the TVB models is measured with the *coefficient of variation root mean square error* (CV-RMSE). The CV-RMSE is calculated as follows:

$$CV-RMSE = \frac{\frac{\sum(\widehat{Y_i}-Y_i)^2}{n-k-1}}{\overline{Y}}$$

where $\overline{Y}$ is the mean of the measured energy consumption in the training data (the reference year). Y_i is the actual energy use in hour i, $\widehat{Y}_i$ is the predicted value of energy use in hour i from the TVB model, estimated on the baseline period. Furthermore, n is the sample size, and k is is the number of features in the model. This performance measure is recommended by the IPMVP [25]. However, the guideline does not propose a definite threshold that defines a good baseline model. The CV-RMSE has received some critique because the RMSE increase with both the average absolute error and the variance error, which is not desirable [29]. Still, since this paper does not compare between models the CV-RMSE is a good indicator to measure the performance of the models between different training data and different buildings. Furthermore, a "rule-of-thumb" where the CV-RMSE should be below 25 % is applied. In the next section when the web application and the reactive framework is presented it is also shown that within the application it is easy to switch reference year (training data) between the last three years, and the corresponding change in CV-RMSE can be seen instantly. Setting up the model and choosing reference year was done in collaboration with both the ESCO and the building owners. Large CV-RMSE (greater than25 %) was immediately inspected through visualizations of the data in the reference period and could in many instances be led back to the food store being out of operation and/or some extraordinary activity, for example in-store promotion activities. Consequently, the CV-RMSE was actively used to understand the modeling results together with the stakeholders, and for many buildings this collaboration led to a change of reference year and/or a better understanding of the actual data used as training data, eg. missing and/or unusual data.

2.2. The reactive framework energy monitoring solution

This section presents the reactive framework that was used to develop the web application that documents and monitors energy savings. The ESCO and their customers typically depends on an energy analyst to deliver the analysis that documents the energy savings in a project. The usual way of delivering results from statistical models is based on imperative programming, e.g. when $c = a + b$ then c is assigned the sum the terms a, b. If a, b changes then c needs to be re-evaluated to change. However, reactive programming will allow c to be updated instantly when a, b change [30]. This reactivity is the main idea behind the R library Shiny [16]. Shiny helps the energy analyst to promote real-time user interaction with the analysis through a user interface (UI). The Shiny reactive framework allows user input to be evaluated dynamically via the user interface, and the library comes with pre-defined templates for web based user interfaces. This avoids the need to learn web based programming languages such as HTML and/or JavaScript. However, in contrast to the reactive framework, an analyst would write a script that runs all the necessary analysis to deliver a report of the energy savings. Then the analyst has to re-do all the analysis on a regular basis, often monthly. The script is re-run and a new updated report is produced. This way of working is referred to as *imperative* programming. One example of this is seen in the below R and SQL code sequence (comments indicated with a #). In short, the analyst;

1. Selects data needed for the analysis. In this case this means writing an SQL code that extracts energy and temperature data (store_id, date_hour, temperature, kWh) from a database table. The buildings id's and the date intervals has to be specified by the analyst. The query is then run to pull data into a dataset train, using an R function from a database connection library DBI::dbGetQuery(pool, sql).
2. A formula for the baseline model is defined (in this case the TVB model).
3. The regression model is run (through the R function lm) and the results is stored in the object TVB_estimates for further inspection.

2.3. Imperative code sequence

```
# Step 1 - pull data from database to R as specified by script
sql <- (
"SELECT
store_id, hour, temperature, kwh
sum(timeseries_interval_observations.value) AS kwh
FROM energy_data
WHERE id IN ('6754','6789')
AND date BETWEEN '2019–01-01' AND '2019–12-31';"
)
train <- DBI::dbGetQuery(pool, sql)

# Step 2 - define the formula for linear regression model
TVB_formula <- as.formula("kwh ~ hour*weekday + month +
hour*temperature + hour*temperature^2 +
hour*temperature^3 + month*temperature +
month*temperature^2 + month*temperature^3")
# Step 3 - run the linear regression
TVB_estimates <- lm(TVB_formula, data = train)
```

Now, let us look at how the imperative programming approach is handled by R using the Shiny library to create a reactive framework with some simple user inputs in a web application. There are 6 steps involved;

1. The analyst set up the input fields that the user can access and interact with in the user interface (UI). Shiny has pre-canned UI elements that is used to define the UI (radio buttons, date range input fields, checkbox, etc). In addition, the data analyst sets up a UI element to display the analysis (tables, graphics, text). In this simplified example this is only a output field for text where the analyst plans to show the model coefficients from the TVB model.
2. The server environment is defined.
3. Data is selected based on the UI inputs from step 1.
4. The formula is set up (same as in the imperative script).
5. The regression model is run. Every time the user chose new input in the UI, this step is instantly re-run.
6. Set up a render statement to be passed back and displayed in the UI. As for this example; the linear regression models coefficients.

This was a basic workflow example. Yet this process only adds a few lines of extra R code, but turns the analysis into an web application that can be used to interactively chose stores and train data for the TVB models. The part of the web application that will be presented in the results section is about 1000 lines of code. This includes functions to handle pulling and pushing data between R and the database and code for error handling. Strictly speaking, it would be possible to write the application with some 500 lines of code. Hence, in contrast to other web application this is quite efficient.

2.4. Reactive code sequence

```
# Step 1 - Define UI
ui <- fluidPage(
titlePanel("Reactive code sequence"),
sidebarLayout(
checkboxGroupInput(ns("selected_meters"), 'Main meter(s)', choices = NULL),
dateRangeInput(
"Referenceperiod:",
start = ref_start_date,
```

(*continued on next page*)

(*continued*)

```
end = ref_end_date,
),
mainPanel(
textOutput("SQL_model_scoring")
)
)

# Step 2 - Define server environment
server <- function(input, output) {

# Step 3 - pull data inta a R dataframe with a reactive expression
train_data <- reactive({
sql <- glue::glue_sql(
"SELECT
hour, month, temperature, kwh
sum(timeseries_interval_observations.value) AS kwh
FROM energy_data
WHERE id IN ({metering_points*})
AND date BETWEEN {date_from} AND {date_to};",
.con = pool,
.envir = list(
metering_points = selected_meters(), # user selected store
date_from = ref_start_date, # user chosen from-date
date_to = ref_end_date # user chosen to-date
)) %>%
DBI::dbGetQuery(pool, sql)

# Step 4 - define formula for linear regression
TVB_formula <- as.formula("kwh ~ hour*weekday + month +
hour*temperature + hour*temperature^2 +
hour*temperature^3 + month*temperature +
month*temperature^2 + month*temperature^3")

# Step 5 - run linear regression and translate the coefficients to SQL.
# Reactive expression. Will re-run every time a user change the input fields.
model_coefficients_as_sql <- reactive({
train <- train()
TVB_estimates <- lm(TVB_formula, data = train_data)
sql_statement <- tidypredict::tidypredict_sql(TVB_estimates, dbplyr::
  simulate_postgres())
return(model_coefficients_as_sql)
})
output$SQL_model_scoring <- renderText({
print(model_coefficients_as_sql)
})
}
```

In Fig. 1 the reactive framework used to design the web application is presented.

The web application framework consists of 7 different steps.

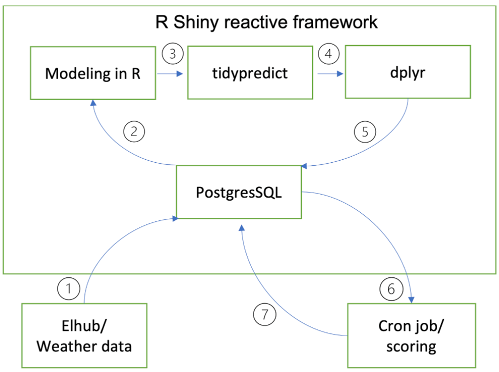


Fig. 1. Reactive framework.

1. Data from the Elhub repository (energy data on an hourly level from the main meter) is stored in a PostgreSQL database. The database is updated daily. This step takes place outside the R/Shiny reactive framework. However, the R library RPostgreSQL [31] handles data transfer between the database and R.
2. Modeling in R. This step is based on user inputs in the Shiny web application. The user chose the relevant building and the period to be used as training data. This is the reference period before the implemented ECMs. Typically, this would be one whole year of data (energy consumption and temperature data) before the installed ECMs. The user then chose the period to be predicted based on the trained model. This involves choosing the date that the ECMs were installed, and the period to be analyzed. Once these choices are made the app automatically runs the TVB model, the CV-RMSE is calculated and presented and the ECM period is scored and visualized (both the actual energy use and the predicted, the difference between these being the savings). Fig. 2 shows the UI that is available to the users. The blue line is the predicted kWh, while the red line is the actual energy consumption.
3. The R library tidypredict [20] reads the current model from step 2, creates a list objects with the necessary components to run predictions, and builds a formula based on the list object.
4. The R library dplyr [32] evaluates the formula through its database backend dbplyr and translates the code into SQL.
5. The translated SQL is stored in a text field in the database with additional details about the building, main meters, and the data used for the models. Not only will this SQL statement be available for continuously scoring new data, but it also works as a documentation for future reference of the model.
6. A cron job (automated Linux job scheduler) wakes up every morning when new data is stored in the database, pulls out the stored SQL from step 5 and use the new data and the SQL statement for new predictions. Hence, the energy savings from the last 24 h in added continuously.
7. The predictions are stored in the database (on an hourly level). This way the database is always updated with the most current data to calculate the energy savings. All the tables and visualizations in the application is based on this; hourly visualizations and aggregates of the results.

See Fig. 2 for a closer look at the UI from step 2 where the user interacts with the web application. In this particular example the reference year was 2018, and the ECM period was 2021. It is important to choose a year for training the TVB model that is representative of each building's energy usage before the installed ECMs. CV-RMSE is the "technical" indicator that guides the model making. However, the interactive user interface that pictures a year of energy and temperature data on an hourly level together with the prognosis from the TVB model made the modeling process more intuitive for the non-analytical building owners and the ESCO. They could observe how changing the reference base impacted the prognosis after the ECMs made the process of setting the reference data for the 40 buildings easier. This interactivity in the process was found to break down the complexity barrier and connect the stakeholders to the modeling process for increased collaborative benefits.

The entire model building process was done together with the ESCO and the building owners. First, a building under scrutiny was chosen, then year of reference was set. The reference year within the portfolio was either 2018, 2019 or 2020. The effect on choosing any of these could be seen immediately by comparing the predicted and the actual values in the time series plot of the UI. For example, one could train a model on energy data from 2018 and then predict the energy consumption in 2019. Given no ECMs in 2019 the actual value and the predictions should follow each other closely. This exercise generated a lot of very useful discussions - both for the building owner and the analyst. For example, the CV-RMSE was above 25 % for many of the models. Often,

Fig. 2. User interface example of TVB in the web application.

this could be related to issues such as; missing data, in-store promotion, fault with the AMI energy meters and/or some other unusual building activity. The visualizations was either performed using the R library ggplot2 [18] or the interface to dygraphs described in [19].

3. Results

3.1. Energy savings

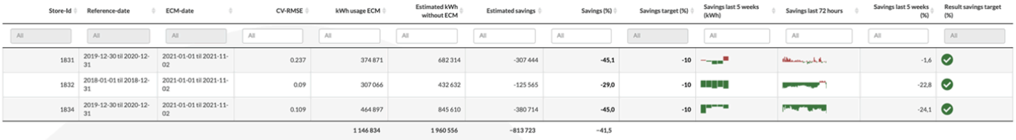
After a user has set up a baseline model with reference data and the ECM period, accepted the CV-RMSE and stored the baseline model, the stores end up in the web application as seen in Fig. 3. This table presents an example from the web application of three of the stores (Store-id 1831, 1832, 1834) with the largest energy savings. The savings are aggregated between 2021 and 01-01 and the current date. The table is updated every day based on the last 24 h. As can be seen the savings range from 45.1 % – 29.0 %. Also, the reference dates, ECM dates and the CV-RMSE for each underlying model are displayed. Additionally, the savings are presented in three different ways. First, the total savings (field: 'Estimated savings'), which is the actual kWh usage minus the predicted in the ECM period. Second, the results the last 5 weeks is presented as a bar plot in the field 'Savings last 5 weeks (kWh).' Each bar represents a week. The bars are colored green if saving energy, and red if actual energy use is above the predicted level. This gives the end user a way of quickly comparing the aggregated result with the results over the last 5 weeks. For example, store '1831′ in the first row has a total energy saving of 45.1 %, but only a 1.6 % savings the last 5 weeks. For this particular store this was somewhat expected as the implemented ECMs was related to winter ECM, since the store had changed the heating system. However, Fig. 3 also shows the savings the last 72 h as bars (each bars equals an hour; green color if saving, and red if higher use than predicted). This indicates that the store needs further investigation into possible causes of higher than expected energy usage.

Table 1 presents the aggregated results for each of the stores ECMs for the first 11 months of 2021, from January 1st 2021 up until November 30th 2021. The three stores with the largest percent energy savings where those with store-id 1522, 1249, and 1520 with energy savings reductions of 46 %, 48 % and 56 %, respectively. However, compared to many of the other stores within this building portfolio the stores undertook quite extensive ECMs, such as both change of lighting, HVAC and the refrigeration system. The three stores with the lowest energy savings where store-id 1572, 1538 and 1653. These three stores actually had an increase in the energy use when comparing with the reference year. However, these results are due to the ESCO replacing oil boilers with heat pumps to heat the buildings (in 2021), hence the increase in electricity consumption. These issues are regulated in the EPC contract. The aggregated results are available to the ESCO at any time in the web application, and updates daily. Every month the ESCO generates a report from the dashboard and has a review meeting together with the customer.

Furthermore, the ESCO wanted a web dashboard that could be used to follow the energy savings in more details from day to day. To deliver on that a visualization that displays the average energy savings for all 24 h of the day, but also split into all week days, ia developed. As can be seen from the top bar plot in Fig. 4 the savings are at its peak at 10:00 with a reduction in energy usage of 25 %. In general it is best at opening hours between 07:00 and 21:00. Further, one may study the savings for each individual day, were Sunday is best (-21,3%) and Tuesday worst (-14.5 %). Also, note that the savings are very low during Tuesday evening and night, and at its worst negative (red bar) on Tuesdays at 23:00. Visual inspection gives a quick overview of the performance of the ECMs and potential for detecting errors and faults.

3.2. Cost-Benefit analysis

So far in this paper one have only considered the actual electricity consumption (kWh) savings. However, from the ESCO point of view, translating these savings into profitability is important. In the EPC project the profitability analysis have been ad hoc and based on manually collecting the relevant data, costs, electricity prices and grid



Store-Id	Reference-date	ECM-date	CV-RMSE	kWh usage ECM	Estimated kWh without ECM	Estimated savings	Savings (%)	Savings target (%)	Savings last 5 weeks (kWh)	Savings last 72 hours	Savings last 5 weeks (%)	Result savings target (%)
All	All	All	All	All	All	All	All	All	All	All	All	All
1831	2019-12-30 til 2020-12-31	2021-01-01 til 2021-11-02	0.237	374 871	682 314	-307 444	-45,1	-10			-1,6	
1832	2018-01-01 til 2018-12-31	2021-01-01 til 2021-11-02	0.09	307 066	432 632	-125 565	-29,0	-10			-22,8	
1834	2019-12-30 til 2020-12-31	2021-01-01 til 2021-11-02	0.109	464 897	845 610	-380 714	-45,0	-10			-24,1	
				1 146 834	1 960 556	-813 723	-41,5					

Fig. 3. Energy savings as presented in the web application.

rent, and then performing the analysis in a spreadsheet. However, since the underlying data mostly is available through web API's and is continuously updated it is a natural next step to develop a profitability dashboard as an extension of the web application. Thus, in this section the available data is presented, prices and rent, and how these can be integrated with the kWh savings (as reported in the previous section). And at last a benefit cost analysis that should be relatively straightforward to implement in the *ShinyRBase* web application is offered.

3.3. Electricity prices and grid rent

Electricity prices in the end-user market in Norway comprise of physical power and grid rent for transmission of the electricity by the local grid company. Nord Pool spot is the marketplace for physical electricity contracts, and is the place where the Norwegian electricity spot prices are set. The electricity prices used in this paper were downloaded from Nord Pools website as a spreadsheet.[2] However, they do have available an API that can be used to automate this step within an application.[3] Furthermore, the second price component of electricity consumption is grid rent. The food retail stores in this paper belongs to different grid owners, and to simplify-one of the largest grid owners in the area, *Vevig AS* is chosen, as a basis for calculating grid rent for all of the relevant stores. A summary of Vevig's tariffs is presented in Table 2 below. For further details please see: https://vevig.no/nettleie-og-vilkår/nettleie-næring.

