UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

PhD Thesis

Data-driven methodologies for evaluation and recommendation of energy efficiency measures in buildings. Applications in a big data environment

Benedetto Grillone

Thesis presented for the degree of Doctor by Universitat Politècnica de Catalunya
PhD program in Electrical Engineering

Thesis directors
Prof. Dr. Andreas Sumper
Dr. Stoyan Danov

Barcelona, 2021

ὁ δὲ ἀνεξέταστος βίος οὐ βιωτὸς ἀνθρώπῳ
Socrates (470–399 BC)

Acknowledgements

This thesis is dedicated to my parents Nunzia and Guglielmo, my brother Lenny and my grandparents Grazia and Felice. Their love and unconditional support helped me become the person I am today and this work would have not been possible without them. I also want to express my gratitude to all my uncles, aunts and cousins, and I would like to extend the dedication of this thesis to the loving memory of my uncle Luciano.

I want to thank my directors, Dr. Stoyan Danov and Dr. Andreas Sumper, for accompanying me on this journey and always advising me in the best possible way. During these three years, I consistently found in you the guidance and support necessary to carry out this work.

Thanks to all the colleagues in BEE Group that I had the pleasure to work with in the last years. This thesis work is also a result of the joint efforts of our team. Thanks especially to Gerard Mor for the guidance and technical advice continuously provided during these three years.

I also want to thank all the teachers and professors that I met until now, who always pushed me to reach my full potential and sparked my passion for lifelong learning.

Finally, to all my close friends and to all the people I met in Barcelona, that made me always feel at home, you're too many to mention, but each of you knows how you contributed to this achievement. This thesis would have been much more difficult if it was not for you.

Abstract

In order to reach the goal set in the Paris agreement of limiting the rise in global average temperature well below 2 °C compared to pre-industrial levels, massive efforts to reduce global greenhouse gas emissions are required. The building sector is currently responsible for about 28% of total global CO_2 emissions, meaning that there is substantial savings potential lying in the correct energy management of buildings and the implementation of renovation strategies. Digital tools and data-driven techniques are rapidly gaining momentum as approaches that are able to harness the large amount of data gathered in the building sector and provide solutions able to reduce the carbon footprint of the built environment.

The objective of this doctoral thesis is to investigate the potential of data-driven techniques in different applications aimed at improving energy efficiency in buildings. More specifically, different novel approaches to verify energy savings, characterize consumption patterns, and recommend energy retrofitting strategies are described. The presented methodologies prove to be powerful tools that can produce valuable, actionable insights for energy managers and other stakeholders.

Initially, a comprehensive and detailed overview is provided of different state-of-the-art methodologies to quantify energy efficiency savings and to predict the impact of retrofitting strategies in buildings. Strengths and weaknesses of the analyzed approaches are discussed, and guidance is provided to assess the best performing methodology depending on the case in analysis and data available. Among the reviewed approaches there are statistical and machine learning models, Bayesian methods, deterministic approaches, and hybrid techniques combining deterministic and data-driven models.

Subsequently, a novel data-driven methodology is proposed to perform measurement and verification calculations, with the main focus on non-residential buildings and facilities. The approach is based on the extraction of frequent consumption profile patterns and on a novel technique able to evaluate the building's weather dependence. This information is used to design a model that can accurately estimate achieved energy savings at daily scale. The method was tested on two use-cases, one using synthetic data generated using a building energy simulation software and one using monitoring data from three existing buildings in Catalonia. The results obtained with the proposed methodology are compared with the ones provided by

a state-of-the-art model, showing accuracy improvement and increased robustness to missing data.

The second data-driven tool that developed in this research work is a Bayesian linear regression methodology to calculate hourly energy baseline predictions in non-residential buildings and characterize their consumption patterns. The approach was tested on 1578 non-residential buildings that are part of a large building energy consumption open dataset. The results show that the Bayesian methodology is able to provide accurate baseline estimations with an explainable and intuitive model. Special focus is also given to uncertainty estimations, which are inherently provided by Bayesian techniques and have great importance in risk assessments for energy efficiency projects.

Finally, a concept methodology that can be used to recommend and prioritize energy efficiency projects in buildings and facilities is presented. This data-driven approach is based on the comparison of groups of similar buildings and on an algorithm that can map savings obtained with energy renovation strategies to the characteristics of the buildings where they were implemented. Recommendation for implementation of such a methodology in big data building energy management platforms is provided.

Resumen

Para alcanzar el objetivo fijado en el acuerdo de París de limitar el aumento de la temperatura media mundial muy por debajo de los 2 °C con respecto a los niveles preindustriales, es necesario realizar esfuerzos masivos para reducir las emisiones mundiales de gases de efecto invernadero. El sector de la edificación es actualmente responsable de alrededor del 28% de las emisiones totales de CO_2 a nivel mundial, lo que significa que existe un potencial de ahorro sustancial en la correcta gestión energética de los edificios y en la aplicación de estrategias de renovación. Las herramientas digitales y las técnicas basadas en datos están ganando rápidamente impulso como enfoques capaces de aprovechar la gran cantidad de datos recopilados en el sector de la edificación y proporcionar soluciones capaces de reducir la huella de carbono del entorno construido.

El objetivo de esta tesis doctoral es investigar el potencial de las técnicas basadas en datos en diferentes aplicaciones destinadas a mejorar la eficiencia energética de los edificios. Más concretamente, se describen diferentes enfoques novedosos para verificar el ahorro de energía, caracterizar los patrones de consumo y recomendar estrategias de rehabilitación energética. Las metodologías presentadas demuestran ser poderosas herramientas que pueden producir valiosos conocimientos para los gestores energéticos y otras partes interesadas.

En primer lugar, se ofrece una visión general y detallada de las distintas metodologías más avanzadas para cuantificar el ahorro de energía y predecir el impacto de las estrategias de rehabilitación en los edificios. Se discuten los puntos fuertes y débiles de los enfoques analizados y se ofrecen orientaciones para evaluar la metodología más eficaz en función del caso en análisis y de los datos disponibles. Entre los enfoques revisados hay modelos estadísticos y de aprendizaje automático, métodos Bayesianos, enfoques deterministas y técnicas híbridas que combinan modelos deterministas y basados en datos.