The effect prices are weighted based on months. For instance, in January, February, November and December these weight equals 1. Further, between March and October the weights vary from 0.9 to 0.6. Hence, given a store with a maximum load of 210 kW in March the effect price is $210 \times 0.9 \times 55.9 = 10\,564$. Hence, there is a "penalty" for larger loads in the colder months. The end user's total electricity bills also consists of a fee earmarked for the energy fund Enova SF's (owned by the Ministry of Climate and Environment) work to reduce greenhouse gas emission and to strengthen security of supply. There is also a variable fee for electricity certificates. This fee depends on the developments in the electricity certificate market. At last there is a consumption tax on electricity.

It is not easy to automate the collection of grid rent data. There are more than 100 grid owners in Norway and many of them have different grid rent pricing strategies. Nonetheless, the pricing scheme often follow the same pattern as presented for Vevig. Hence, summer prices versus winter prices, and typically a penalty for larger loads, in particular in the winter. Probably the best solution would be to develop a user interface in the web application where the grid rent is based on user input.

3.4. Benefit cost analysis

In this section a closer look at the benefit cost analysis is presented. Unfortunately, due to confidentiality issues we are not allowed to share the details behind the actual cost elements. However, it is still found useful to propose a general method on how to approach this EPC project, in particular as the customer pays a monthly fee that includes the guaranteed energy savings, but also service and maintenance over a 10 year period. The results presented in this section is for the period January 1, 2021 - November 30, 2021, thus an 11 month period.

Usually when investigating the profitability of a project one can look at the Present Value (PV). This is defined as:

$$PV = X_0 + \sum_{t=1}^{T} \frac{X_t}{(1+r)^t}$$

where X_0 is the initial cost of the project, X_t is the net cash flow generated by the project for $t = 1, \dots, T$ periods, and r is the discount rate. However in this case, there is no initial cost like an investment in new equipment. This is done by the ESCO, and the *customer* pays a monthly lump sum for all 40 stores for the 10 years of the energy performance contract.

This "subscription" to energy savings and new improved equipment through the energy contract makes the standard way of looking at cost/benefit measures obsolete. The monthly lump sum cost paid by the *customer* to the ESCO can be considered an annuity over the 10 year period. It is like leasing a car. The monthly payment covers the ESCO's investments in the contract in addition to their profits. The benefits for the *customer*, since there is no initial investments, are "avoided" costs. These benefits for the *customer* are taken into account by looking at the investments costs by the ESCO. The *customer* gets benefits from reduced electricity costs and avoided administrative costs for equipment maintenance.

In terms of the actual implemented ECMs, all the stores changed to more efficient LED lightning, nine stores got new refrigeration systems, five stores got new heating, ventilation and air conditioning (HVAC). In addition the *customer* avoids insurance costs on the refrigeration system.

In order to calculate the benefits the present value (P) of an annuity based on the investments (PV) by the ESCO is calculated as:

$$P = PV \times \frac{\frac{r}{k}}{1 - \left(1 + \frac{r}{k}\right)^{-nk}}$$

where r is the annual interest rate, k is the number of compounds per year (12 months), and n is the number of years.

As an example, if you need to invest $space(100000)$ (PV) in new lights in a store today, the present value (P) of an annuity at monthly installments (k) for 10 years (n) at 2.5 % interest (r) is 942.7 per month.

The project has run for 11 months. In Fig. 5 (panel **A)** the monthly benefit cost ratio (BCR) is calculated. The benefits consists of two parts. The value of savings in energy costs. This is a variable part. And the fixed benefits from "avoided" costs as discussed above.

The average BCR is 1.48. Further, there is an increase in the ratio the last four months. The average percentage savings in kWh is some 13.52 % (panel **B**). This indicates that the variable component in the BCR is the price of electricity. Also, panel **C** shows that there is an increase in the average price of electricity. In the same panel (**C**) one also see that the variable portion of the BCR is above one in September.

The finding that electricity costs is contributing a relative high share in the benefit cost calculations is interesting, even when the energy savings in this period is between 11 and 16 %. One might expect higher electricity prices in the future, making the contributions of energy savings relatively large in the cost benefit perspective. Note also that before the building owner singed up for the EPC project they handled the

[2] https://www.nordpoolgroup.com/4ab28c/globalassets/marketdata-excel-files/elspot-prices_2021_hourly_nok.xls.

[3] https://www.nordpoolgroup.com/trading/api/.

Table 1
Energy savings as of November 30 2021.

Store-ID	Ref. year	CV-RMSE	Tot. kWh	Tot. prog. kWh	Savings (kWh)	Savings (%)
1522	2018	0,178	413 439	934 539	−521 100	−55,760
1249	2018	0,086	262 384	508 378	−245 994	−48,388
1520	2018	0,078	505 378	940 101	−434 723	−46,242
1546	2018	0,088	100 448	183 416	−82 968	−45,235
1560	2019	0,131	388 999	579 186	−190 187	–32,837
1555	2018	0,139	259 624	378 003	−118 379	−31,317
1551	2018	0,114	652 130	901 184	−249 054	−27,636
1524	2018	0,09	345 700	476 985	−131 284	−27,524
1573	2019	0,124	1 429 274	1 967 543	−538 269	−27,357
1562	2018	0,136	861 749	1 172 593	−310 844	−26,509
1563	2019	0,108	361 473	467 376	−105 903	–22,659
1526	2018	0,086	350 440	452 670	−102 230	–22,584
1557	2018	0,077	285 757	369 029	−83 272	–22,565
1531	2019	0,073	184 752	229 118	−44 366	−19,364
1556	2018	0,096	222 500	275 873	−53 373	−19,347
1566	2018	0,106	561 216	693 902	−132 686	−19,122
1681	2018	0,138	182 268	223 251	−40 983	−18,357
1569	2020	0,213	978 826	1 192 884	−214 057	−17,945
1548	2018	0,122	109 926	133 574	–23 648	−17,704
1536	2018	0,119	91 541	110 778	−19 237	−17,366
1540	2018	0,101	204 746	241 196	−36 450	−15,112
1528	2018	0,125	346 505	398 251	−51 746	−12,993
1731	2018	0,143	1 324 183	1 512 011	−187 828	−12,422
1564	2018	0,087	426 011	484 579	−58 569	−12,087
1552	2018	0,149	549 568	623 214	−73 646	−11,817
1554	2018	0,129	255 964	289 778	–33 814	−11,669
1533	2019	0,171	877 886	988 021	−110 135	−11,147
1529	2018	0,107	132 523	147 924	−15 400	−10,411
1542	2018	0,121	80 213	88 055	−7 842	−8,905
1549	2018	0,093	104 823	112 850	−8 027	−7,113
1545	2018	0,122	217 618	233 760	−16 143	−6,906
1547	2018	0,136	213 985	223 827	−9 842	−4,397
1553	2019	0,118	198 001	206 148	−8 147	−3,952
1640	2019	0,073	718 989	744 872	−25 883	−3,475
1558	2018	0,102	259 053	258 983	70	0,027
1532	2018	0,127	100 878	99 774	1 104	1,106
1527	2018	0,1	2 594 299	2 547 885	46 414	1,822

Table 1 (*continued*)

Store-ID	Ref. year	CV-RMSE	Tot. kWh	Tot. prog. kWh	Savings (kWh)	Savings (%)
1653	2018	0,11	129 735	127 324	2 411	1,893
1538	2018	0,095	157 216	152 955	4 261	2,786
1572	2019	0,189	580 417	545 210	35 208	6,458

insurance, service and maintenance internally. This was often a time involving process, and for many food retailers taking away the burden of handling this is also important to recognize - an element difficult to integrate into the benefit cost analysis.

All the analysis in this section was conducted using the R library lifecontingencies [33], and as such the library is easy to fully integrate into the Shiny web application. Also, the library has many other options for financial analysis that may be incorporated.

4. Discussion

Developing the *ShinyRBase* framework and application outlined in this paper has been a two year research project in close cooperation with the ESCO. Weekly status meetings were done throughout the process. The application was in fully operation in January 2021. Hence energy savings for the 40 food retail stores has been monitored on a day to day basis since then. Note that the savings are estimated in the context of measuring energy use at the whole facility. The interaction when setting up the baseline model has resulted in knowledge sharing for both sides. First and foremost, the main objective from the ESCO points of view was to have a web application that could be used to closely follow the energy savings, and to make sure that they complied with the guaranteed savings as agreed upon in the EPC contract. The ESCO did not share information about how they approached the energy audit and the calculations of the potential energy savings within the buildings under study. Often, in the audit stage of the retrofitting project simulations of the energy savings could be conducted through software such as EnergyPlus.[4] It would be useful to have access to such simulations as a comparison between the simulations and the predicted savings could be used to adjust the simulations, and to enhance the understanding of the effect of the retrofitting. Also, the provided energy data was only for the whole building, and access to sub-meters (e.g., ventilation, refrigeration, lighting) could further improve the energy saving analysis.

Furthermore, during 2021 the application was used to discard training data due to high CV-RMSE (visual inspection). In this process the application detected: lights that were on during the night, ventilation that was running in day-mode during night, in-store promotions in the training data, AMI meters that stopped working or had an technical error that resulted in to much reported electricity, and errors in the Elhub data repository that gave zero electricity reported, but also some very high levels of energy consumption (large hourly peaks). Some of these errors would have been possible to detect with a standard energy monitoring system (EMS) and/or through more standard reports from a energy analyst. However, the continuously real-time aspect that the web application enabled has been a great advantage to facilitate such findings. At the same time, building owners may benefit from other automated approaches such as occupant-building interaction via smart zoning of thermostatic loads or demand management via distributed control. Still, extending the use of energy savings baseline models such as TVB (within a reactive programming paradigm) may prove additional information that can be used to detect potential errors in the technical infrastructure. For instance, given that a model was trained on data from, for example, the refrigeration system within a time span when the

4 https://energyplus.net.

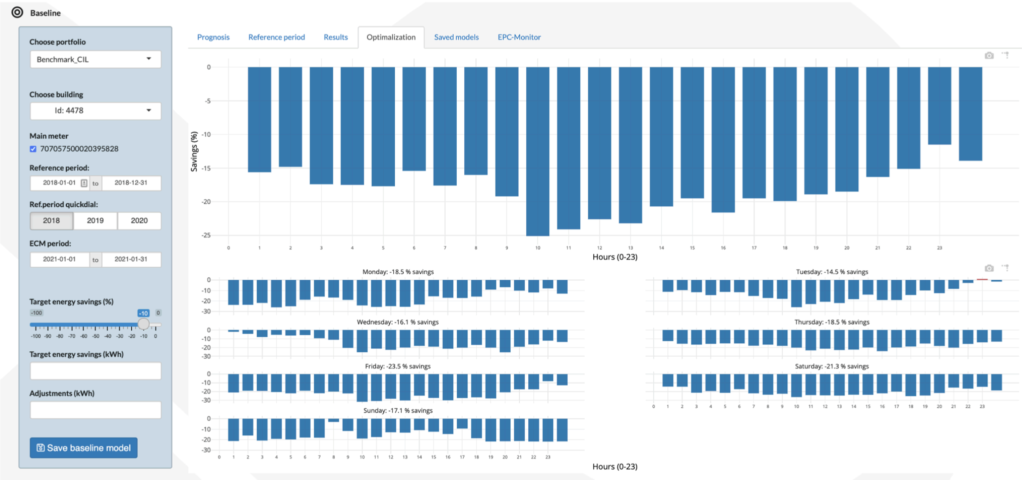


Fig. 4. Optimization - following the energy savings over days and hours.

Table 2
Vevig grid rent tariffs.

Description	Price
Fixed price (NOK, monthly)	2048
Effect price (NOK)	55.9
Winter (Øre)	0.063
Summer (Øre)	0.042
Consumption tax (Øre)	0.1669
Enova SF energy fund (NOK, yearly	800

system was running under optimal settings, the prediction can be set into production and the running difference between the actual energy consumption may be an indication of some non-optimal setting in the system.

Working together with the different stakeholders, the ESCO and the building owners, has surely increased the overall knowledge about the drivers of energy savings within the EPC project. These synergy effects ended up in a more usable web application, and as a tool for error detection. The app in itself has integrated the different stakeholders in ways that otherwise would have been difficult. Hence, the reactive framework streamlines the M&V process and deliver significant value. These findings echo the study by [34]. Furthermore, an important feature of M&V 2.0, as [12] also notes, in many energy efficiency projects there is a time lag between implementation of the ECMs and the evaluation of the savings. This lag hinders on-going changes in the ECMs that may further reduce the energy savings, e.g. optimize the control units in a HVAC. For ESCOs that has energy saving contracts the opportunity to identify and correct these failures may increase payments.

The *ShinyRBase* application was used with the following workflow. The user choose a baseline model based on a reference year. Then the model was evaluated using the CV-RMSE and by visual inspection of the actual versus predicted energy consumption on an hourly level. When the user was happy about the quality of the model they saved the model. The model was then automatically scored on a daily basis when new energy and temperature data was stored in the database. This was repeated for each of the 40 stores. The stakeholders used a web dashboard where they could monitor energy savings, both aggregated, the last 5 weeks and the last 72 h. This process was not flawless. There were several models whose predictions made little sense. However, this was either when the CV-RMSE was larger than 25 %, or when predictions was made using temperatures in the ECM period that was not present in the training data. Today, the application handle this with some basic rules that rolls back the predictions made when the feature space is not "fully covered." The roll back is very basic and just takes the last predictions (1 h back). But so far this has been a successful workaround.

4.1. Alternative baseline models

What about the TVB model and the potential need for more advanced models in operation? In this research this is an issue approached very carefully. Current research from field experience show that interpretability of models may keep the clients from accepting a black-box model (ex. artificial neural network) [35]. Furthermore, throughout this project it has been important to balance the scientific perspective with the stakeholders practical perspective. For example, it has not been easy to argue for more advanced models as the end-user already had concerns about the TVB model. Typically, the users wanted to use available development resources to enhance the visualization and layout in the application. Nonetheless, the modeling approach should be further developed taking into consideration more recent research findings. For instance, as [36] points out, a one-size-fits-all model is not realistic to be reliable across different building types. Hence, the web application should make it easier to compare models of different complexity because the user instantly gets a visualizations of the black-box model, for example a time-series plot with the actual values versus the predicted values. The visuals is a potential solution to break down this barrier. For instance, particle swarm optimization, similar to what was implemented in [14] could be integrated into the reactive framework through R libraries such as pso or psoptim [37,38].

The methods that are currently reviewed and tested is based on the main findings from the ASHRAE 2019 Kaggle competition "Great Energy Predictor III. How much energy will a building consume?"[5] This 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 are presented in [39]. The top 5 solutions were reproduced by [39] and the

[5] https://www.kaggle.com/c/ashrae-energy-prediction.

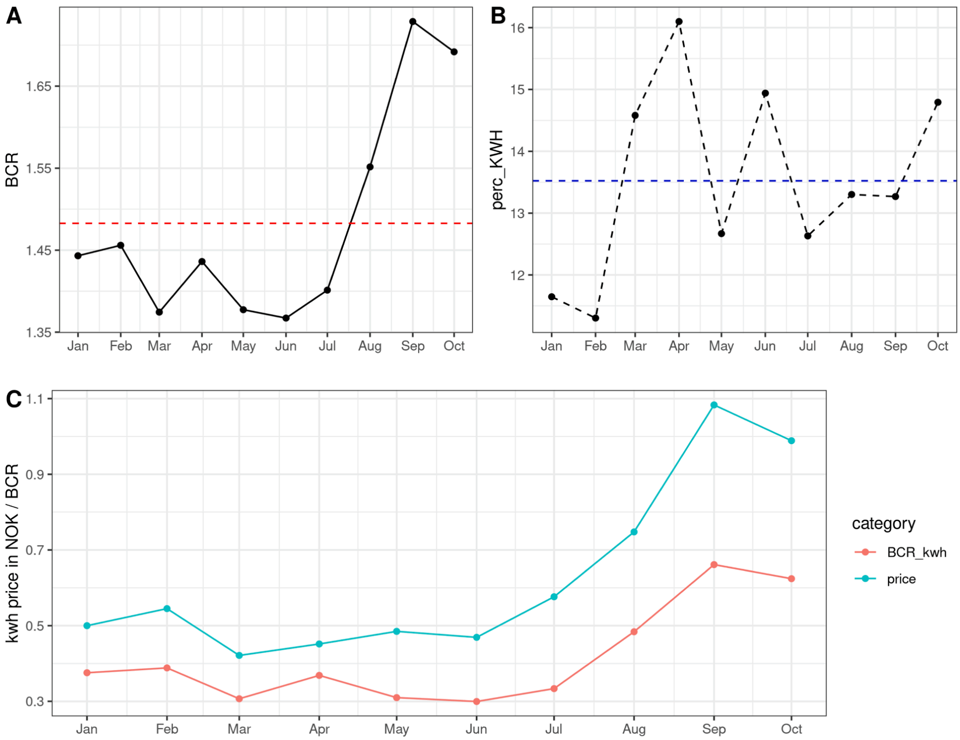


Fig. 5. BCR and energy savings in the first 11 months of the EPC project.

accompanying code can be found on github.[6] The winning solutions are presented in Table 3. As can be seen 4 out of 5 used multiple methods and post-processing of data with ensembling and weighting. All the winning solutions used Light GBM, and three of the winning teams used Catboost, and two used XGBoost.

The methods used to compete in the Kaggle competitions are quite technical. However, the winning solutions where all coded in Python [40], which is also open-source, and the code for the winning solutions can found in the above mentioned github repository for easy reproduction. The teams used pre-canned solutions (modules) to train the models, and replication in other settings should therefore be relatively straightforward. For example, all models are also available in R; lightgbm [41], xgboost [42], catboost (not part of CRAN yet, but can be downloaded as a development version from the catboost.ai website[7]), MLP (multi-layer perceptron) [43]. Most of the ensambling and weighting can be handled in the stacks package [44]. In a bi-lingual R and Python team, or if a method is only available in Python it is possible to use the R package reticulate [45], which provides a comprehensive set of tools for interoperability between Python and R. For example you can call Python from within R, translate between R and Python objects and there are flexible bindings to different versions of Python, both virtual and Conda environments.

As previously mentioned it is beneficial to have the coefficients from the models saved into a database for in-database continuously scoring

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 weather 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

[6] https://github.com/buds-lab/ashrae-great-energy-predictor-3-solution-analysis.
[7] https://catboost.ai/en/docs/installation/r-installation-binary-installation.

and monitoring of the ECMs. This is in the application handled by the package tidypredict, where xgboost, random forest and tree models are already a part of the tidypredict library. However, catboost, light-GBM and MLP is presently not possible to translate into SQL. Still, running predictions inside databases based on this is relatively straightforward to implement. While future work will carefully review these methods potential to improve the baseline models, the question raised by the ESCO several times along the project, "will more advanced modeling approaches enable the application to deliver more reliable results?" must also be recognized. To answer this there is a need to balance predictive accuracy versus the value of a slightly better model. For instance, [29] points out that because of cost and constraints, the stakeholders are less prone to embrace innovations from a modeling point of view. Often they chose a very simplified model to estimate energy savings (eg. just one temperature variable). This is important to take into consideration. For example, the ESCO in this study was used to using energy - temperature curves from weekly aggregated data with a regression model based on only the average weekly temperature as an independent variable. In this study, the web application made it much easier for the ESCO's transition from this model to an hourly based TVB model. To illustrate, they were given the chance to play around with the TVB model and to compare the overall results with the ET-curves (which gave very similar results). This comparison gave the user confidence as well as they could also see the extra benefits of looking at the results from an hourly perspective.

4.2. Baseline models in web applications - advantages and disadvantages

The advantages and disadvantages to deliver analytics using a web application instead of as a static report from an energy analyst can be divided into two phases; the development phase and the phase when the application is in production. During development one find that:

1. The stakeholder is closely involved in deciding what information will be presented in the application, and is able to try early versions of the app interactively. This leads to a strong ownership for the end product. However, this phase requires more involvement and is time consuming.
2. Different parties have different skill-sets, and working together during the development process means that those skill-sets are reflected in the application. These are collaborative benefits that may be challenging to achieve otherwise. Because the application is real-time and dynamic there is an instant feedback from the users that may be difficult to capture in a workflow where the analysis is delivered as a static report. This interactivity was found to "trigger" curiosity and need for more information. It is easier for the user to "play" with the application than to order a new report with more information.

Several advantages were found during the phase where the application is finished and set into production:

1. The results are available at the users convenience.
2. Increased efficiency for the analyst and the stakeholder. Less repetitive work for the analyst and no user dependence on the analyst to deliver.
3. The models are documented in the database for easy reproduction and daily scoring.
4. Because it is possible to closely follow the ECM's actual versus predicted consumption on an hourly level in real-time, the application is not only used to monitor the energy, but also to optimize parts of the technical system.
5. Several errors in the technical system was quickly detected.

Furthermore, the reactive framework as offered by the Shiny application and adjoining R libraries enabled fast prototyping of different solutions, web dashboards and ways to visualize and report on the TVB model. Compared to the imperative programming scheme with static reports and a strict dependence on the energy analyst the proposed framework has proven valuable.

5. Conclusions

This paper demonstrates the development of a web application, *ShinyRBase*, using a reactive framework to document and continuously monitor and benchmark energy savings for 40 food retail stores in Norway. Using open source tools, R, Shiny and adjoining libraries, this process was relatively straightforward, compared to the more standard way of delivering energy savings report. There is no need to know HTML, CSS or JavaScript to do this. The reactive framework within the Shiny library and the automated way of developing a user interface handled those aspects. The end-user was trained to make them self-sufficient in terms of setting up baseline models for the different buildings and to continuously monitor the energy savings. The baseline models was based on a well specified linear regression, the Tao Vanilla Benchmarking model.

Complex methodologies was instantly used by the end user without the need of advanced computation skills. The development and the use of the application promoted collaboration between practitioners (the ESCO and the customers) and the researcher/analyst. This collaboration resulted in an app that was fit for purpose because of the advise and the on-going interactive use from the collaborators. The advantages was twofold. First, during development the stakeholders took part of the process, which resulted in increased ownership and engagement. The different participating parties had different skill-sets, and working together during development those skill-sets ended up in a final application that was more relevant. Second, after the application was in production there were several other advantages compare to using a standard report to follow the energy savings. The users could look into the results at their own convenience, and always had fresh and current updates that was easy to monitor. Thus, a more efficient workflow for both the energy analyst (less repetitive work) and the end users (self-sufficient). The parameter estimates from the linear regression models for the different stores was saved in a table in the database (as a SQL query). This worked both as a documentation of the models, and as a useful way of scoring and updating the results every day. Hence, having the models documented in a database gave both easy and reliable reproduction of the models. Finally, since the results was always current and it was easy to closely monitor the savings, it was also easy to detect when the savings trended negatively. This enabled the users to quickly detect several errors in the technical system, such as ventilation in day-mode during night and lights that was not turned off. Some caveats should be mentioned. Even though the Shiny app simplifies the process of setting up a reactive framework there are still some added complexity. This way of working is more involved and time-consuming during the development phase, and running the app in production mode requires knowledge about setting up a server environment.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

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. IPCC Secretariat: WMO; 2021.
[2] Sánchez-Garcia D, Rubio-Bellido C, del Río JJM, Pérez-Fargallo A. Towards the quantification of energy demand and consumption through the adaptive comfort approach in mixed mode office buildings considering climate change. Energy Build 2019;187:173–85.
[3] Hawken P, editor. Buildings and cities: retrofitting. In: Drawdown: The most comprehensive plan ever proposed to reverse global warming. Penguin; 2017. p. 102–3.
[4] Statistics norway building stock. https://www.ssb.no/en/bygg-bolig-og-eiendom/bygg-og-anlegg/statistikk/bygningsmassen.
[5] United states building energy efficiency retrofits: Market sizing and financing models. Rockefeller Foundation; 2012.
[6] Wages NA, Petroni GR. A web tool for designing and conducting phase I trials using the continual reassessment method. BMC Cancer 2018 Feb;18(1):133.
[7] Klein T, Samourkasidis A, Athanasiadis IN, Bellocchi G, Calanca P. webXTREME: R-based web tool for calculating agroclimatic indices of extreme events. Comput Electron Agric 2017 Apr;136:111–6.
[8] Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, et al. A manifesto for reproducible science. Nat Hum Behav 2017;1(1).
[9] Lobaccaro G, Carlucci S, Löfström E. A review of systems and technologies for smart homes and smart grids. Energies 2016 May;9(5):348.
[10] Google powermeter [Internet]. Google; 2016. Available from: https://developers.google.com/powermeter/.
[11] EmonCMS. Open-source energy visualisations [Internet]. EmonCMS; 2021. Available from: https://emoncms.org/.
[12] Franconi E, Gee M, Goldberg M, Granderson J, Guiterman T, Li M, et al. The status and promise of advanced M&V: An overview of “M&V 2.0” methods, tools. and applications 2017.
[13] Fagiani M, Severini M, Valenti M, Ferracuti F, Ciabattoni L, Squartini S. rEMpy: A comprehensive software framework for residential energy management. Energy Build 2018 Jul;171:131–43.
[14] Ke M-T, Yeh C-H, Su C-J. Cloud computing platform for real-time measurement and verification of energy performance. Appl Energy 2017 Feb;188:497–507.
[15] R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/.
[16] Chang W, Cheng J, Allaire J, Sievert C, Schloerke B, Xie Y, et al. Shiny: Web application framework for r [Internet]. 2021. Available from: https://CRAN.R-project.org/package=shiny.
[17] Li J, Cui B, Dai Y, Bai L, Huang J. BioInstaller: A comprehensive R package to construct interactive and reproducible biological data analysis applications based on the R platform. PeerJ 2018 Oct;6:e5853.
[18] Wickham H. ggplot2: Elegant graphics for data analysis [Internet]. New York: Springer-Verlag; 2016. Available from: https://ggplot2.tidyverse.org.
[19] Vanderkam D, Allaire J, Owen J, Gromer D, Thieurmel B. Dygraphs: Interface to 'dygraphs' interactive time series charting library [Internet]. 2018. Available from: https://CRAN.R-project.org/package=dygraphs.
[20] Kuhn M. Tidypredict: Run predictions inside the database [Internet]. 2020. Available from: https://CRAN.R-project.org/package=tidypredict.
[21] Kasprzak P, Mitchell L, Kravchuk O, Timmins A. Six years of shiny in research - collaborative development of web tools in R. The R Journal 2020 Dec;12(2):20–42.
[23] 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.
[24] Hong T, Pinson P, Fan S. Global energy forecasting competition 2012. Int J Forecast 2014 Apr;30(2):357–63.
[25] EVO Efficiency Valuation Organization. International performance measurement and verification protocol. 2016.
[26] 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. https://www.eceee.org/library/conference_proceedings/eceee_Summer_Studies/2019/4-monitoring-and-evaluation-for-greater-impact/quantification-of-energy-savings-from-energy-conservation-measures-in-buildings-using-machine-learning/.
[27] Severinsen A, Myrland Ø. Statistical learning to estimate energy savings from retrofitting in the Norwegian food retail market. Renew Sustain Energy Rev 2022;167:112691.
[28] The PostgreSQL Global Development Group. PostgreSQL [Internet]. 2021. Available from: https://www.postgresql.org.
[29] Agenis-Nevers M, Wang Y, Dugachard M, Salvazet R, Becker G, Chenu D. Measurement and verification for multiple buildings: An innovative baseline model selection framework applied to real energy performance contracts. Energy Build 2021;249:111183.
[30] Grolemund G. Shiny - how to understand reactivity in r [Internet]. 2016. Available from: https://shiny.rstudio.com/articles/understanding-reactivity.html.
[31] Conway J, Eddelbuettel D, Nishiyama T, Prayaga SK, Tiffin N. RPostgreSQL: R interface to the 'PostgreSQL' database system [Internet]. 2017. Available from: https://CRAN.R-project.org/package=RPostgreSQL.
[32] Wickham H, François R, Henry L, Müller K. Dplyr: A grammar of data manipulation [Internet]. 2020. Available from: https://CRAN.R-project.org/package=dplyr.
[33] Spedicato GA. The lifecontingencies package: Performing financial and actuarial mathematics calculations in R. Available from Journal of Statistical Software [Internet] 2013;55(10):1–36. https://www.jstatsoft.org/v55/i10/.
[34] Granderson J, Price PN, Jump D, Addy N, Sohn MD. Automated measurement and verification: Performance of public domain whole-building electric baseline models. Appl Energy 2015 Apr;144:106–13.
[35] Molnar C. Interpretable machine learning a guide for making black box models explainable. https://leanpub.com; 2021.
[36] Cui C, Wu T, Hu M, Weir JD, Li X. Short-term building energy model recommendation system: A meta-learning approach. Appl Energy 2016 Jun;172:251–63.
[37] Bendtsen CP. Available from: Particle swarm optimization [Internet] 2012. https://CRAN.R-project.org/package=pso.
[38] Psoptim CK. Available from: Particle swarm optimization [Internet] 2016. https://CRAN.R-project.org/package=psoptim.
[39] Miller C, Arjunan P, Kathirgamanathan A, Fu C, Roth J, Park JY, et al. The ASHRAE Great Energy Predictor III competition: Overview and results. Science and Technology for the Built Environment 2020;26(10):1427–47.
[40] Python Core Team. Python: A dynamic, open source programming language [Internet]. Python Software Foundation; 2019. Available from: https://www.python.org/.
[41] Ke G, Soukhavong D, Lamb J, Meng Q, Finley T, Wang T, et al. Lightgbm: Light gradient boosting machine [Internet]. 2020. Available from: https://CRAN.R-project.org/package=lightgbm.
[42] Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, et al. Xgboost. Available from: Extreme gradient boosting [Internet] 2021. https://CRAN.R-project.org/package=xgboost.
[43] Bergmeir C, Benítez JM. Neural networks in R using the stuttgart neural network simulator: RSNNS. Available from Journal of Statistical Software [Internet] 2012; 46(7):1–26. https://www.jstatsoft.org/v46/i07/.
[44] Couch S, Kuhn M. Stacks: Tidy model stacking [Internet]. 2021. Available from: https://CRAN.R-project.org/package=stacks.
[45] Ushey K, Allaire J, Tang Y. Reticulate: Interface to 'python' [Internet]. 2021. Available from: https://CRAN.R-project.org/package=reticulate.
[46] Kontokosta CE. Modeling the Energy Retrofit Decision in Commercial Office Buildings. Energy Build 2016;131(November):1–20. https://doi.org/10.1016/j.enbuild.2016.08.062.

6.4 Paper 4: A 3-step framework to benchmark potential and actual energy savings in retrofitting projects

A 3-step framework to benchmark potential and actual energy savings in retrofitting projects

2022-07-27

Abstract

This paper demonstrates a 3-step benchmarking framework to document the effect of energy savings and efficiency from retrofitting Norwegian food retail stores. This is accomplished in collaboration with an energy service company (ESCO). During autumn 2020 the ESCO undertook a retrofitting project for 34 food retail stores. The proposed framework follows the same order as a retrofitting project typically is conducted: the audit, implementation, and measurement and verification. In the first step (during the audit) a data envelopment analysis (DEA) is used to establish the energy efficiency before any of the energy conservation measures (ECMs) are installed. At the same time a novel energy saving baseline model is developed, the Broken line model (BL). Baseline models are normally introduced after implementing the ECMs in the measurement and verification phase. However, training baseline models during the audit phase may reveal important information about the condition of the technical infrastructure in terms of the demand for cooling and heating. In the second step the Tao vanilla benchmarking model (TVB) is used to estimate the energy savings on an hourly level. The results are used to adjust the ECMs during the implementation phase. This optimize the energy saving potential and reveal plausible non-routine events. In the third and last step during the measurement and verification phase the DEA is redone with data one year after the implemented ECMs. In this step the energy savings are esimated using both the BL and the TVB model to assure the energy saving results. This enables a perspective where it is possible to investigate the efficiency change together with the energy savings. This last step use the weights from the DEA multiplier model to analyze the change in the input variables (demand for heating and cooling) across the retrofitting project. The results from the proposed framework show that DEA, together with baseline models, can be a valuable tool for the ESCO to monitor and advance the different stages of a retrofitting project, and eventually give new insights into how to prioritize and adjust ongoing ECMs.

Keywords— Energy efficiency, savings evaluation, data envelopment analysis, building energy retrofitting, Measurement and verification, Data driven models, Broken line models, Tao Vanilla Benchmark model

Abbreviations— ASHRAE;American Society of Heating; Refrigerating and Air-Conditioning Engineers, DEA; Data Envelopment Analysis, 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, CV-RMSE; coefficient of variation root mean square error

1 Introduction

The latest IPCC Sixth Assessment Report clearly states that it is critical to initiate measures to be able to reduce human-induced climate change, and that human activities is causing pervasive disruption in nature that affects billions of people. Hence, it is of major importance to quickly make extensive cuts in greenhouse gas emissions (GHG).[1]

Globally, buildings consume some 40% of all produced energy and as such cause large scale GHG emissions [1]. Moreover, a substantial amount of this energy consumption may have been wasted due to faults in the design and the construction of the buildings, and in particular building operation [2,3]. Among the different building categories food retail stores are one of the largest energy users. In Norway the average energy consumption for food retail stores are 540 kWh/m^2, and is by far the category with the largest energy use intensity [4]. As such, to increase energy efficiency renovation projects has an important role, and food retail stores represent an attractive target to reduce energy consumption and increase energy efficiency [5]. Still, uncertainty regarding the expected savings has previously deterred new retrofitting projects [6]. Consequently, for building owners with a large portfolio of buildings it is fundamental to have reliable methods to analyze the *potential* energy savings within the existing buildings, and moreover, to document the *actual* energy savings after implementation of energy conservation measures (ECMs). One may answer questions such as: are the accomplished savings as expected given the initial efficiency? What was the main driving parameters behind the savings, and are there any identified variables that can increase the savings?