Posteriormente, se propone una novedosa metodología basada en datos para realizar cálculos de medición y verificación, centrada principalmente en edificios e instalaciones no residenciales. El enfoque se basa en la extracción de patrones de perfiles de consumo frecuentes y en una técnica innovadora capaz de evaluar la dependencia climática del edificio. Esta información se utiliza para diseñar un modelo que puede estimar con precisión el ahorro energético conseguido a escala diaria. El método se ha probado en dos casos de uso, uno con datos sintéticos

generados mediante un software de simulación energética de edificios, y otro con datos de monitorización de tres edificios existentes en Cataluña. Los resultados obtenidos con la metodología propuesta se comparan con los proporcionados por un modelo de última generación, mostrando una mejora de la precisión y una mayor robustez ante la falta de datos.

La segunda herramienta basada en datos que se desarrolló en este trabajo de investigación es una metodología de regresión lineal Bayesiana para calcular las predicciones de línea base de energía horaria en edificios no residenciales y para caracterizar sus patrones de consumo. El enfoque se probó en 1578 edificios no residenciales que forman parte de un gran conjunto de datos abiertos de consumo energético de edificios. Los resultados muestran que la metodología Bayesiana es capaz de proporcionar estimaciones precisas de la línea de base con un modelo explicable e intuitivo. También se presta especial atención a las estimaciones de incertidumbre, que son inherentes a las técnicas bayesianas y que tienen gran importancia en las evaluaciones de riesgo de los proyectos de eficiencia energética.

Por último, se presenta una metodología conceptual que puede utilizarse para recomendar y priorizar proyectos de eficiencia energética en edificios e instalaciones. Este enfoque basado en datos se basa en la comparación de grupos de edificios similares y en un algoritmo que puede asociar los ahorros obtenidos con las estrategias de renovación energética a las características de los edificios en los que se aplicaron. Se recomiendan las aplicaciones de esta metodología en plataformas de gestión energética de edificios de big data.

Table of contents

Nomenclature

Acronyms

$ADVI$ Automatic Differentiation Variational Inference

ANN Artificial Neural Network

$ASHRAE$ American Society of Heating, Refrigerating and Air-Conditioning Engineers

$BART$ Bayesian Additive Regression Trees

BES Building Energy Simulation

BIC Bayesian Information Criterion

BPD Building Performance Database

CDD Cooling Degree Days

COP Coefficient Of Performance

$CV(RMSE)$ Coefficient of Variation of the Root Mean Squared Error

DHW Domestic Hot Water

EEM Energy Efficiency Measures

EUI Energy Usage Intensity

FRL Fallen Rule List

GA Genetic Algorithm

GAM Generalized Additive Model

GBM Gradient Boosting Machine

GHG Greenhouse Gas

GHI Global Horizontal Radiation

GMR Gaussian Mixture Regression

GOF Goodness of Fit

GP Gaussian Process

HDD Heating Degree Day

HDI High Density Interval

$HVAC$ Heating, Ventilation and Air Conditioning

$IPMVP$ International Performance Measurement and Verification Protocol

$M\&V$ Measurement and Verification

$MCEM$ Montecarlo Expectation Maximization

$MCMC$ Markov Chain Monte Carlo

$NMBE$ Normalized Mean Bias Error

NRE Non-Routine Event

$NSGA$ Non Sorted Genetic Algorithm

$NUTS$ No-U-Turn Sampler

$NZEB$ Nearly Zero Energy Building

PDF Probability Density Function

$RMSE$ Root Mean Squared Error

SVM Support Vector Machines

Chapter 1

Introduction

1.1 Background and motivation

The latest reports from the Intergovernmental Panel on Climate Change (IPCC) have demonstrated how the rising temperatures and extreme weather events that humanity is experiencing in the last decades are of anthropogenic origin and largely related to the emission of greenhouse gases (GHG) [1]. Figure 1.1, taken from the IPCC Summary for Policymakers report, clearly shows the impact of human activities during the last 150 years on the rise of global surface temperatures [2]. In panel a), changes in observed temperatures from the period between 1850 and 2020 are compared with reconstructed temperatures from paleoclimate archives. The vertical bar on the left shows the estimated temperature during the warmest multi-century period in the last 100.000 years, which occurred approximately 6500 years ago. In panel b), observed changes in global temperature are compared with two climate model simulations, one including both human and natural drivers (brown) and the other including only natural drivers (green).

In order to limit the potentially disastrous consequences of climate change on humankind, the representatives of 195 states gathered in Paris in 2015 to draft an agreement aimed at reducing global GHG emissions, possibly reaching net-zero in the second half of the 21st century. This international treaty, which goes by the name of the Paris Agreement, defines a long-term goal to limit the rise in global average temperature to well below 2 °C compared to pre-industrial levels, and preferably to 1.5 °C [3]. Unfortunately, despite the efforts made in the last years, recent analyses have shown how the policies currently in place around the world are not going to be enough to set us on the 1.5 °C or 2 °C pathway. Figure 1.2 shows data from the Climate Action Tracker, an independent scientific analysis that

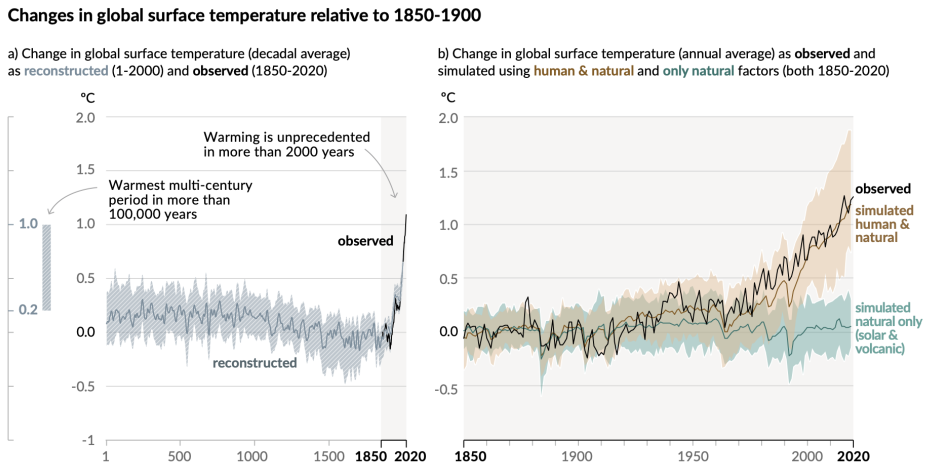


Figure 1.1: Impact of human activities on global surface temperatures, according to the Intergovernmental Panel on Climate Change

studies climate change mitigation pledges and assesses whether countries are on track to meeting them [4]. From the figure, it appears clear that current policies and government pledges are not going to be sufficient and that a very steep reduction in global GHG emissions will be necessary over the next decades in order to reach the 1.5 °C or 2 °C goals.

Global GHG emissions have been steadily increasing during the last century, driven by substantial growth of the world's population and energy consumption. In parallel with these trends of population and emissions growth, urbanization rates have also seen a substantial increase. The latest UN projections show that, by 2050, 68 per cent of the world's population will be living in urban areas. This would represent an increase in the global urban population of almost 2.5 billion dwellers in the course of the next 30 years [5]. One of the consequences of this process is that the world's economic, societal and environmental challenges of the next century will have to be tackled mainly within cities and urban areas.

All of this opens vast challenges related to the management of the built environment. As cities expand and have to face rapid adaptation to new situations, the task of providing reliable, clean, safe and affordable energy to everyone is not of simple solution [6]. In this framework, the role played by buildings is of crucial importance. Figure 1.3 shows a breakdown of the main sources of CO_2 emissions for the year 2019 according to data from the Global Alliance for Buildings and Construction [7]. The building sector accounted for approximately 28% of

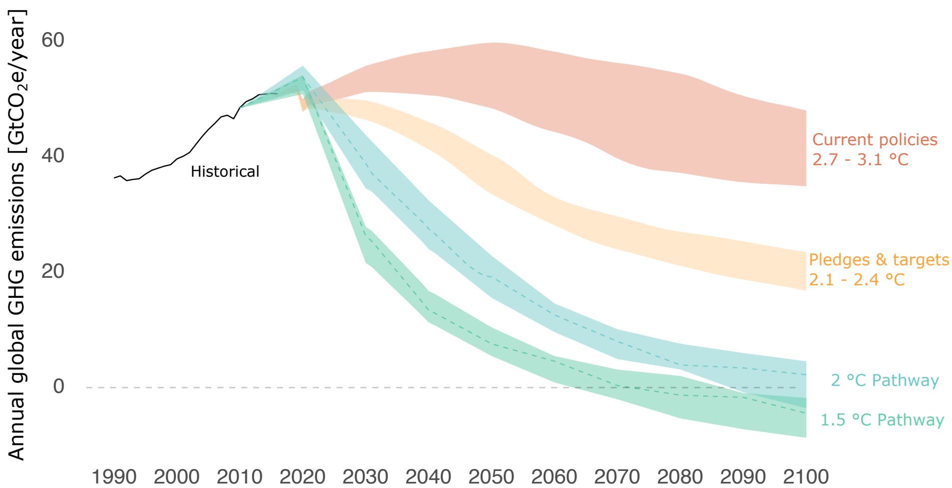


Figure 1.2: Emissions and warming scenarios based on pledges and current policies, according to data from the Climate Action Tracker

global carbon dioxide emissions (about 10 $GtCO_2$), or 38% if we also include the construction sector. This shows that enormous potential to avoid emissions lies in sustainable building design and the implementation of energy conservation strategies.

From Figure 1.3, it appears evident that buildings have a central role in the energy transition. While it is important to focus on sustainable design for new buildings, different studies have highlighted that up to three-quarters of the existing built environment will still be standing in 2050 [8]. This highlights the vast untapped energy savings potential that lies in the correct energy management of buildings and in the implementation of renovation strategies. Efficiency is a topic of great interest in the energy field, and reports from the International Energy Agency (IEA) have shown how the right energy efficiency policies could result in more than 40% of the emissions reductions required to reach the Paris Agreement [9]. In the building sector, the role of efficiency has been highlighted by different reports, with the European Commission estimating that retrofit projects in buildings have the potential of lowering Europe's total CO_2 emissions by 5% [10]. The United States Department of Energy also states that energy efficiency could save up to 741 TWh of electricity until 2035 [11]. At the same time, the integration of renewable energy generation technologies in buildings and the implementation of demand-side management solutions is enabling buildings to cover an active role in the energy

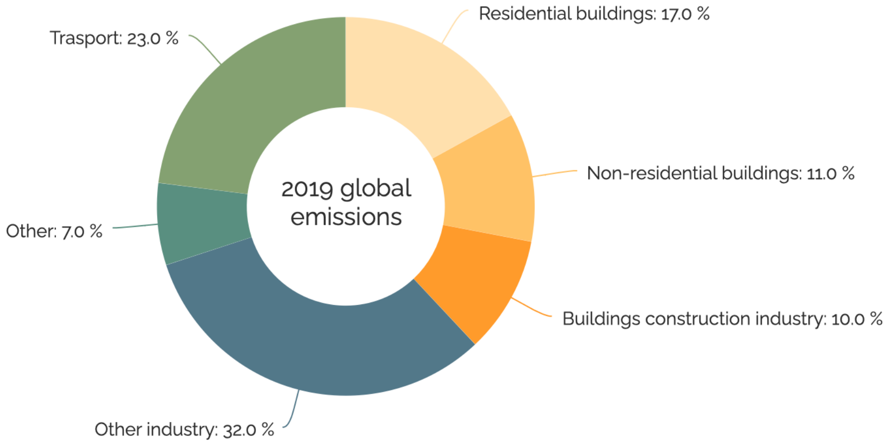


Figure 1.3: Share of global CO_2 emissions by sector for the year 2019, according to data from the Global Alliance for Buildings and Construction

transition. Several industry and academic projects are now investigating the grid services that energy efficiency measures can provide and how efficiency can translate into demand flexibility. Pay for Performance (P4P) [12] and Pay for Load Shape (P4LS) [13] are two concepts that are being investigated in this regard. In these innovative energy efficiency schemes, users are dynamically compensated to modify their electricity usage load shape according to the grid conditions. These new approaches are the result of a need to reduce the overall energy usage intensity of the building stock while also creating a permanent shift of the demand curve that matches the hours of highest renewable energy production [14].

In order to enable improved building energy performance management, informed decision making for energy retrofitting strategies, and dynamic measurement and verification of renovation programs, advanced techniques able to harness the data generated from automated metering infrastructure are required. A recent IEA report showed how digital tools are having a fundamental role in current energy efficiency projects, enabling new data sources and facilitating policy-making by providing quick, cost-effective ways of analyzing energy efficiency potential [15]. In this thesis, we present a range of novel data-driven techniques that can be used to analyze and improve the energy performance of the built environment.

1.2 Objectives and contribution

The objective of this thesis is to investigate how the use of novel data-driven statistical and machine learning techniques can contribute to the improvement of the energy performance of the existing built environment. More specifically, the investigated techniques are aimed at finding ways to quantify energy efficiency savings, analyze energy performance, characterize consumption, and recommend energy retrofitting strategies.