The energy performance in a building demonstrates the quality of a building in energy use [7,8]. The performance is often assessed through Energy Performance Indicators (EPI). The most prevalent for many buildings, food retail stores included, is energy use intensity (EUI), e.g. kWh/m^2. Floor area is a strong indicator of energy use. However, as [9] points out, a limitation of this method is that other energy related features are not taken into account. For instance, the weather is an important component to fully understand the energy use within a building. When it is cold outside heating increase, and on warm days the demand

[1] https://www.ipcc.ch/report/ar6/wg2/

for cooling increase. Because of this, EUI, is often complemented by comparing buildings within the the same climatic zone, hence, a normalized EUI. Still, floor area and weather are not the only variables that affect the energy consumption in a building. Depending on building category, opening hours, occupancy and technical infrastructure may be other important components to fully understand the energy use.

To study the impact of different variables on energy efficiency in retrofitting projects one approach is to use data envelopment analysis (DEA). Nonparametric DEA is a mathematical programming technique used to find an efficiency frontier which consists of the most energy efficient buildings. DEA has the advantage that it can be used to identify factors that can be used to create effective renovation strategies.

In general, several papers have conducted DEA to investigate efficiency in the energy sector. For instance, [10] research the energy efficiency in 277 Chinese cities with the use of slack-based DEA. They find that there is major disparity in energy efficiency between the cities. Moreover, [11] investigate the efficiency of rice farmers with regard to energy use in rice production in India. The findings from this paper show that 11.6% of the total input energy could be saved through better use of power tillers and improved machinery. Furthermore, [12] use DEA to study energy inputs and cucumber yield in selected greenhouses in Iran and finds that 12 out of 18 greenhouses were efficient and that 8.5% of the overall resources could be saved by improving the performance of the inefficient greenhouses.

More specifically, within the building sector a number of different benchmarking approaches have been proposed. For instance, [13] use DEA to benchmark green building attributes to achieve maximum green points with limited capital. Furthermore, [14] use DEA to evaluate the efficiency of 47 government office buildings in Taiwan. The study finds that the average energy performance was 65%, hence an average saving potential of 35%. The analysis uses standard regression analysis to temperature adjust the energy consumption before the conducted DEA. Further, in a subsequent study [15] prepares the data for DEA with the use of cluster analysis to classify buildings into different climate cluster. Across the climatic zones the study finds efficiency scores of 0.5, 0.56, and 0.56. Thus, a substantial energy savings potential. Furthermore, [16] propose a two-stage DEA energy benchmarking method. The first stage conducts a DEA which integrate a common degree-day approach to take into account temperature differences between the buildings, and the second stage use Tobit-regression for a detailed efficiency analysis. In the study they demonstrate the benchmarking methods in 189 residential buildings. In a more recent study within the food retail sector, [17] conduct DEA for 137 Norwegian retail stores. The detected inefficiency was 28% compared against an efficient rank of 32 stores.

The above benchmarking studies all reveal substantial inefficiencies and as such a promising capacity for energy savings within the building sector. Nonetheless, while it is vital to detect any inefficiencies during the audit phase of a renovation project, to fully understand how to approach the estimated inefficiencies actual implemented energy conservation measures (ECM) and estimation of the energy savings after a retrofitting project provide useful insights and a more coherent view on the expected savings.

In this paper we demonstrate a novel benchmarking framework that is designed to document energy savings and energy efficiency in retrofitting projects of food retail stores. The framework is designed to be used throughout the different phases of the renovation project: the energy audit, implementation, and at last, measurement and verification (M&V). To accomplish this, we work together with a medium sized Norwegian energy service company (ESCO) that has specialized in retrofitting within the food retail sector. The ESCO recently signed an energy performance contract (EPC) for retrofitting 34 Norwegian food retail stores. The contract includes a yearly guaranteed energy savings target, was active from 1 January 2021 and has a 10-year duration. Within this period the ESCO handles all the energy infrastructure within the stores, including service and maintenance. The proposed framework is used in the project to document both the potential energy savings as documented in the initial energy audit, to optimize implementation of the energy conservation measures (ECMs), and finally to establish the actual energy savings in the M&V phase. The framework consists of 3 stages illustrated in 1. The illustrated stages are,

1. A broken line model (BL) is trained on a reference year before the implemented ECMs. Weekly data is used. The BL model has three purposes during the audit phase. First, it is used to understand the buildings heating and cooling needs, including the changing point temperature (CPT). Second, the estimated kWh cooling and heating is passed on to the DEA analysis for efficiency analysis. The BL model is later used as a baseline model to estimate the energy savings after the implemented ECMs.

2. The Tao Vanilla benchmarking (TVB) model is trained on the same data as the BL model, however hourly data is used. The objective of the TVB model is, as with the BL model, to estimate energy savings after the implemented ECMs. The estimated accuracy of the TVB model should coincide with the BL model, and provide additional reliability for the estimated energy savings. However, in this particular step the model is used to optimize installation of the ECMs during the implementation phase. For instance, the hourly level allows us to study the energy savings in terms of the actual hours when the savings is at its highest versus lowest, changing opening hours (open versus a closed store), and in weekdays and weekends. In particular, this is important during the implementation stage for ongoing optimization of the ECMs.

3. In the final step the DEA model is redone one year after the implemented ECMs. The constant returns to scale (CRS) model is used to calculate efficiency scores before and after the implemented ECMs.

Hence, one can demonstrate the change in efficiency as a result of the ECMs. The model outputs are the size of the buildings (m^2) and opening hours; two important variables that may explain the energy consumption in food retail stores. In addition, the following three inputs are included in the DEA model: energy consumption for heating, cooling and the remainder. Furthermore, to strengthen interpretability the DEA multiplier model find how these three inputs relate to the efficiency scores. This final step allows an investigation into how the energy savings relate to the efficiency. For example, with this perspective, one finds groups of food retail stores within the building portfolio with the same energy savings, the same implemented ECMs, but with different DEA efficiency. Those results may be used to improve the ECMs to gain more efficient buildings in terms of energy consumption. The calculated multiplier model weights (ux) for heating, cooling and the remainder can be used to compare both across the food retail stores, and between different measurements (before and after the ECMs). In this way it is possible to gain a better understanding of the underlying drivers that determine the efficiency. For instance, a low weight on heating in relation to cooling and the 'rest' means that this store should investigate the causes behind the energy consumption consumed by heating. In the results section we will indicate "worst" performers and enhance the interpretation of the ux weight with energy-temperature signatures from the BL model.

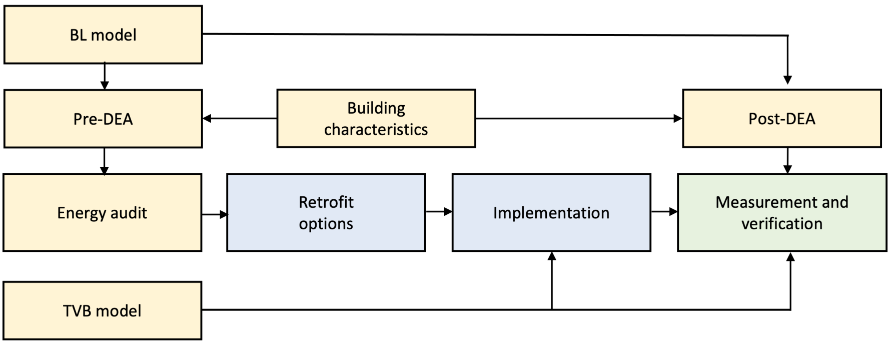


Figure 1: Framework for benchmarking energy efficiency and savings

1.1 Novelty of the paper

Previous building energy efficiency studies that use DEA primarily focus the attention to the audit phase of renovation projects. Although being able to identify the energy performance is essential to create effective retrofitting strategies, it is significant to be able to track the efficiency change throughout a retrofitting project. In the present study a DEA is conducted during the audit of the building portfolio, and than re-done one year after the ECMs were installed. Additionally, the DEA multiplier model is used to review the efficiency changes of the DEA input variables. Hence, it is possible to gain a better understanding of the underlying drivers that determine the efficiency (e.g. an inefficient cooling or heating system), and identify untapped energy savings. As far as the authors has been able to establish this is the first paper that study the change in the efficiency before and after a implemented renovation project, and that apply the multiplier model to facilitate a more comprehensive study of efficiency and energy savings.

Second, the audit phase and the benchmarking of the energy efficiency within the building portfolio prompts a list of recommended energy conversations measure (ECMs). In the current study 34 buildings are retrofitted, and the energy savings are documented using the BL and the TVB model. Typically, baseline models are developed during the M&V phase. However, output from these models is used as input to the DEA model to improve energy savings. The integrated use of DEA as a benchmarking methods, combined with the baseline models (TVB and BL) advance knowledge about the relationship between energy efficiency and energy savings. The present paper offers a collective perspective for efficiency versus savings.

The following sections first describe the ESCO and the implemented ECMs. The methods are then outlined; the broken line model, the TVB model and the CRS-DEA multiplier model. Furthermore, the results are presented. At last, a discussion of the findings, limitations, and the conclusion.

1.2 Data pipeline - electric load, weather data and building characteristics

The launch of advanced metering infrastructure (AMI) has created the opportunity to analyze energy data in near-real time and on a more granular level, often referred to as M&V 2.0. In Norway this is handled by a central repository, Elhub (elhub.no), that daily collects energy use on an hourly level (15 min interval will be available in May 2022) for commercial and household buildings using the AMI system. It is mandatory for all the Norwegian grid operators with daily update to a central repository. Furthermore, temperature data is collected 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. The size of the buildings was collected from an API set up by the store owners. These automated processing and access to data enable delivery of current and updated savings to stakeholders [18].

The collected data and the analytics conducted in this paper is presented in fig 2. The data from Elhub, met.no and the store owners API is continuously saved in a database, and further imported into `R` [19]. Finally, the DEA, TVB and BL analysis was conducted in `R` and the accompanying `Benchmarking` library

[20].

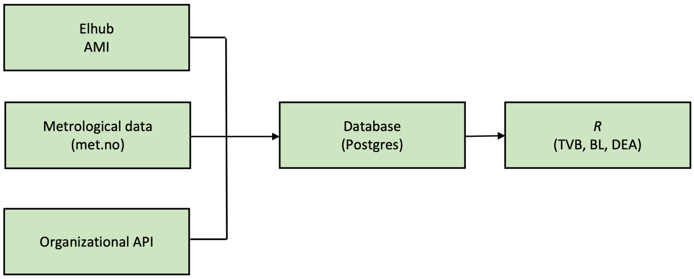


Figure 2: Data pipeline for benchmarking

2 Methods

2.1 Broken Line Models

To estimate the relationships between energy consumption and temperature for each building broken-line (BL) models will be used. For instance, when the outside temperature increases this commonly leads to an increase in a buildings energy consumption due to increased use of cooling. By the same token, the energy consumption increases due to cold temperature in the winter because of increased use of heating. The change point, changing point temperature (CPT) is the point at which there is no demand for cooling or heating.

Furthermore, traditional methods used to account for non-linear effects, such as regression splines or polynomial regression are not suitable because the CPT values are set a priori. Also, the estimated regression parameters from can not be interpreted directly [21]. When the CPT parameters must be estimated, standard likelihood-based inference is complicated due to the log-likelihood is only piecewise differentiable and the classical regularity conditions are not met [22–24].

The is approached by reducing the above issues to a linear scheme. The CPT relationship between the mean response $\mu = E\left[\Upsilon\right]$ and the variable K is modeled by including the linear predictor for

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

Where $(K_i - \psi)_+ = (K_i - \psi) \times I(K_i > \psi)$ and $I(\cdot)$ is the indicator function equal to one when the statement is true. Hence, β_1 is the left slope, β_2 is the difference-in-slopes, and ψ is the CPT value. A number of challenges has been described by [25]. Often, grid-search algorithms have been used to estimate similar models, for instance fitting several linear models and searching for the value that coincide to the model with the better fit. Still, this is not the best approach when there are numerous changing points, or a large dataset.

In this paper the open-source *R* package *segmented* is used to estimate the BL models [26]. The package estimate model (2) by repeatedly fitting the linear model

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

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 $\psi = \psi + \tilde{\gamma}/\tilde{\beta}_2$. An applied example of this approach can be seen in [27], where energy savings in 5 different Norwegian food retail stores were estimate, and the model proved reliable results.

2.2 Baseline Model - Estimating the energy savings

The TVB models has previously been validated as easy to implement and produce accurate results [27,28]. Furthermore, the models has also proven reliable when the expected energy savings target is below 10%, for instance in smaller implemented ECMs [29].

The following specification is used:

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

where Y_t is the load for hour t, β_i are the estimated coefficients from the least squares regression; M_t, W_t and H_t are month of the year, day of the week and hour of the day, respectively. Furthermore, T_t is the outside temperature for time t. In Hong's TVB model past loads and trend was included as explanatory variables. Nonetheless, in the present research the TVB model will model the energy consumption in stores based on a reference period, hence lagged variables and trend are not included as features.

The accuracy of the models is measured through the *coefficient of variation root mean square error* (CV-RMSE). The CV-RMSE is calculated in the following way,

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

where $\bar{Y}$ is the mean of the energy demand in the reference data. Furthermore, Y_i is the energy use in hour i, $\hat{Y}_i$ is the predicted energy use in hour i from the TVB model. Also, n is the sample size, and k is is the number of explanatory variables in the model.

2.3 Data Envelopment Analysis

The non-parametric DEA method first introduced by [30] measures the relative efficiency between homogeneous units by estimating a composite score for each unit under consideration. The Charnes, Cooper and Rhodes (CCR) model assumes constant returns to scale (CRS), allowing possible scaling of units in the analysis. This implies allowing units of different size to be compared. Further, [31] developed the model to account for variable returns to scale (VRS). Assuming variable return to scale, the Banker, Charnes and Cooper (BCC) model ensures that units in the analysis will be compared to other units of similar size.

The method calculates efficient utilization of resources by applying mathematical programming. One of the advantages of the method is the ability to incorporate multiple inputs and multiple outputs. DEA, unlike parametric methods, do not require an a priori functional form. The main disadvantage is its sensitivity to individual units that are not comparable, and as such a thorough investigation of possible outliers is necessary for a reliable result. Efficient units obtain a score of 1 (100 percent), while inefficient units receive a score less than one but greater than zero. The objective is to minimize input (maximize output) holding output (input) fixed.

Investigating energy efficiency, the term decision-making unit (DMU) represents the different food retail stores, and the objective is to minimize energy consumption, hence an input oriented model. Considering a set of K observations of DMUs, each DMU^k $(k \in K)$, uses m inputs $x^k = (x_1^k, ..., x_m^k) \in \mathbb{R}_+^{\gtrdot}$ to produce n outputs $y^k = (y_1^k, ..., y_n^k) \in \mathbb{R}_+^{\ltimes}$. By creating a piece-wise linear approximation DEA determines an efficient frontier or best practice frontier by these K observations and returns an efficiency estimate for each DMU.

Following [32] the input-oriented VRS multiplier model is specified for DMU^0 as follows:

$$
\begin{aligned}
&\max_{u,v,\phi} \quad vy^{\circ} + \phi \\
&\text{s.t.} \quad ux^{\circ} \leq 1 \\
&-ux^{k} + vy^{k} + \phi \leq 0, \quad k = 1, \dots, K \\
&\phi \in \Phi(\gamma).
\end{aligned}
\tag{3}
$$

where $\Phi(vrs) = \mathbb{R}$ and $\Phi(crs = 0)$. For the input-oriented CRS multiplier model there is no restriction on $\lambda^1, ..., \lambda^k$ and the problem may be formulated as

$$
\begin{aligned}
&\max_{u,v,\phi} \quad vy^{\circ} \\
&\text{s.t.} \quad ux^{\circ} \leq 1 \\
&-ux^{k} + vy^{k} \leq 0, \quad k = 1, \dots, K \\
&\phi \in \Phi(\gamma).
\end{aligned}
\tag{4}
$$

The scalar ϕ is the cost for not having constant returns to scale. The value weights ϕ, u and v selected by the DEA program put the evaluated unit in the best possible light compared to the other units. The dualization thus supports the popular view that DEA puts everyone in the best possible light [32].