The specific contributions of this thesis are:

- The detailed review of deterministic and data-driven methodologies present in literature to quantify energy efficiency savings and predict retrofitting scenarios in buildings.
- The conceptualization and validation of a novel data-driven technique aimed at accurately quantifying daily energy savings in non-residential buildings using weather and energy consumption data from automated metering infrastructure.
- The conceptualization and validation of a novel data-driven technique to generate hourly baseline energy predictions for non-residential buildings and characterize consumption, with a special focus on uncertainty estimation to de-risk investments.
- The conceptualization of a methodology for recommendation and prioritization of energy efficiency strategies within big data repositories of residential and non-residential buildings.

All the mentioned methodologies have the final goal of producing actionable insights that can be used by energy managers and other stakeholders to reduce the carbon footprint of the building sector and help pave the pathway towards a sustainable energy transition.

1.3 Outline of the Thesis

The present doctoral thesis by compendium of publications contains six chapters, including the introduction and a chapter of conclusions. Figure 1.4 shows the structure of the thesis.

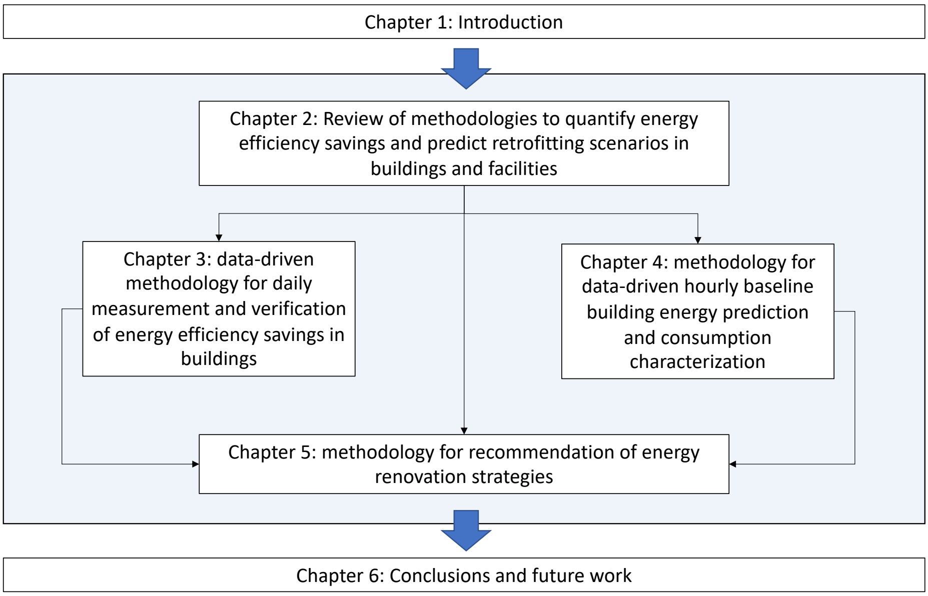


Figure 1.4: Thesis overview flowchart

The present chapter serves as an introduction to the topics that are going to be discussed in this dissertation. The research background is presented, and the objectives and contributions of the thesis are introduced. Chapter 2 presents an extended literature review of state-of-the-art methodologies aimed at quantifying energy savings in buildings and predicting retrofitting scenarios. Both deterministic and data-driven approaches are classified and analyzed in detail. Chapter 3 and 4 introduce two novel data-driven methodologies, developed in the course of this thesis research work, that can be used to characterize and predict building energy consumption at daily and hourly levels. Chapter 2, 3 and 4 correspond to the three publications that have served as a requirement for the presentation of this thesis. The articles were published in three different journals: *Renewable and Sustainable Energy Reviews*, *Applied Energy* and *Energies*. The JCR metrics for the year 2020 of these three journals can be found in Table **??**.

Chapter 2 takes as its springboard the concept that the lack of accurate information regarding the impact of retrofitting actions is one of the barriers preventing a widespread application of energy conservation strategies worldwide. For this reason, state-of-the-art methodologies to quantify energy savings from implemented measures and to predict renovation program scenarios are analyzed in

Table 1.1: JCR journal metrics for Renewable and Sustainable Energy Reviews, Applied Energy and Energies

Journal Name	JIF 2020	Q 2020	Rank Category
Renewable and Sustainable Energy Reviews	14.982	Q1	Green & Sustainable Science & Technology
Applied Energy	9.746	Q1	Energy & Fuels
Energies	3.004	Q3	Energy & Fuels

detail, as well as their strengths and weaknesses. The review work shows that, as more and more building data is collected, with increasing accuracy and granularity, data-driven approaches can become an exceptional tool to perform low-cost large scale analyses within building energy management platforms and repositories.

Chapter 3 introduces a novel data-driven methodology aimed at performing measurement and verification of daily energy efficiency savings in non-residential buildings. The approach is based on a pre-processing phase in which a characterization of building use patterns is performed thanks to the implementation of a clustering technique that allows the extraction of typical consumption profiles. The profiles are then used, in combination with an innovative weather dependence analysis technique, to design a model that provides accurate dynamic estimations of achieved energy savings at daily scale.

In Chapter 4, a Bayesian linear regression approach to predict hourly baseline energy consumption in non-residential buildings is presented. Bayesian inferential techniques are found to be especially fit to solve measurement and verification problems because of their intuitive approach and probabilistic nature. The results obtained show that this novel approach is able to provide accurate high granularity predictions of baseline energy consumption while also being intuitive and explainable. At the same time, a characterization of the consumption of the analyzed buildings is obtained, that provides actionable information to energy managers. The possibility of providing results in the form of probability distributions and of estimating Bayesian credible intervals represents a fundamental feature that allows the implementation of the results obtained in risk assessments for energy efficiency projects and programs.

Chapter 5 presents a concept methodology that employs the techniques introduced in the previous chapters to provide recommendations for renovation strategies within building energy management platforms. Although no real data is analyzed in this section, the structure and the data requirements of the algorithms used to provide recommendations are discussed.

Finally, Chapter 6 presents the conclusions and future work for this doctoral thesis. The practical contributions of this thesis work are summarized and different research lines to improve and further validate the methodologies presented are outlined.

1.4 Projects and publications related

The research carried out for this doctoral thesis has been elaborated in the framework of multiple international projects developed in CIMNE - BEE Group between 2018 and 2021 and led to publications in different journals.