Hence, the idea is to ensure that high outputs accrue from low inputs. The calculated weights (ux) for heating, cooling and 'the rest' can be used to compare both across the food retail stores, and between different measurements (before and after the ECMs). In this way it is possible to gain a better understanding of the underlying drivers that determine the efficiency. For instance, a low weight on heating in relation to cooling and the 'rest' means that this store should investigate the causes behind the energy consumption consumed by heating. In the results section we will indicate "worst" performers and enhance the interpretation of the ux weights with energy-temperature signatures from the BL model.

3 Results

This section starts with a presentation of the results from the hourly TVB model. The model was used to give detailed insights on an hourly level how the ECMs performed. Furthermore, we look at the results from the standard approach that the ESCO used to document energy savings, the BL model. At last, the efficiency scores from the CRS-DEA model together with the cooling and the heating weights are presented. This part also pulls in two aggregate energy savings from the TVB model and the BL model - necessary to validate the two models, and to be able to compare the energy savings with the efficiency scores.

3.1 TVB Baseline model

In figure 3 we present an example of how the results from the TVB hourly model can be applied. This particular model was trained for store-id '1522,' using 2018 as a reference year, and the example show predictions between December 1st 2021 and December 17th 2021. The stippled line is the actual hourly kWh energy use, and the solid line is the prognosis. The difference between these lines are the actual savings (given that the building performs as in the reference year). For instance, December 3rd at 08:00 the prognosis was 185 kWh and the actual usage 87. Hence, at this particular hour a 98 kWh saving. However, at December 5th 14:00 the actual usage was 54 kWh and the prognosis 74 kWh, hence a 20 kWh saving. The TVB model thus demonstrates that the savings can be narrowed down on a detailed level. After working together with the building owners and the ESCO to establish reference years for the 34 food retail stores, we trained the TVB model for all the buildings and estimated the savings for 2021 (up until December). The aggregated savings are presented in table 1.

Furthermore, the model reliability was tested against the CV-RMSE, which is recommended by the IPMVP, and accordingly should be below 25% for the model to be accepted [33]. Large CV-RMSE (>25%) was flagged and inspected through visualizations of the data in the reference period. Often large CV-RMSE could be led back to the food retail store being out of operation and/or some other uncommon activity, for instance in-store promotion activities. Consequently, the CV-RMSE was actively used to understand the modeling results together with all the involved stakeholders. As a results, this collaboration led to a change of reference year for many buildings, and/or a better understanding of the actual data used as training data, e.g. missing and/or unusual observations. Also, the TVB was use during implementation of the ECMs. This according to the second step in the framework. In this phase the TVB improved several of the ECMs as the ESCO was able to quickly analyse the savings and accordingly adjust the ECMs.

3.2 Broken line model

As discussed previously the standard approach for the ESCO was to use a one variable model to estimate the energy savings, using a weekly aggregate for the energy consumption and the outside temperature. To be able to compare the results from this approach with the TVB model we have estimated the BL model for

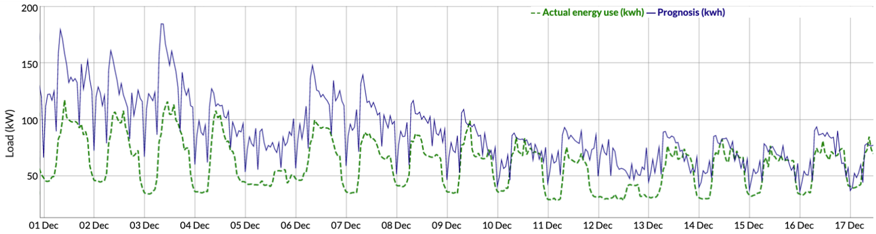


Figure 3: TVB baseline model for store-id '1522' between December 01 2020 - December 17 2021. TVB model based on data for 2018

the same 34 buildings with the same reference year. Figure 4 show 8 of these models, the 4 top performers in terms of energy savings in the left row and the bottom 4 in the right row. For instance, for store-id '1522' the y-axis shows the weekly kWh consumption, and the x-axis shows the average outside temperature. The BL-model was estimated using the reference data (the solid dots in the ET - curve). Further, we also plot the weeks up until December 2021 (the 'x'- dots = the ECM period). The distance from the different weeks in the ECM periods (the 'x') up to the ET - curve line is the energy savings for that week. The aggregated savings from the BL model are presented together with the corresponding savings from the TVB model in table 1.

3.3 Data envelopment analysis - efficiency and energy savings

3.3.1 Energy savings

The overall energy savings effect of the implemented ECMs, as presented in table 1 gives an average energy savings of -17.6%. There is relatively little practical difference between the results from the hourly TVB and the weekly BL models. For instance, the Pearson's product-moment correlation between the two models estimated energy savings is 0.997, with p-value < 0.000. Also, there does not seem to be any pattern in the difference; no over or under prediction. Still, there are some noteworthy differences, e.g. for store-id '1560' the TVB model gives an energy savings of -33.5% versus -30.9% from the BL model. The CV-RMSE for the TVB model was 14%, while the corresponding value for the BL model was 17%. A further inspection of the potential cause for these differences revealed that the first two weeks of 2018 (the reference year) had an unusual low energy consumption compared to other comparable weeks, likely due to some errors in the technical system. In figure 4 these two weeks for id '1560' is displayed in the ET - curve (circled in with stippled lines). Discussing these anomalies with the ESCO and the customer (who agreed this was not normal) led to an agreement to impute those two weeks. Likewise, for store-id '1548' there was

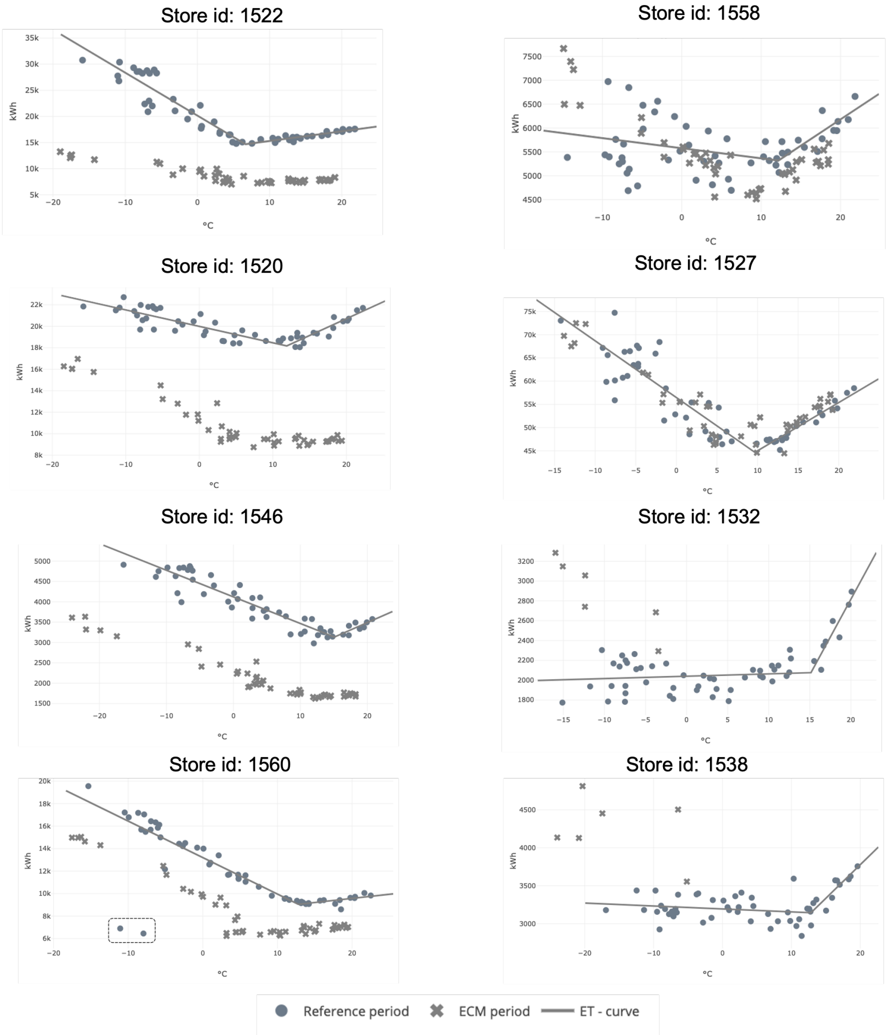


Figure 4: ET curve for the reference periode estimated with the broken line model with weeks after ECMs displayed as ’x’. Top 4 performers in terms of energy savings on the left top, bottom 4 on the right row.

a difference of 1.7%, the TVB model savings was estimated at -17.5% (CV-RMSE = 14%) and the BL at -15,8%. (CV-RMSE = 17%). Thus, the TVB model had a somewhat higher precision for both of these examples.

3.3.2 Energy efficiency

The efficiency scores from the DEA is presented in table 1. The column *'CRS Ref.'* gives the efficiency scores before the installed ECMs, and the column *'CRS 2021'* after. The average efficiency scores both before and after the installed ECMs were 74%. Given that the retrofitting project generated an average energy savings of 17,6%, it is a surprising finding that the overall efficiency was not affected. This finding will be further examined in the discussion.

The results are sorted by the percent energy savings as estimated by the TVB baseline model. We can see that store-id '1522' and '1520' has a reduction in the energy consumption of -55.8% and -46.2%, respectively. Both of these stores had implemented three different ECMs; change to more efficient LED-lights, and change of the HVAC and the refrigeration system. Furthermore, the DEA revealed that these stores had an efficiency of 43 and 41% before the implemented ECMs *(CRS Ref)*, indicating that the stores had great potential for energy reduction, and store-id 1522 had a positive change in the efficiency (49,5%) that corresponded to the energy savings (55,8%/TVB model).

In terms of energy savings several of the stores did actually have a slight increase in the energy consumption (Store-id '1558,' '1527,' '1532' and '1538'). However, two of these stores, '1532' and '1538' had the oil boiler that was used for heating replaced by a heat pump. Thus, electricity and not oil is used for all the heating. Hence, the increase in the overall electricity consumption. These issues are also regulated in the EPC contract between the ESCO and the customer. This finding is reflected in the DEA result as both of these stores were on the DEA front before the implemented ECMs and ended up with a lower efficiency score in 2021 (from 1 to 0.76 and 0.92, respectively). Nonetheless, we find that these two stores according to the DEA analysis had a low potential for energy savings given that they were already at the front. Furthermore, store '1558' had changed their lights to efficient LED, but did not show any decrease in electricity consumption and the efficiency score was approximately the same in the reference year and in 2021. It turns out that the ESCO had several technical issues with this particular store. The control system for the ventilation did not work optimal and several fixes are currently being set in operation to deal with this issue.

Table 1: Aggregate results from TVB, broken line and DEA-CRS

Store-Id	% change TVB	% change BL	CRS Ref.	CRS 2021	% change eff.	Refrigeration	HVAC
1 522	-55,8	-55,4	0,43	0,65	49,5	1	1
1 520	-46,2	-46,2	0,41	0,54	30,9	1	1
1 546	-44,9	-45,3	0,89	1,00	13,0	1	0
1 560	-33,5	-30,9	1,00	0,73	-26,9	0	0
1 555	-30,9	-30,3	1,00	1,00	0,0	1	0
1 551	-28,8	-29,5	0,51	0,44	-14,1	0	0
1 573	-27,4	-26,7	0,85	0,89	4,4	0	0
1 524	-26,7	-26,8	0,81	0,63	-22,2	1	0
1 562	-26,4	-26,5	0,65	0,81	24,7	0	0
1 563	-22,8	-22,4	0,60	1,00	65,5	0	0
1 557	-22,4	-22,2	0,51	0,57	11,2	0	0
1 556	-19,5	-19,5	0,59	0,72	21,1	0	0
1 531	-19,1	-19,2	0,75	0,67	-11,7	0	0
1 566	-19,1	-19,2	0,58	0,73	26,3	0	0
1 569	-18,2	-18,7	0,92	0,74	-19,2	0	0
1 548	-17,5	-15,8	0,89	1,00	12,4	0	0
1 536	-17,3	-16,7	1,00	1,00	0,0	0	0
1 540	-15,4	-15,4	0,69	0,67	-3,2	0	0
1 528	-13,0	-13,6	0,69	0,60	-12,1	1	0
1 731	-12,4	-12,5	1,00	1,00	0,0	0	0
1 564	-12,0	-11,7	0,65	0,70	7,0	0	0
1 552	-11,9	-11,9	0,50	0,67	33,3	0	0
1 554	-11,3	-11,2	0,66	0,64	-3,4	0	1
1 533	-10,8	-10,6	0,44	0,30	-31,9	0	0
1 529	-10,5	-11,5	0,81	0,72	-11,4	0	0
1 542	-9,1	-11,0	1,00	1,00	0,0	0	0
1 549	-7,4	-7,2	1,00	1,00	0,0	0	0
1 545	-6,9	-5,7	0,74	0,66	-10,1	0	0
1 547	-4,2	-4,9	0,72	0,77	7,9	0	0
1 640	-3,1	-1,9	0,60	0,50	-16,7	0	0
1 558	0,4	-2,4	0,66	0,60	-8,0	0	0
1 527	1,6	1,1	0,61	0,50	-19,4	0	0
1 532	2,0	2,6	1,00	0,76	-23,7	0	0
1 538	2,7	2,5	1,00	0,92	-7,5	0	0

3.3.3 UX weights

In table 2 we present the UX weights from the multiplier model before (ref.) and after (2021) the implemented ECMs. For example, the weight for heating (UX Heat ref.) was 0,0 for store-id '1522' and in the reference period that weight increased to 0,789 after the implemented ECMs (UX Heat 2021). This implies that part of the increase in the DEA-CRS efficiency scores can be attributed to a improvement in the heating system. The same finding can be seen for store-id '1546.' Note that the weights sum to 1. These findings are further summarized in figure 5.

To perform research an on actual EPC retrofitting project poses some challenges as there are bound to be events that are difficult to foresee. Some of these issues are summarized in table 3. Note for instance the results for store-id '1532' and '1538.' Both of these had their oil boilers phased out, and a heat pump installed for heating. Hence, this particular ECM actually increased the overall energy consumption of these two stores. During the ECM period there was also several examples of technical errors that substantially affected the energy consumption. For instance, store-id '1560' had an error in the ventilation control unit that prevented the switch between night and day mode. Also, store-id '1563' had to repair the steering unit for the snow smelting system. Both of these issues were detected from the TVB baseline model (the model was automatically updated every day to be able to quickly detect errors). Some of these issues are naturally regulated in the EPC project contract, but it is still much easier to discuss these issues with the customer if quickly detected.

In figure 5 we present the change in the UX weights after the ECMs for the top 4 and bottom 4 performers in terms of energy savings. For example, store-id '1522' had a 56% energy savings from the implemented ECMs (new LED lightning, refrigeration and HVAC). The store had an increase of 0.789 for the heating weights (ux.heating), hence the increase in efficiency can partly be explained by improvement in the energy consumption that was used for heating. Referring back to the ET - curves in figure 4 storeid '1522'had a much lower increase in the energy consumption in the colder months compared to the reference period (after the installed ECMs). The same comments applies for store-id '1546' where ux.heating increased with 0.709.