Chapter 2

The review of existing methodologies to quantify energy efficiency savings and predict retrofitting scenarios was carried out in the framework of two projects, called SHERPA and SENSEI, financed by the European Commission through the Interreg MED program and the Horizon 2020 research and innovation program (grant agreements 581 and 847066).

This chapter was also published as a journal article: Grillone B., Danov S., Sumper A., Cipriano J., Mor G. A review of deterministic and data-driven methods to quantify energy efficiency savings and to predict retrofitting scenarios in buildings. Renewable and Sustainable Energy Reviews, 2020; 131: 110027, https://doi.org/10.1016/j.rser.2020.110027

Chapter 3

The methodology for estimation of daily energy savings was developed in the framework of the SENSEI project, financed by the European Commission through the Horizon 2020 research and innovation program.

This chapter was also published as a journal article: Grillone B., Mor G., Danov S., Cipriano J., Sumper A. A data-driven methodology for enhanced measurement and verification of energy efficiency savings in commercial buildings. Applied Energy 2021; 301: 117502, https://doi.org/10.1016/j.apenergy.2021.117502

Chapter 4

The Bayesian methodology for building characterization and hourly energy baseline prediction was developed in the framework of the EN-TRACK project, financed by the European Commission through the Horizon 2020 research and innovation program (grant agreement 885395).

This chapter was also published as a journal article: Grillone B., Mor G., Danov S., Cipriano J., Lazzari F., Sumper A. Baseline energy use modeling and characterization in tertiary buildings using an interpretable Bayesian linear regression methodology. Energies 2021; 14, 5556, https://doi.org/10.3390/en14175556

Chapter 5

The recommendation methodology presented in Chapter 5 was designed to be implemented in the building repositories developed in the framework of the BIGG and EN-TRACK projects, financed by the European Commission through the Horizon 2020 research and innovation program (grant agreements 957047 and 885395).

Chapter 2

Review of methods to evaluate the impact of retrofitting strategies in buildings

This chapter was published as a journal article:

Grillone B., Danov S., Sumper A., Cipriano J., Mor G. A review of deterministic and data-driven methods to quantify energy efficiency savings and to predict retrofitting scenarios in buildings. Renewable and Sustainable Energy Reviews, 2020; 131: 110027, https://doi.org/10.1016/j.rser.2020.110027

2.1 Introduction

Low energy performance of the built environment is one of the main barriers to reach the 2030 European energy efficiency targets [16]. One of the most successful ways to address low building energy efficiency is a massive and affordable implementation of energy renovation strategies [17, 18]. However, at present, there are still several barriers hindering the adoption of procedures and technologies that improve energy efficiency, and limiting the investments in this field. Tuominen et al. [19] found a low impact of renovations on property prices, lack of trusted information, and small prioritization for energy performance improvements, to be frequently cited as the main barriers, in the case of privately owned residential buildings. On the other hand, Kontokosta [20] identified information asymmetry between project partners, uncertainty over expected savings, and shortage of expertise in energy technologies, as the main obstacles in the retrofitting decision making process for commercial office buildings. In the latter case, the author also highlights that these issues have been worsened by case-study oriented approaches, many times because of lack of extensive data and comprehensive pre/post analyses of load profiles following an energy efficiency measure (EEM) implementation.

For commercial and public buildings, applied EEMs can have a significant impact, but the evaluation of this impact with certainty and reliability is no easy task. At the same time, no consolidated framework exists to evaluate ex-ante the effect of different energy retrofitting strategies over buildings. Several techniques to find the most cost-efficient set of measures for a particular building have been developed [21], but scaling up such methods proves to be a major technical challenge, since the effectiveness of retrofitting actions depends on many parameters and this is a clear constraint for any evaluation method.

The objective of this review chapter is to establish the state of knowledge related with the modeling-based approaches used to support the planning and evaluation of building energy retrofitting strategies. More specifically, the chapter aims at reviewing methods, as well as tools, to:

- determine the energy savings obtained through an energy retrofitting program (commonly referred to as measurement and verification).
- support the process of identification of the most appropriate energy renovation action according to the specific features of the analyzed building (in this chapter referred to as prediction and recommendation).

Although few reviews already exist, partially covering the topics addressed in this chapter, to the best of the authors' knowledge no published review provides an in-depth and comprehensive analysis such as the one presented here. In no other review work the measurement and verification, and the prediction and recommendation processes are analyzed together and in a structured way as in the present review. The details of this analysis are described in Chapter 2.3. This review work focuses mainly on data-driven methods, although some deterministic and hybrid methods are also analyzed. The reason for this is that, in the last years, a surge in the number of smart energy monitoring devices has significantly increased the amount of building energy performance data available. This made possible the setting up of many publicly available databases containing energy consumption data and building characteristics of hundreds of thousands of buildings. Data-driven methods are hence becoming of increasing interest, as they are able to harness such huge amount of information for both evaluating the applied energy retrofitting measures and predicting the energy savings potential of new EEMs [22]. Moreover, traditional deterministic methods not based on data have to face an important issue related with their scalability, since the results obtained are usually only valid for the specific building under analysis. This means that using these methods to develop large scale retrofitting strategies can be a major challenge [23]. It's also important to point out that data-driven techniques are being already widely employed in building energy efficiency, and several interesting applications are arising, such as control optimization in demand response, efficiency improvement of HVAC systems, energy efficient operation of different types of buildings, and more [24–26].