4 Discussion

The discussion is divided into four parts according to the steps in the proposed framework. In the first section the output from the BL model, and the DEA efficiency scores from the audit phase is presented. Second, the results from the energy savings as estimated in the BL and the TVB model are presented. Third,

Table 2: UX weights for cooling, heating and rest from the DEA Multiplier model before (ref) and after implemented ECMs (2021)

Store-Id	UX Cool ref.	UX Cool 2021	UX Heat ref.	UX Heat 2021	UX Rest ref.	UX Rest 2021
1 522	0,106	0,083	0,000	0,789	0,89	0,129
1 520	0,175	0,000	0,017	0,126	0,81	0,874
1 546	0,030	0,021	0,000	0,709	0,97	0,270
1 560	0,035	0,000	0,000	0,000	0,96	1,000
1 555	0,306	0,024	0,000	0,881	0,69	0,095
1 551	0,746	0,000	0,254	0,000	0,00	1,000
1 573	0,237	0,000	0,000	0,071	0,76	0,929
1 524	0,637	0,000	0,363	0,158	0,00	0,842
1 562	0,866	0,527	0,134	0,473	0,00	0,000
1 563	0,226	0,184	0,000	0,485	0,77	0,331
1 557	0,040	0,074	0,000	0,385	0,96	0,541
1 556	0,125	0,569	0,006	0,264	0,87	0,167
1 531	0,217	0,000	0,001	0,044	0,78	0,956
1 566	0,168	0,597	0,014	0,403	0,82	0,000
1 569	0,285	0,000	0,001	0,000	0,71	1,000
1 548	0,124	0,094	0,000	0,597	0,88	0,309
1 536	0,010	0,352	0,056	0,080	0,94	0,568
1 540	0,475	0,282	0,525	0,336	0,00	0,382
1 528	0,237	0,000	0,001	0,114	0,76	0,886
1 731	0,253	0,651	0,000	0,244	0,75	0,105
1 564	0,201	0,353	0,000	0,464	0,80	0,183
1 552	0,146	0,167	0,008	0,679	0,85	0,154
1 554	0,084	0,046	0,000	0,454	0,92	0,500
1 533	0,162	0,000	0,000	0,000	0,84	1,000
1 529	0,064	0,000	0,000	0,066	0,94	0,934
1 542	0,179	0,530	0,036	0,103	0,78	0,367
1 549	0,016	0,024	0,059	0,548	0,93	0,428
1 545	0,000	0,000	0,000	0,000	1,00	1,000
1 547	0,085	0,000	0,000	1,000	0,92	0,000
1 640	0,000	0,000	0,113	0,062	0,89	0,938
1 558	0,085	0,000	0,000	0,044	0,92	0,956
1 527	0,238	0,000	0,000	0,107	0,76	0,893
1 532	0,112	0,083	0,007	0,459	0,88	0,458
1 538	0,055	0,225	0,005	0,300	0,94	0,475

Table 3: Unplanned events and other issues affecting the energy efficiency and savings

Store-Id	Comment
1560	Repair of the ventilation control unit
1551	Got HVAC and refreigeration system from discontinued store. Almost new system
1562	Optimized heat exchanger, district heating kicked in, thus electric consumption down
1563	Repair of the steering unit for the snow smelting system
1557	Cooling disks changed, customer own cost
1531	Oil boiler phased out, expected larger increase in energy consumption
1558	A difficult store. Lots of minor technical issues not resolved
1532	Oil boiler phased out, heatpump installed
1538	Oil boiler phased out, heatpump installed

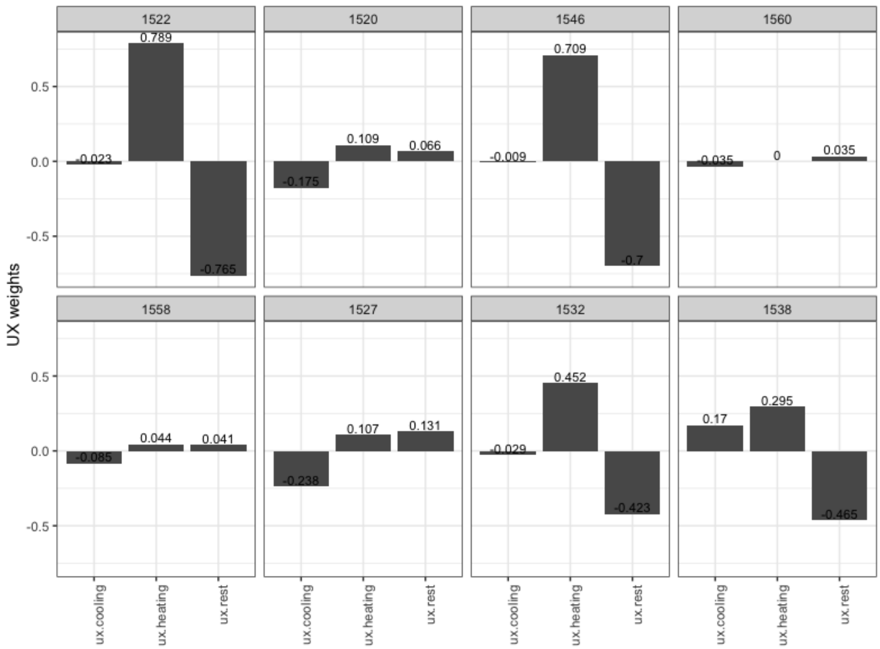


Figure 5: Change in UX weights after ECM

the DEA efficiency scores from when the ECMs have been in effect for a year is discussed, together with a comparison of the change in efficiency from the year before the implemented ECMs. The multiplier model is applied to accommodate this discussion. At last a general discussion of the framework is offered.

4.1 Step 1: Broken line and energy efficiency

The M&V process is often conducted after the ECMs are installed. However, the results from baseline modeling, such as the BL model, may be useful before implementation of the ECMs. In the BL model the changing point temperature (CPT), and the demand for cooling and heating could be useful input when the ESCO performs a complete review of all the buildings technical infrastructure (the audit phase). For instance, in figure 4 the BL model was used to produce the ET-curves for 8 different stores (sample of the portfolio under study). While the main objective of this plot was to visualize and estimate the energy savings, the ET-curve for the year before the installed ECMs give useful insight about the heating and cooling demand. This may be seen by comparing store id 1527 and 1538. While the CPT value is about the same around 10°C, the demand for heating is quite different, and the ET-curve for store 1527 has a much steeper curve for heating. Ranking the building portfolio based on heating and cooling demand across the portfolio may give useful insights into the condition of the technical infrastructure (HVAC), but may also indicate differences in the building envelope. Additionally, training the baseline models during the audit phase, and not postpone this step until the M&V phase, gives two prevalent advantages. First, during the audit phase the ESCO has full attention to the buildings under contract, and at this stage it is likely to be easier to understand any issues with the data that will be used to train the baseline models. Often the energy data has concerns that needs to be considered (technical issues with the meters, building/technical infrastructure non-operative for periods, promotional events). Getting attention to these issues is easier at a time when the ESCO is already working on related issues.

Also note that previous studies use default published values for CPT. In the U.S. this value is 18 degree Celsius, which is the value used in a previous benchmarking study of residential buildings [16]. However, food stores often have a technical system that is used for cooling, heating and heat recovery. Depending on building material, geographic location, building size and the age and condition of the technical system the CPT values may be quite different between buildings, thus using a default value may lead to incorrect calculation of weather effects on the energy consumption. Hence, estimating the CPT values from the BL model gives reliable insight into the CPT differences. For instance, [17] finds that the median CPT value for 95 food retail stores in Norway was 6.8°C, while [34] find the equivalent for office buildings to be 11 °C.

Furthermore, the CPT has often been used as a reference to calculate the degree days, both heating and/or cooling degree days. The degree days represents a measure that is often used to normalize weather effects between buildings in different temperature zones. For example, this is the approach that [16] use in their benchmarking study. Nonetheless, since the CPT values varies across food retail stores using a fixed CPT value may decrease the precision of the estimated cooling and heating degree days.

At last, previous research has established that the stakeholders, due to cost and constraints, are less likely to embrace innovations from a modeling point of view, and that a one temperature variable model to estimate energy savings is quite standard [35]. Also, as [36] finds, interpretability is important to gain acceptance from clients. As such, the BL model was easy to interpret through the ET-curves, and the results was communicated effectivly.

4.2 Step 2: TVB and Optimization

In figure 3 an example of the TVB hourly model was presented. The TVB model shows that the savings can be constricted on a detailed level. Furthermore, the data pipeline allowed the frequency of the analysis to be updated on a daily level. This approach enabled detection of issues that may otherwise have been overlooked, e.g. lights that are on during the night or a ventilation system that runs in day-mode during night, or other non-optimal settings. Hence, the ESCO could closely monitor the performance of the newly installed ECMs and make ongoing adjustments.

Furthermore, the model reliability was tested against the CV-RMSE, which is recommended by the IPMVP, and accordingly should be below 25% for the model to be accepted [33]. Large CV-RMSE (>25%) was flagged and inspected through visualizations of the data in the reference period. Often large CV-RMSE could be led back to the food retail store being out of operation and/or some other uncommon activity, for instance in-store promotion activities. Consequently, the CV-RMSE was actively used to understand the modeling results together with all the involved stakeholders. As a results, this collaboration led to a change of reference year for many buildings, and/or a better understanding of the actual data used as training data, e.g. missing and/or unusual observations.

4.3 Step 3: DEA efficiency and the energy savings

In table 1 the energy savings from the BL and the TVB model were presented. The results across the portfolio of 34 food retail stores varies between an effect of the ECM where the lowest performing store had an increase of +2.7% energy consumption (store-id 1538) to the best performing store which had a decrease in energy consumption of 55% (store-id 1522). Still, there is no practical difference between the results from the TVB and the BL model. This is reassuring and gives the results credibility. Furthermore, this also corresponds with the findings in [27] where the BL model was compared with the TVB model across 5 different food retail buildings. Note that the only ECM implemented in store-id 1538 was optimization of the unit that controlled the HVAC. Discussing this finding with the ESCO it was found that some of the control units did not work as expected, and ongoing work to address this was commenced. While the retrofitting project generated an average energy savings of 17,6%, the average efficiency scores both before and after the installed ECMs were 74%. The expectation up front from the ESCO was that the efficiency would increase. Despite, during the project many unplanned events occurred. In particular, substituting oil boilers with heat pumps in three of the stores increased the energy consumption. More details about these events are given in table 3. Thus, it would make sense to estimate what the energy consumption would have been in the reference year if the heating was not handled by an oil boiler. However, the information about this came about during the project, and was initially not planed.

Furthermore, investigating the individual stores efficiency score is useful to better understand the effect of the savings. For example, comparing store-id '1522' and '1520' with the other stores the efficiency scores attained for these two stores had the largest potential, hence it was logical that the actual reductions in energy savings reflected the potential. We also find that the efficiency scores compared to the reference period with the ECM period increased for these two stores by 30.9% and 49.5%. Interestingly the largest efficiency gain was for the store with a somewhat lower energy reduction. It is not easy to point to a particular cause for this difference, however it may be sensible for the ESCO to analyze in further detail the implemented ECMs in these two stores to gain a better understanding of the differences to improve results for similar projects in the future.

Finally, table 2 and figure 5 present the UX weights from the multiplier model before (ref.) and after (2021) the implemented ECMs. For example, the weight for heating (UX Heat ref.) was 0,0 for store-id '1522' and in the reference period that weight increased to 0,789 after the implemented ECMs (UX Heat 2021). As mentioned previously, this may imply that the part of the increase in the DEA-CRS efficiency scores can be attributed to an improvement in the heating system. The same finding can be seen for store-id '1546.' Note that the weights sum to 1. Furthermore, a low weight on heating in relation to cooling means that this store should investigate the causes behind the energy consumption consumed by heating.

4.4 General discussion

A retrofitting project often starts with an audit of the building portfolio under contract. This involves a complete review of the technical infrastructure in the buildings. It is important to map the condition of the different components and to clarify which ECMs are prioritized and how the installation should progress. The potential for energy savings are often substantial. For instance, previous studies find that energy consumption from HVAC could save 40% of building energy without compromising occupant thermal comfort [37]. Other studies within the Norwegian food sector has documented savings between 25% and 55% [27]. Furthermore, during the audit phase the energy performance is often been evaluated with energy use intensity (EUI, e.g. kWh/m2). However, a number of other important variables may affect energy consumption. The IEA Annex 53 project sets fourth six factors that determine energy performance: (1) climate, (2) building envelop, (3) building services and energy systems, (4) building operation and maintenance, (5) occupants' activities and behavior and (6) indoor environmental quality provided [38]. In this paper the attention was focused on climate, opening hours and the difference in demand for cooling and heating (a proxy for building envelop, but also as part of the energy system).

When an ESCO signs an EPC contract several issues can complicate the partnership with the customer. The contract for the 34 stores studied within this paper has a length of 10 years. Consequently, it is vital to have a strict regulated contract that covers as many different scenarios as possible. However, is it not easy to foresee all possible issues that may arise over such a time span. Because of this it is essential to have reliable and appropriate methods to document the savings and the energy efficiency as a result of the project. The demonstrated framework within the paper may be used to accomplish such a task, leading to a more more holistic analytically approach where the joint effect of both estimated savings and efficiency are taken into account.

The collaboration with the ESCO showed that very little analytically resources went into the audit phase, and only after implementing the ECMs the demand for baseline models occurred. The ESCO had substantial knowledge about the technical infrastructure, however, when benchmarking the energy efficiency the only performance indicator used was energy intensity (kWh/m2). Our proposed 3-step framework find that it may be useful to extend this perspective when benchmarking energy efficiency in buildings. The framework offers a tool that the ESCOs can apply to document energy efficiency and energy savings documentation; we find the methods to be valuable tools to monitor efficiency and savings throughout the retrofitting project.

4.5 Limitations and future research

The IEA Annex 53 project sets fourth six factors that determine energy performance: (1) climate, (2) building envelop, (3) building services and energy systems, (4) building operation and maintenance, (5) occupants' activities and behavior and (6) indoor environmental quality provided [38]. In the present study only climate, opening hours and cooling and heating demand was used to benchmark the efficiency. However, more detailed information about the building envelop, the occupants, and maintenance would improve the efficiency scores. Additionally, it would be interesting to use data from sub-meters as input variables in the DEA (e.g. ventilation, lighting, refrigeration). Inclusion of these new input variables could enhance insight about how individual ECMs contribute to energy efficiency. Hence, future research that integrates these features into the benchmarking analysis will be useful.

Further, trustworthy calculations of energy savings are critical to convince stakeholders in energy efficiency projects of the benefit and the cost-effectiveness of the investments [39]. Recent research have offered several solutions that automate this process and delivers energy savings in near real-time [40–42]. However, previous research within this area focus the attention solely on energy savings, and not efficiency. It would be very useful if such tools integrated the efficiency perspective as that would give a broader use case through the retrofitting project.

At last, a drawback of DEA is that it relates residuals to inefficient units. Moreover, due to its nonparametric nature it can not provide a specific equation that relates the input and output [43]. To improve these defects, the stochastic frontier analysis (SFA) approach has been tested in several previous studies to evaluate energy efficiency performance [44–46]. A replication of the efficiency scores that was produced with DEA within this paper using SFA would be a welcome future contribution.

5 Conclusions

In this paper we have demonstrated a 3-step framework that Energy Service companies can use to document both the energy savings and the energy efficiency in retrofitting projects. This has been accomplished using data from a Norwegian ESCO who signed an EPC project for 34 food retail stores. The chosen framework was designed and tested in close cooperation with the ESCO to enable optimized outcome of the project, and the framework follows the logic workflow in retrofitting proejcts: (1) audit, (2) implementation, (3) Measurement and Verification. In the starting phase during the energy audit of the building portfolio the CRS-DEA efficiency is useful to enhance understanding of the potential energy savings. Furthermore, using the ET - curves from the BL model was useful to gain a better understanding of the demand for heating and cooling in the food retail stores.

During implementation the TVB model provided detailed information on an hourly level, which enabled the ESCO to continuously optimize and adjust the ECMs during the implementation phase.

Furthermore, after the implemented ECMs both the TVB and the BL model served several purposes. First, both models was used to estimate the aggregated energy savings. The overall results from the models were similar, which gave the results reliability. The final step consisted of conducting a CRS-DEA on the data after the implemented ECMs, and to compare the results of the efficiency scores before and after the ECM. In that context the aggregated savings from the TVB/BL model was presented, and the ESCO could then relate the actual energy savings to the potential savings, and the change in efficiency as a result of the ECMs, and through the DEA multiplier model at the same time relate cooling and heating to efficiency.

Documenting energy savings and efficiency in large building portfolios over many years requires a solid framework that can be used both as a documentation tool for the customer, and at the same time continuously give useful insights that the ESCO can use on a day-to-day basis. As such, the proposed 3-step framework may serve that objective well.

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] Claridge DE, Culp CH, Liu M, Deng S, Turner WD, Haberl JS. Campus-wide continuous commissioning of university buildings. In: Proceedings of the 2000 ACEEE summer study of energy efficiency in buildings [Internet]. 2000. p. 101–12. Available from: https://www.eceee.org/library/conference_proceedings/ACEEE_buildings/2000/Panel_3/p3_8/

[3] Xinhua X. Model based building performance evaluation and diagnosis [Internet] [PhD thesis]. Ph. D. Dissertation, Hong Kong Polytechnic University; 2005. Available from: https://ira.lib.polyu.edu.hk//handle/10397/84709

[4] Enova byggstatistikk [Internet]. Enova SF; 2017. Available from: https://www.enova.no/download?objectPath=upload_images/5C6245BC2AD74248BB629BFA95145AA3.pdf&filename=Enovas%20byggstatistikk%202017.pdf

[5] 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' [Internet]. Vol. 69, Energy and Buildings. 2014. p. 515–24. Available from: https://doi.org/10.1016/j.enbuild.2013.11.004

[6] Kontokosta CE. Modeling the energy retrofit decision in commercial office buildings. Energy Build [Internet]. 2016 Nov;131:1–20. Available from: https://doi.org/10.1016/j.enbuild.2016.08.062

[7] EN 15217, energy performance of buildings—methods for expressing energy performance and for energy certification of buildings [Internet]. 2007. Available from: https://www.standard.no/no/Nettbutikk/produktkatalogen/Produktpresentasjon/?ProductID=283426

[8] Poel B, Cruchten G van, Balaras CA. Energy performance assessment of existing dwellings. Energy and Buildings [Internet]. 2007;39(4):393–403. Available from: https://doi.org/10.1016/j.enbuild.2006.08.008

[9] Wang S, Yan C, Xiao F. Quantitative energy performance assessment methods for existing buildings. Energy Build [Internet]. 2012 Dec;55:873–88. Available from: https://doi.org/10.1016/j.enbuild.2012.08.037

[10] Yu A, You J, Zhang H, Ma J. Estimation of industrial energy efficiency and corresponding spatial clustering in urban china by a meta-frontier model. Sustainable Cities and Society [Internet]. 2018 Nov;43:290–304. Available from: https://doi.org/10.1016/j.scs.2018.08.037

[11] Chauhan NS, Mohapatra PKJ, Pandey KP. Improving energy productivity in paddy production through benchmarking—an application of data envelopment analysis. Energy Convers Manage [Internet]. 2006 Jun;47(9):1063–85. Available from: https://doi.org/10.1016/j.enconman.2005.07.004

[12] Omid M, Ghojabeige F, Delshad M, Ahmadi H. Energy use pattern and benchmarking of selected greenhouses in iran using data envelopment analysis. Energy Convers Manage [Internet]. 2011 Jan;52(1):153–62. Available from: https://doi.org/10.1016/j.enconman.2010.06.054

[13] Vyas GS, Jha KN. Benchmarking green building attributes to achieve cost effectiveness using a data envelopment analysis. Sustainable Cities and Society [Internet]. 2017 Jan;28:127–34. Available from: https://doi.org/10.1016/j.scs.2016.08.028

[14] Lee W-S, Lee K-P. Benchmarking the performance of building energy management using data envelopment analysis. Appl Therm Eng [Internet]. 2009 Nov;29(16):3269–73. Available from: https://doi.org/10.1016/j.applthermaleng.2008.02.034

[15] Lee W-S, Kung C-K. Using climate classification to evaluate building energy performance. Energy [Internet]. 2011 Mar;36(3):1797–801. Available from: https://doi.org/10.1016/j.energy.2010.12.034

[16] Wang E, Shen Z, Alp N, Barry N. Benchmarking energy performance of residential buildings using two-stage multifactor data envelopment analysis with degree-day based simple-normalization approach. Energy Convers Manage [Internet]. 2015 Dec;106:530–42. Available from: https://doi.org/10.1016/j.enconman.2015.09.072

[17] Severinsen A, Holst HMS. Using machine learning and mathematical programming to benchmark energy efficiency of buildings. In: ECEEE SUMMER STUDY PROCEEDINGS [Internet]. Hyeres, Nice: ECEEE; 2017. p. 1083–9. Available from: https://www.eceee.org/library/conference_proceedings/eceee_Summer_Studies/2017/5-buildings-and-construction-technologies-and-systems/using-machine-learning-and-mathematical-programming-to-benchmark-energy-efficiency-of-buildings/

[18] Franconi E, Gee M, Goldberg M, Granderson J, Guiterman T, Li M, et al. The status and promise of advanced M&V: An overview of “M&V 2.0” methods, tools, and applications [Internet]. 2017. Available from: https://doi.org/10.2172/1350974

[19] R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2013. Available from: http://www.R-project.org/

[20] Bogetoft P, Otto L. Benchmarking with DEA and SFA. 2020.

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

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

[23] Kuchenhoff H, Ulm K. Comparison of statistical methods for assessing threshold limiting values in occupational epidemiology. Computational Statistics [Internet]. 1999;12:249–64. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4804

[24] Feder PI. The log likelihood ratio in segmented regression. The Annals of Statistics [Internet]. 1975;3(1):84–97. Available from: https://projecteuclid.org/journals/annals-of-statistics/volume-3/issue-1/The-Log-Likelihood-Ratio-in-Segmented-Regression/10.1214/aos/1176343000.full

[25] Hinkley DV. Inference in Two-Phase regression [Internet]. Vol. 66, Journal of the American Statistical Association. 1971. p. 736–43. Available from: https://doi.org/10.2307/2284220

[26] Muggeo VMR. Segmented: An R package to fit regression models with Broken-Line relationships. R News [Internet]. 2008;8:20–5. Available from: https://cran.r-project.org/doc/Rnews/Rnews_2008-1.pdf

[27] Severinsen A, Myrland Ø. Statistical learning to estimate energy savings from retrofitting in the norwegian food retail market. Renewable and Sustainable Energy Reviews [Internet]. 2022;167:112691. Available from: https://doi.org/10.1016/j.rser.2022.112691

[28] 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

[29] Severinsen A, Hyndman RJ. Quantification of energy savings from energy conservation measures in buildings using machine learning. In: ECEEE summer study proceedings [Internet]. 2019. p. 757–66. Available from: https://www.eceee.org/library/conference_proceedings/eceee_Summer_Studies/2019/4-monitoring-and-evaluation-for-greater-impact/quantification-of-energy-savings-from-energy-conservation-measures-in-buildings-using-machine-learning/

[30] Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. European Journal of Operational Research [Internet]. 1978;2(6):429–44. Available from: https://doi.org/10.1016/0377-2217(78)90138-8

[31] Banker RD, Charnes A, Cooper WW. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management science [Internet]. 1984;30(9):1078–92. Available from: https://doi.org/10.1287/mnsc.30.9.1078

[32] Bogetoft P, Otto L. Benchmarking with DEA, SFA, and R. In Springer Science & Business Media; 2010. p. 132–6.

[33] EVO Efficiency Valuation Organization. International performance measurement and verification protocol. 2016.

[34] Pedersen L. Load modeling of buildings in mixed energy distribution systems [Internet] [PhD thesis]. Doctoral thesis, Norwegian University of Science; Technology.; 2007. Available from: http://hdl.handle.net/11250/233327

[35] Agenis-Nevers M, Wang Y, Dugachard M, Salvazet R, Becker G, Chenu D. Measurement and verification for multiple buildings: An innovative baseline model selection framework applied to real energy performance contracts. Energy Build [Internet]. 2021 Oct;249:111183. Available from: https://doi.org/10.1016/j.enbuild.2021.111183

[36] Molnar C. Interpretable machine learning a guide for making black box models explainable [Internet]. https://leanpub.com; 2021. Available from: https://christophm.github.io/interpretable-ml-book/

[37] Rahman MM, Rasul MG, Khan MMK. Energy conservation measures in an institutional building in sub-tropical climate in australia. Appl Energy [Internet]. 2010 Oct;87(10):2994–3004. Available from: https://doi.org/10.1016/j.apenergy.2010.04.005

[38] IEA ECBCS Annex53, Annex53 total energy use in buildings: Analysis & evaluation methods [Internet]. IEA; Available from: http://www.ecbcsa53.org

[39] Granderson J, Touzani S, Custodio C, Sohn MD, Jump D, Fernandes S. Accuracy of automated measurement and verification (M&V) techniques for energy savings in commercial buildings. Appl Energy [Internet]. 2016 Jul;173:296–308. Available from: https://doi.org/10.1016/j.apenergy.2016.04.049

[40] Fagiani M, Severini M, Valenti M, Ferracuti F, Ciabattoni L, Squartini S. rEMpy: A comprehensive software framework for residential energy management. Energy Build [Internet]. 2018 Jul;171:131–43. Available from: https://doi.org/10.1016/j.enbuild.2018.04.023

[41] Ke M-T, Yeh C-H, Su C-J. Cloud computing platform for real-time measurement and verification of energy performance. Appl Energy [Internet]. 2017 Feb;188:497–507. Available from: https://doi.org/10.1016/j.apenergy.2016.12.034

[42] Severinsen A, Myrland Ø. ShinyRBase: Near real-time energy saving models using reactive programming. Norwegian University of Life Sciences; 2022. (NMBU working paper).

[43] Ding Y, Liu X. A comparative analysis of data-driven methods in building energy benchmarking. Energy Build [Internet]. 2020 Feb;209:109711. Available from: https://doi.org/10.1016/j.enbuild.2019.109711

[44] Buck J, Young D. The potential for energy efficiency gains in the canadian commercial building sector: A stochastic frontier study. Energy [Internet]. 2007 Sep;32(9):1769–80. Available from: https://doi.org/10.1016/j.energy.2006.11.008

[45] Filippini M, Hunt LC. US residential energy demand and energy efficiency: A stochastic demand frontier approach. Energy Econ [Internet]. 2012 Sep;34(5):1484–91. Available from: https://doi.org/10.1016/j.eneco.2012.06.013

[46] Yang Z, Roth J, Jain RK. DUE-B: Data-driven urban energy benchmarking of buildings using recursive partitioning and stochastic frontier analysis. Energy Build [Internet]. 2018 Mar;163:58–69. Available from: https://doi.org/10.1016/j.enbuild.2017.12.040

6.5 Paper 5: Forecasting and Technoeconomic Optimization of PV-Battery Systems for Commercial Buildings

Forecasting and technoeconomic optimization of PV-battery systems for commercial buildings

Jonathan Fagerström, Kari Aamodt Espegren
& Josefine Selj
Institute of Energy Technology
Instituttveien 18
NO-2007 Kjeller
Norway
jonathan.fagerstrom@ife.no

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

Keywords

photovoltaics, batteries, energy model, system simulation, peak load, load management

Abstract

The cost structure of electricity tariffs varies among countries. In Norway, it is being modified from a two-part tariff, where the cost is divided between a fixed installation cost (EUR/installation) and a cost for consumed electricity (EUR/kWh), to a three-part electricity tariff where customers additionally pay demand charges for capacity usage (EUR/kW). To combat demand charges, commercial customers are looking into supplementing PV installations with batteries to more efficiently reduce peak electricity demand, i.e. peak shaving. A crucial part of the complete energy system is also the energy management, where forecasting improves efficiency and economics. The objective of this work was to investigate the profitability with peak shaving in Norway for a commercial building. A forecasting algorithm for load prediction was developed, and the economic value of forecasting was determined for a PV-battery system. The load forecasting was developed using component-wise gradient boosting and the results from the model were verified against a renowned benchmarking load forecasting model. The economic value of forecasting was determined through simulations with Homer Energy Software that optimizes the net present cost of the systems. The results showed that battery storage was only economically beneficial when forecasting was deployed. Moreover, the cost savings came mainly from reduced demand charges, not from increased self-consumption of PV electricity. It was also discussed that the application of forecasting in an energy management system could be divided into three phases. One phase where forecasting is deployed to dimension energy system components in an early stage, one monthly forecast overview that identifies height and frequency of maximum peaks, and finally one high-resolution forecast that operates the battery on an hourly basis. Altogether, such an energy management system could additionally also be used by utility grid owners to schedule demand response actions for power quality control.

Introduction

The cost structure of electricity tariffs varies among countries. In Norway, it is being modified from a two-part tariff, where the cost is divided between a fixed installation cost (EUR/installation) and a cost for consumed electricity (EUR/kWh), to a three-part electricity tariff where customers additionally pay demand charges for capacity usage (EUR/kW). The demand charge reflects purchased energy per time unit, i.e. kWh/h. Simshauser (2016) argues that this three-part tariff is more efficient and reflects both cost elements of electricity distribution, capacity and energy. Fridge et al. (2018) took this study further and performed an analysis of how different electricity tariffs affect cost distribution between micro-grid owners and electricity distribution grid owners. They found that two-part tariffs encourage grid destabilization. Although the three-part tariff has not been fully implemented in Norway, commercial customers are billed for demand charges. A shift to the three-part tariff would create winners and losers (Simshauser 2016), and thus, there is a need for assessing the potential for cost savings by cutting the peak demand, i.e. peak shaving.

To combat demand charges, commercial customers are looking into supplementing PV installations with batteries to more efficiently perform peak shaving. However, the current PV installations do not usually include batteries, but as battery prices decline (IRENA 2017), it becomes interesting to consider a co-optimization of PV-battery systems. Comello et al. (2018) analyzed the profitability for PV-battery systems and found that systems with low-cost storage would be profitable. Indeed, the cost of storage is vital for profitability. To the authors' knowledge, no similar study has been conducted for the commercial sector in Norway.

Installation of the physical PV and batteries alone will not result in optimal solutions. A crucial part of the complete system is also the energy management, which is the control that eventually will improve efficiency and economics while reducing emissions. One conceptual framework for such a system was presented by Zhao et al. (2010) and includes both a cyber and a physical system. Forecasting is a crucial feature of the cyber-part and it has been shown to increase revenues from PV-battery systems, although it was highlighted that the actual benefits are strongly dependent on site-specific boundary conditions such as feed-in-limit, feed-in-tariff etc. (Litjens 2018).

Therefore, this work was initiated to investigate the profitability with peak shaving in Norway for a commercial building. There were two specific objectives, first to develop a forecasting algorithm for predicting electric load, and second to determine the economic value of using forecasts for efficient battery control.

Method

LOAD FORECASTING

To forecast the energy load for 2018, data from 2017 was used to train the model. Hourly electricity usage was collected from electric meters from the advanced metering infrastructure (AMI) system. Further, weather data on an hourly level was collected from the Norwegian Metrological Service[1].

Two different modelling techniques were used. First, the Tao vanilla benchmark model (TVB). This model was first published in Hong (2010) and was later used as a benchmarking model in the GEFCom2012 load forecasting competition (Hong, Pinson, and Fan 2014). The model performed among the best 25 of 100 teams. In the commercial software package SAS Energy Analytics, the TVB model is integrated as a standard load forecasting method. Further, because of the relatively straightforward specification and proved predictive performance the model is a good candidate to test other models against. The model is a multiple regression model

$$\begin{aligned} Y_t = {} & \beta_0 + Y_{t-1} + \beta_1 M_t + \beta_2 W_t + \beta_3 H_t + \beta_4 W_t H_t + \beta_5 T_t \\ & + \beta_6 T_t^2 + \beta_7 T_t^3 + \beta_8 T_t M_t + \beta_9 T_t^2 M_t + \beta_{10} T_t^3 M_t \\ & + \beta_{11} T_t H_t + \beta_{11} T_t^2 H_t + \beta_{11} T_t^3 H_t \end{aligned} \tag{1}$$

where Yt 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 we make two different TVB models, one with and one without the lagged dependent variable, Y_{t-1}.

This has some very important implications for how it is possible to apply the model in production. Without Y_{t-1} it is possible to predict as long as a year ahead (given that the model was trained on one year of data) and that the model is fed some realistic temperatures series for the different seasons. However, using Y_{t-1} we have to continuously score the model based on the latest data each hour. This will 'predict' any sudden "peaks" after the actual "peak".

In the next section, we present the gradient boosting approach. Previous research with boosting demonstrates excellent prediction performance within statistics and machine learning (Schapire and Freund 2012). Further, Bühlmann and Yu (2003) developed *component-wise gradient boosting (CW-GB)* 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). Also, boosting is robust against multicollinearity and flexible in terms of modelling different types of effects (Mayr and Hofner 2018). A similar approach was used by Taieb and Hyndman (2014) in the Kaggle global energy forecasting competition 2012 and ranked fourth out of 105 participating teams. Next, we provide a more detailed overview of the procedure.

We label the outcome variable, energy consumption, y and the predictors (temperature variables and calendar data) x_1, …, 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 generalized additive model 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

$$f^* := argmin_f \mathbb{E}_{y,x}[\rho(y, f(x^T))], \tag{2}$$

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

1. Initiate the function estimate $\hat{f}^{[0]}$.
2. Determine the set of *base-learners*. Each of the base-learners acts as a modelling 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 us the negative gradient vector
   $$\mathbf{u}^m = (u_i^{[m]})_{i=1,\ldots,n} := (\frac{\partial \rho}{\partial f}(y_i, \hat{f}^{[m-1]}(x_i^T)))_{,i=1,\ldots,n} \, .$$
   b. Fit each of the base learners individually to the negative gradient vector. We estimate the negative gradient $\mathbf{u}^m$ for all the vectors of the predicted values P.
   c. This step selects the base-learner that fits $\mathbf{u}^m$.

1. www.met.no

d. The current estimate is updated by setting $\hat{f}^{[m]} = \hat{f}^{[m-1]} + \nu\hat{u}^{[m]}$ where $0 < \nu \leq 1$.

4. Steps 3 and 4 are iterated until m_{stop} is reached.

and p is the number of features in the model. In step 3c) and 3d) the algorithm performs variable and model selection. There are two hyperparameters that need to be estimated, M, the number of steps, and ν, a step length factor. However, Friedman (2001) shows that a small ν can prevent over fitting. We set ν=0.15 and M=400. Further, the CW-GB had 32 different variables available (temperature data, holidays, calendar data) and the algorithm then chose the best set of variables from these.