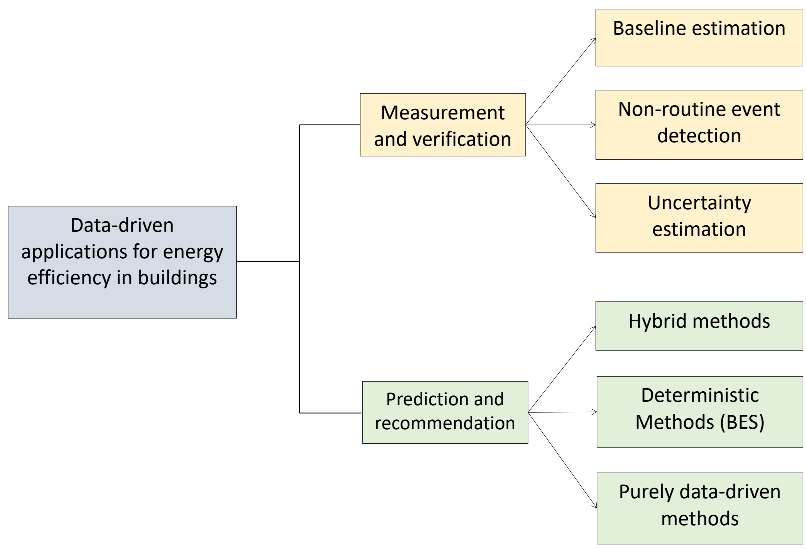


Figure 2.1: Illustration of chapter structure

The chapter is organized as follows. Section 2.2 introduces the reader to different key concepts and how they are used in the context of this review work: the measurement and verification process, the prediction and recommendation process, and the distinction between data-driven and deterministic models. In Section 2.3, a concise overview of previous studies focused on building energy consumption modeling and forecasting techniques is provided. In Section 2.4, a review of existing M&V protocols, as well as data-driven energy baseline estimation methods is presented. State-of-the-art techniques for non-routine event detection and uncertainty estimation are also reviewed in that section. Section 2.5 includes a detailed review of methods to predict the effect of energy efficiency measures and to plan energy retrofitting strategies. Finally, in Sections 2.6 and 2.7 the discussion and conclusions of this review work are outlined. The structure of the chapter is also illustrated in Fig. 2.1, where the two main processes reviewed are highlighted, together with the different applications studied in each case.

2.2 Background

In this section, some concepts which can help to better understand the full content of the review, are introduced, namely: the measurement and verification process, the prediction and recommendation process, and the difference between data-driven methods and deterministic methods.

2.2.1 The measurement and verification process

Measurement and verification (M&V) is the process of using measurements to accurately estimate real savings generated in a facility thanks to the implementation of an energy management strategy [27].

Baseline modeling

Since savings can't be directly measured, as they represent the absence of energy usage, they are determined by comparing measured energy consumption before and after the implementation of a retrofit measure, considering the relevant adjustments for changes in conditions. In order to carry out a comparison between the energy usage before and after the EEM application, a model of the consumption prior to the implementation of the measures needs to be developed. This is model is called the baseline energy model. The baseline model can be defined as the energy characterization of the starting situation and has a fundamental role in the determination of energy savings. In fact, the baseline model allows to isolate the effects of a retrofit intervention from the effects of other parameters that can simultaneously affect the energy consumption, therefore reducing the uncertainty with which savings are estimated. In this chapter, the most common data-driven methods used to develop baseline models are reviewed.

Advanced measurement and verification (M&V 2.0)

In recent years, M&V has been transitioning to a new state, known in the field as “advanced measurement and verification" (or M&V 2.0). This new form of M&V is a result of the breakthroughs in advanced metering infrastructure systems and automated analytics techniques. In M&V 2.0, high granularity datasets with increased sampling frequency, volume, and resolution, are analyzed, in order to perform an estimation of energy efficiency savings which is almost in real-time [28]. This is enabling M&V to advance from a static and cumbersome process to a more dynamic one, that translates into hourly energy insights, maximized savings and great benefit for all the parts involved in the energy retrofitting programs [29]. One of the main drivers of M&V 2.0 is the development of accurate baseline models for real-time savings estimation, through the application of advanced statistical and machine learning techniques. The new features of M&V 2.0 are not only limited to savings evaluation, in fact, most of the advanced M&V tools currently on the market also provide a range of different services, such as analysis and visualization of energy monitoring data, system-level fault detection and diagnostics, and building energy benchmarking [30].

2.2.2 The prediction and recommendation process

The term *prediction* refers to a group of techniques used to predict the effect of an hypothetical EEM application on an individual building or facility. The prediction results are then used to *recommend* the application of specific EEMs over others, and to plan optimal energy retrofitting scenarios. Thanks to prediction and recommendation techniques, it’s possible to answer many different questions, such as: “What is the return on investment for a specific EEM?", “Which EEM would perform best in the selected building, given its characteristics?", “Which low capital cost measures can be applied to increase the energy performance of the selected building?", “Of all the buildings belonging to the considered stock, which ones would benefit the most from an energy renovation program?", “Which EEM would yield the highest energy savings, in a 30 years time span?" etc. All these questions are commonly answered by an engineer, after performing a building energy audit, although the results obtained with the audit can be very uncertain. In Section 2.5, an overview is provided of practical data-driven and deterministic methods to predict EEM impact and plan energy retrofitting strategies for an individual facility or a group of buildings.

2.2.3 Data-driven models and deterministic models

Having clear the goals of the two main processes that are going to be studied in this review, let's now define the two categories of methods under analysis: data-driven models and deterministic models.

Data-driven models are statistical models that find relationships between state variables of the analyzed system (inputs and outputs) without explicit or detailed knowledge of its physical behaviour. In the case of models built for M&V, for example, typical input variables can be external air temperature, wind speed and direction, solar irradiance, building occupancy rate, while typical output variables can be the total electrical or thermal load of the building. Depending on the level of physical significance of the parameters used, these models are usually referred to as *grey-box* or *black-box* models.

The other class of methods reviewed in this work are deterministic methods: detailed building energy simulation models based on the differential equations of the energy transfer flows occurring in the control volumes (rooms or spaces) of the buildings. These physics-based models are usually referred to as *white-box* models.

While for the measurement and verification process, the methods reviewed are exclusively data-driven, in the prediction and recommendation section, both data-driven and deterministic models are analyzed, as well as “hybrid" models, in various which data-driven techniques are used to analyse results obtained with deterministic methods.

2.3 Existing review studies

This paragraph gives a concise but complete overview of previously published review works regarding the different topics treated in this chapter. To the best of the authors' knowledge, there is no published review that addresses the same topics presented in this chapter, that is: an up-to-date and detailed analysis of data-driven and deterministic methodologies used to verify the effect of EEMs in buildings and to predict the impact of future energy retrofitting strategies. The existing data-driven and machine learning techniques used to model and forecast building energy consumption have been thoroughly analyzed in a wide range of reviews published over the last years: [31–38].

Deb et al. [39] divided state-of-the-art forecasting methods in nine different categories and compared them in terms of length of training, data needed, accuracy, and computation time required for the estimation. Wei et al. [40] extended this analysis to other applications, such as energy pattern profile identification, energy-usage mapping, benchmarking of the building stock, and the definition of extensive retrofitting plans. Data-driven techniques related to the development of retrofitting strategies were also studied in the same review(artificial neural networks, genetic algorithms, and clustering techniques).

Harish and Kumar [41] carried out an analysis of different approaches to model and simulate building energy systems and to evaluate the impact of energy retrofitting strategies. Different dynamic modeling techniques were reviewed, including the forward approach (white-box), the data-driven approach (black-box) and the hybrid grey-box approach. The different methods were then classified according to the model type, the parameters used, the simulation period and, the method of validating the results. A list of building energy simulation software, together with their strengths and limitations is also presented in that paper.

Lee et al. [42] reviewed retrofitting analysis toolkits for commercial buildings, classifying them in 3 main categories: toolkits using data-driven methods, toolkits using normative calculations, and toolkits using physics-based energy models. From the analysis, it appears that there is still room for improvement of these methods, especially regarding: (i) mitigation of the high degree of uncertainty associated with these tools, (ii) interoperability between the different tools, (iii) incorporation of human behaviour in the models, (iv) extension of output parameters. An overview of the current state of advanced measurement and verification tools was also provided by Granderson and Fernandes [30]. The authors reviewed sixteen different commercially available tools and classified them according to various criteria: the standard protocol employed, the type of baseline models used, the input data granularity required, the possibility to provide uncertainty estimates, and more. Granderson et al. [43] also compared the accuracy of ten different baseline energy use models for automated measurement and verification of energy savings. The techniques were tested on 537 commercial buildings in the US using training periods of different lengths and without any non-routine adjustment. Two different error metrics: normalized mean bias error (NMBE) and coefficient of variation of the root mean squared error (CV(RMSE)) were calculated and compared, showing similar performances for the ten models. Results of this analysis showed that data-driven statistical techniques are better candidates for scaling up the adoption of whole-building energy savings evaluations using advanced metering infrastructure. In a subsequent publication [44], the same authors applied one of the ten methods (the *time of the week and temperature* baseline model) on a set of 84 buildings, in an attempt to test the applicability of these M&V approaches on a larger scale. It was found that 70% of the buildings of the data set were well fit to be analyzed with the automated approach, and in 80% of the cases savings and uncertainties were quantified to levels above the minimum acceptable thresholds defined by the ASHRAE Guideline 14 [45], a standard protocol used for M&V.

Although the presented review works are of great importance, there is still a shortage of studies covering specifically, and in detail, the processes of measurement and verification of energy savings, and of energy retrofitting planning. Practitioners looking at different options for these two processes, will find in this review a thorough, as well as detailed, overview of the different methods that can be used. Guidance is also provided to determine which method could work best depending on the specific case under analysis. At the same time, it's important to highlight how this review work is mainly focused on data-driven approaches. Considering the growing attention that statistical and machine learning techniques are now receiving in the field of building energy performance analysis, such a study appears essential to identify the research gaps and to highlight future research lines.

2.4 Measurement & Verification: review of methods and data-driven applications

This section aims at reviewing the most popular M&V methods currently in use, with special focus on the data-driven techniques used to estimate baseline energy models. In the first part of the section, four frequently employed M&V protocols are introduced. Following, a review of state-of-the-art data-driven techniques to develop baseline energy models and estimate retrofit savings is presented. The last paragraphs of the section present a review of data-driven approaches to the problems of non-routine event detection and savings uncertainty estimation.

2.4.1 Measurement & Verification protocols

M&V is an evolving science and various methods and best practices were drawn up and documented in different guidelines. Attempts have been made to create a unique standard for the M&V process, but depending on the analyzed facility's geographical location, principal use (residential, commercial, industrial, etc.), and type of metering data available, practitioners still employ different protocols. The optimal degree of standardization that will ultimately be required for advanced M&V is an open issue and currently under discussion among stakeholder groups [30].

International Performance Measurement and Verification Protocol (IPMVP)

The International Performance Measurement and Verification Protocol [27], proposed by Efficiency Valuation Organization (EVO), defines standard terms and suggests best practices to quantify energy savings following the application of one or more energy efficiency measures. According to this protocol, four different options are available to determine energy efficiency savings:

- Option A: Partially Measured Retrofit Isolation. This option involves the use of measurement instruments to monitor the consumption of the equipment affected by the applied EEM, isolated from the energy usage of the rest of the building. In this option, only partial measurement is used, meaning that some parameter(s) are estimated rather than measured.

- Option B: Retrofit Isolation. This case is equivalent to option A, with the exception that no estimations are allowed and full measurement of all the relevant parameters is required.

- Option C: Whole Building. In this approach, utility meters are used to evaluate the energy performance of the whole building. Option C determines the total savings of all implemented EEMs and is only applicable in projects where savings are expected to have a substantial impact, making them distinguishable from energy variations unrelated to the applied measures.

- Option D: Calibrated Simulation. This option involves using building energy modeling software that allows the prediction of energy consumption in different scenarios. The models used for this scope are first calibrated, making sure that the predicted energy load of the building matches the real (metered) data.

ASHRAE Guideline 14

The ASHRAE Guideline 14 for measurement of Energy, Demand and Water Savings [45], published by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), also specifies three different approaches to determine energy savings:

- Retrofit Isolation Approach, similar to IPMVP option B
- Whole Facility Approach, similar to IPMVP option C
- Whole Building Calibrated Simulation Approach, similar to IPMVP option D

Furthermore, the ASHRAE guideline provides different metrics to evaluate the validity of the applied models, such as thresholds for net determination bias or the maximum acceptable uncertainty of the estimated savings.

DOE Uniform Methods Project

The US Department of Energy (DOE), is also building a set of protocols to assess savings due to energy renovation programs. These protocols, joined together under the name 'Uniform Methods Project' [46], provide a simple and clear method to determine energy savings for residential, industrial, and commercial buildings. The protocols are based on IPMVP, but supplementary practices are included, that can be used to aggregate savings from single retrofitting actions and assess program-wide effects.

CalTRACK

CalTRACK [47] is a protocol that was born from the efforts of the California Energy Commission and the California Public Utilities Commission to have a standardized protocol for the evaluation of energy savings in the residential sector. CalTRACK specifies a set of methods to measure and report changes in the energy consumption of a building following the application of an EEM. These methods have the goal of estimating the energy that would have been consumed in the building if the intervention had not taken place. The techniques implemented have been empirically tested by a technical team with several different stakeholders and developed under an open-source license model. The data required to apply the CalTRACK methods includes one full year of consumption data before the EEM application, local weather data, and the date of implementation of the measure.

2.4.2 Data-driven baseline estimation methods

Several baseline energy modeling approaches, using both monthly billing and interval meter data, are presented in the next paragraphs. The reviewed methods are classified into statistical learning, machine learning, and Bayesian techniques, Fig. 2.2 shows an overview of how this section is structured, and Table 2.1 summarizes the characteristics of all the models analyzed.