ENERGY SYSTEM OPTIMIZATION

A grid-connected commercial building in the retail sector located in Norway was chosen for this study. The yearly electricity consumption was about 2,900 MWh. The volatility of the consumption profile can be used as an indication of the profitability with peak shaving. Lind et al. (2017) found that the coefficient of variation (CV) can be used to give the consumption profile a score, where buildings with high CV-values have a higher probability of benefiting from peak shaving. CV-value is calculated as the standard deviation to the average value, and the object in this work had a low score of 0.37.

Homer Energy

The economic value of forecasting was determined using the commercial software Homer Pro and Homer Grid. Homer is an acronym for Hybrid Optimization of Multiple Energy Resources. Both software programs were developed by Homer Energy to simulate, optimize, and perform a sensitivity analysis of on- and off-grid micro grids (Lambert et al., 2006, Bahramara et al. 2016]. Homer optimizes the system based on minimizing the objective function Net Present Cost (NPC), which is the value of all the costs the system incurs over its lifetime, minus the present value of all the revenue it earns over its lifetime. Costs include capital costs, replacement costs, O&M costs, and the costs of buying power from the grid. Revenues include salvage value and grid sales revenue.

Economic value of forecasting

Four cases were designed to determine the value of forecasting.

Case A was simulated in Homer Pro with a cycle charging dispatch strategy, which is common today in systems with little renewable power generation. Cycle charging means that whenever a generator is running, in this case, grid or PV, the battery is charged until it reaches a specified state of charge, in this case, 95 %. Moreover, there is no forecasting applied, and hence no control to capture excess PV electricity or to avoid grid charging of battery. During case A, the capacity of both PV and battery were optimized to determine optimal component dimensioning for a case without forecasting.

Case B was simulated in Homer Grid with a forecasting dispatch controller. The forecasting feature is not included in Homer Pro. The intention with case B was to determine the optimal battery size if forecasting was applied to a building where PV had already been installed based on optimization without forecasting. Therefore, the optimal PV size from case A was applied and only the battery component was optimized. The forecasting controller sees 48 h ahead and determines how to use the system components for demand charge reductions and energy arbitrage while serving the electrical load.

Case C was similar to Case B but here also the PV component was optimized, thus a complete co-optimization of PV and battery using the forecasting controller in Homer Grid.

Case D was simulated in Homer Pro to show how the project economics are affected if forecasting is not applied to a system that was dimensioned based on an optimal case, i.e. case C.

Modelling constraints

The system components included in the optimization are PV, battery, converter, load, and power grid. The PV component is modelled as polycrystalline silicon 60 cell module (Jinko JKM275-60). The model included temperature effects and a derating factor (losses from wiring, soiling, snow cover, and degradation) of 92 %. Solar irradiance and temperature data were imported through Homer from the NASA Surface Meteorology and Solar Energy database and included monthly global horizontal radiation, averaged from July 1983 to June 2005. The modules were simulated to face south with a tilt of 20 °. The installation cost was set to 1,020 EUR/kWp and the costs related to operation and maintenance were neglected. Lifetime was set to 25 years, even though the effective lifetime of the PV system may be substantially longer.

The battery component was modelled as a generic Li-ion with 90 % round-trip efficiency and with a C-rate of 1 and 3 for charging and discharging, respectively. The initial state of charge was set to 50 % and minimum state of charge to 10 %. The lifetime of the battery was set to either 15 years or 3,000 cycles, whichever comes first. The cost of installation was set to 310 EUR/kWh and replacement of the battery was set to 150 EUR/kWh (IRENA 2017). The converter was modelled as a generic system converter with an efficiency of 96 %. The inverter and rectifier capacities were equally large and the installation cost was included in the cost of PV and battery components.

The grid component was designed to reflect the local conditions. Electricity prices (EUR/kWh) consisted of spot prices, utility fee, demand charge fee, and a specific cost/benefit price for the building owner. Historical prices from 2017 were imported from Nordpool (Nordpool). The cost of power (EUR/kW) was set to 15 EUR/kW for December–February, 8 EUR/kW for March and November, and 2 EUR/kW for April–October.

The project lifetime of the simulation was set to 25 years, the interest rate to 3.5 %, and the inflation rate to 2 %. The forecasting dispatch strategy applied in the optimizations uses both load, PV, and price forecasting. This paper shows results on successful load forecasting, but do not analyse the possibilities of PV and price forecasting. However, price forecasts for the next 24 h are available on Nordpool (Nordpool) and can be incorporated into an actual dispatch strategy. Forecasting of PV production has been studied elsewhere (Chun Sing et al. 2017) and seems to give accurate results. The incorporation of these three forecasting algorithms would enable an optimization similar to the one used by Homer Grid. It should still be mentioned that real forecasts are not 100 % accurate, in contrast to the "perfect foresight" that Homer applies.

Results and discussion

LOAD FORECASTING

The results of the forecasts from the two different modelling strategies are described in the following figures. Figure 1 shows the TVB and the CW-GB model when the dependent variable lagged one hour is used as an explanatory variable. The TVB model is included because it has a previous track record of good load forecasting abilities; hence it is useful to compare the CW-GB results against a benchmark. As can be seen from Figure 1 both the TVB and the CW-GB models follows the actual load (kW) closely. The CW-GB and the TVB has a CV(RMSE) equal to 0.114 and 0.124, respectively. For example, 'ASHRAE' specifies that the CV(RMSE) should be less than 25 % if 12 months of post-measure data are used (American Society of Heating, Refrigeration and Air Conditioning Engineers 2014). The CV(RMSE) for both the models are well below the ASHRAE requirements. However, from a practical perspective using Y_{t-1} is challenging in production. The models need to be updated every hour and will not be able to predict a sudden "peak" in demand.

Figure 2 shows the TVB and the CW-GB model without Yt-1 as an explanatory variable. Both the models follow each other relatively close, but the predictions are not as good as the models with the lagged dependent variable. Also, the predictions are far from the actual loads the first 9 days of January but perform somewhat better for the rest of January. The CV(RMSE) for the CW-GB is 0.323 and 0.367 for the TVB. However, the actual building that these two models were developed for had a lot of different equipment installed, many of which were used on an ad hoc basis, thus difficult to predict.

ECONOMIC VALUE OF FORECASTING

Table 1 presents the four different cases that were evaluated and highlights four main points based on optimization of NPC. First, the NPC was lowest for case C where all components were co-optimized using a 48 h forecasting horizon. During a 25-year period it would save about EUR 10,200 compared to case A with standard cycle charging battery control, and EUR 61,300 compared to case D where forecasting is not applied. Second, the battery is only economically beneficial if forecasting is applied. The difference in NPC is small but applying forecasting do also allow for a larger PV capacity. Third, PV and battery should be co-optimized since battery size affects optimal PV capacity. Fourth, installation of a system optimized using forecasting results in the highest NPC if standard cycle charging control is used.

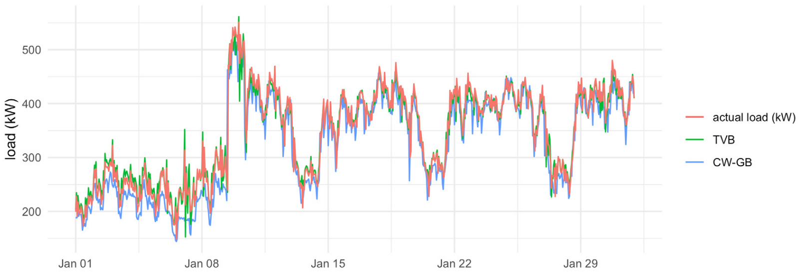


Figure 1. Actual loads (kW) for January 2018, and the predicted loads from TVB and CW-GB models, both models with the dependent variable lagged 1 hour.

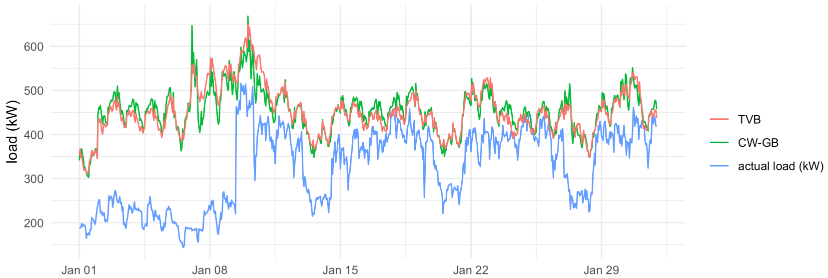


Figure 2. Actual loads (kW) for January 2018, and the predicted loads from TVB and CW-GB models, both models without the dependent variable lagged 1 hour.

Table 1. System dimensioning and project lifetime economics.

Control	PV (kWp)	Battery (kWh)	COE* (EUR/kWh)	NPC** (M EUR)	Optimization
A) Cycle charging	240	0	0.0652	3.92	All components
B) Forecasting	240	135	0.0650	3.91	Only battery
C) Forecasting	322	135	0.0649	3.91	All components
D) Cycle charging	322	135	0.0658	3.97	No components

** COE denotes Levelized Cost of Electricity. ** NPC denotes total Net Present Cost.*

Table 2. Monthly level of peak shaving reduction for the different cases, in % and in EUR.

Month	Case A	Case B	Case C		Case D
	Peak reduction (%)	Peak reduction (%)	Peak reduction (%)	Demand Charge Saving (EUR)	Peak reduction (%)
January	0	12	12	1,680	5
February	0	9	9	950	0
March	0	11	11	630	0
April	0	6	6	60	0
May	9	15	15	170	9
June	10	27	27	350	10
July	6	17	18	170	8
August	7	20	22	200	8
September	14	23	24	170	15
October	7	29	32	220	9
November	9	26	26	830	11
December	0	17	17	1,580	0

The benefits of forecasting arise from both increased self-consumption of PV electricity and reduced costs with peak shaving. Sales of PV electricity was reduced from 512 kWh (Case A) to 94 kWh (Case B) when forecasting was applied and shows consequently that increased self-consumption is not the reason to why battery and forecasting make economic sense. Total PV production for Case C was about 338 MWh. Table 2 summarizes monthly peak shaving levels for the different cases, as well as monthly demand charge savings for Case C. The maximum peak shaving occurred for Case C in October where the co-optimized PV-battery system shaved 32 % of the monthly peak, which is in line with results from Leadbetter and Swan (2012) that presented peak reduction between 28 and 49 %. There was however some months with low peak reductions. In terms of economics, it is seen in Table 2 that Case C saves between 1,680 and 60 EUR/month due to reduced demand charges. Table 2 further shows that Case D did not achieve efficient peak shaving even though the system components were the same size as Case C. This shows that accurate forecasting is crucial to achieving a low-cost system.

Energy storage behind-the-meter, as shown in this paper, is a way to cut costs for the building owner through peak shaving. Another possibility for building owners to cut payback time of behind-the-meter storage is to rent storage capacity to the power grid to enable control of power grid stability in front-of-meter. Whether the battery capacity presented in this paper would be useful for this purpose, was not analysed. Chun Sing et al. (2017) studied large-scale PV-storage installations and concluded that energy storage could limit stability issues related to frequency and voltage. A lab-scale experiment for such a system was conducted by Young-Jin et al. (2017) with promising results. The implementation of such features would require new business models, but it is speculated that through an energy management solution as presented in this work, it would be possible for buildings to have a time-stamped forecast of net power purchase from the grid. This way, it would be possible to schedule actions to control power grid stability, not only by the temporal shutdown of equipment as is the case for certain larger industrial customers, but also by distributing power from behind-the-meter battery to power grid. Ranaweera et al. (2017) presented a battery control method that could serve such a purpose.

Setting up an efficient energy management system for a building might consist of three phases. First, a robust dimensioning of system components (PV, battery, inverters) in an early stage. This phase is covered in the current paper. Second, there is a need for a monthly overview that identifies the maximum peak that will set the cost for the month. Both height of peak (kW) and timing (day) of month should be identified. This forecast should also provide a frequency of these peaks, i.e. whether they occur once or several times a month. Results from this paper show that both TVB and CW-GB with Yt-1 is efficient for this overview. Third, a higher resolution forecast, preferably down to 15 minutes, should identify how to operate the battery on a day-to-day basis. Results from this paper show that model TVB and CW-GB without Yt-1 may have potential, but the building in the current study had some unexplained variation that was difficult to predict. Forecasting both the monthly peak and the day-to-day high-resolution peak is important to get the most economic gains out of the system. There might be no reason to discharge the battery during hours where the demand is much lower than the monthly peak, although it might be beneficial to charge the battery during hours with low electricity price or to capture excess PV electricity. However, if new business models are introduced that allows the energy storage owner to sell electricity back to the power grid in order to control power grid quality, it might be more beneficial to use the battery for both peak shaving and power grid control.

Conclusion

Based on the results from this work, the following conclusions are highlighted:

- Accurate forecasting of electricity demand can be performed with both the TVB and the CW-GB model, but for the building in this study Yt-1 is crucial as a predictor, hence the model will be challenging in production
- During the design of PV-battery systems, the components should be co-optimized.
- Battery storage was only economically beneficial when forecasting was deployed.
- Energy management with forecasting improved profitability and potentially between EUR 10,200–61,300 can be saved during a 25-year period.
- For the optimal case, most of the savings came from peak shaving, not from increased self-consumption.

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.

Bahramara et al., 2016, Renewable and Sustainable Energy Reviews 62 (2016) 609–620.

Bühlmann, Peter, and Bin Yu. 2003. "Boosting with the L 2 Loss: Regression and Classification." J. Am. Stat. Assoc. 98 (462). Taylor & Francis: 324–39.

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

Comello et al., 2018, Renewable and Sustainable Energy Reviews 92 (2018) 744–756.

Fridgen et al., 2018, Applied Energy 210 (2018) 800–814.

Friedman, Jerome H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." *Ann. Stat.* 29 (5).

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.

Irena (2017), Electricity Storage and Renewables: Costs and Markets to 2030, International Renewable Energy Agency, Abu Dhabi.

Lambert et al., 2006, Integration of Alternative Sources of Energy by Farret and Simões, chapter 15 Micropower system modelling with Homer.

Leadbetter Jason and Swan Lucas, 2012, Energy and Buildings 55 (2012) 685–692.

Lind et al., 2017, eceee Summer Study proceedings, 5-327-17.

Litjens et al., 2018, Applied Energy 221 (2018) 358–373.

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

Nordpool. https://www.nordpoolgroup.com/historical-market-data/, access 02.01.2019

Ranaweera et al. 2017, Renewable Energy 113 (2017) 1099–1110.

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

Simshauser, 2016, Energy Economics 54 (2016) 108–122.

Sing, Chun et al., 2017, Renewable and Sustainable Energy Reviews 78 (2017) 439–451.

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

Young-Jin et al. 2017, IEEE Transactions on Power Systems, VOL. 32, NO. 1, January 2017.

Zhao et al., 2010, 2010 9th IEEE/IAS International Conference on Industry Applications.

Alexander Severinsen

School of Economics and Business
Norwegian University of
Life Sciences (NMBU)
P.O Box 5003
N-1432 Ås, Norway

Telephone: +47 993 77455
e-mail:
alexander.severinsen@gmail.com
alexander.severinsen@nmbu.no

Thesis number: 2022:60
ISSN: 1894-6402
ISBN: 978-82-575-2009-0

Alexander Severinsen was born in Tromsø, Norway, in 1974. He holds a master's degree in fishery science from the Norwegian College of Fishery Science.

The IPCC Sixth Assessment Report leaves little doubt that we urgently need to respond to be able to reduce human-induced climate change. The report clearly states that human activities is causing alarming and widespread disruption in nature and is affecting billions of people. Floods, heatwaves, and droughts are seen more often than ever, and unfortunately, people who are least able to struggle through are most affected. To avoid ascending loss of life, infrastructure, and biodiversity we must quickly make major cuts in greenhouse gas emissions (GHG) 1.

Buildings worldwide consume some 40% of all produced energy and are significant contributors to GHG emissions. Hence, energy efficiency retrofitting is a fundamental step in reducing energy consumption. However, one important barrier that hinders renovation projects is uncertainty regarding the expected savings. The main objective of this thesis is to contribute to lower that barrier and to deliver reliable methods to be used to document and monitor energy savings in retrofitting projects.

Professor Atle Guttormsen and Professor Øystein Myrland were Alexander's supervisor.

Alexander Severinsen currently works as a Machine Learning Engineer at the AI department at Aneo AS.

ISBN: 978-82-575-2009-0
ISSN: 1894-6402

Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no

116451 / ANDVORDGRAFISK.NO