Statistical learning techniques

Statistical learning is a branch of data-driven modeling that is based on building a statistical model by inferring relationships between different variables in the analyzed dataset. This model is then used to make predictions on other datasets supposed to be similar to the one used to build the model.

Linear and nonlinear regression

Regression analysis has been the first implemented statistical method for the evaluation of energy savings in buildings. Its origins can be traced back to the development of the PRInceton Scorekeeping Method (PRISM) [48], a statistical procedure formulated to include weather normalization in the estimation (scorekeeping) of energy savings. This model is obtained by applying a regression technique that takes into account different variables frequently having an impact on energy usage, such as occupancy, climate, and equipment operation. Common variables chosen for the regression can be: average outdoor temperature, relative humidity, cooling degree days (CDD), heating degree days (HDD), building occupancy and building working days.

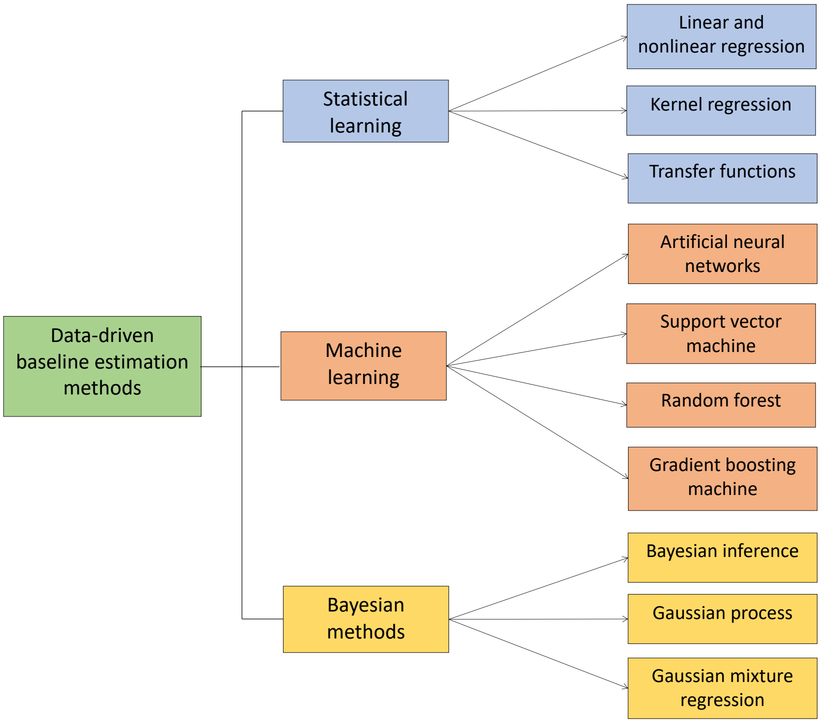


Figure 2.2: Load profiles identified for the office building

Mathieu et al. [49] used linear regression to estimate building energy baselines using high granularity (15-min-interval) consumption data. The model proposed includes an indicator variable that marks the hour of the week and a piecewise-linear temperature regressor having fixed change points. In addition, two different regression models are fit for when the building is considered occupied or unoccupied. This method has been shown to be highly accurate [43] and has been used as a benchmark model in several recent publications regarding measurement and verification methods [50][51][52].

Mohd et al. [53] also tested a linear regression approach to evaluate the effect of an EEM over the HVAC system in an office complex in Malaysia. Both single variable and multivariate linear regressions were fitted, using monthly billing data, temperature readings, and occupancy details. A similar approach was followed by Wang et al. [54] , who tested different linear and nonlinear regression models to assess the energy savings caused by a mechanical system retrofitting in a healthcare facility in Dallas, Texas. The models were fitted with electricity and gas monthly billing data and using average outdoor temperature and degree-day as independent variables. The regression model approach was also tested in the industrial sector: Kissock and Eger [55] built a baseline energy model with multivariable piece-wise linear regression, that was used to disaggregate savings in an industrial facility. The facility's consumption was supposed linearly dependent on its production and on the outdoor air temperature.

Regression analysis is appealing for its simplicity and the possibility of applying it even when low resolution data is available. On the other hand, the linear approach can sometimes be too simple to capture complex relationships between variables.

Kernel regression

Kernel regression belongs to a special class of regression models, called time-varying coefficient models, where the regressors are not considered constant, but dynamically changing over time. The use of kernel regression to estimate building energy consumption baselines was first proposed by Brown et al. [56], with the goal of improving the predictive accuracy of standard linear regression models. The idea behind kernel regression is that the regressors are not estimated using the whole historical dataset. Instead, the regressors are evaluated for each timestep, by estimating a weighted average of all the timesteps with the nearest values of the regression parameters (e.g. weather conditions, time of the day, etc.). The main advantages of kernel regression are an increased estimation accuracy, compared to standard linear regression, and the ability to provide robust and reasonable results even in case of small training sets. On the other hand, since the coefficients are evaluated for a rolling time window and not considering the whole timeseries dataset, when making predictions for longer time frames (e.g. one year or more) the model might not be able characterize the existing seasonal variations and generalize properly .

Transfer functions

Transfer functions have been shown to be capable of accurately estimating the thermal parameters of buildings [57–60], their application for verification of energy savings is now also being tested. The great advantage of transfer functions is the possibility to take into account the building dynamics connected to its thermal inertia. Furthermore, the coefficients of the transfer function model are coupled with the features of the building, thus avoiding the requirement of large amounts of data to obtain reliable results. One of the drawbacks of the method is that the calculations are based on the internal temperature of the building, which is not always known when performing M&V. This baseline estimation methodology was first suggested by Díaz et al. [61], who combined two transfer function models to assess energy efficiency savings in a building of the University of Granada.

Machine learning techniques

The term machine learning (ML) identifies algorithms that make use of statistical models in order learn from data without any specifically programmed instruction. ML algorithms identify patterns in the dataset through iteration and are then able to harness the gained information to make predictions.

Artificial neural networks
Artificial neural networks (ANN) have been applied in several cases to develop baseline energy models [32–35]. The black-box nature of these models makes them very popular, since they can be easily applied to many different problems after just a quick data pre-processing phase. But their simplicity comes at the expense of feature interpretability, making the process of debugging and model improvement considerably more difficult. Low model interpretability and the need for large amounts of training data are the main drawbacks of ANNs. Yalcintas [62] tested ANN models using Levenberg-Marquardt back-propagation to evaluate energy retrofitting savings in two hotel buildings. Adnan et al. [63] used an Hybrid Artificial Neural Network, in combination with Evolutionary Programming, to quantify the savings achieved for a chiller unit in Malaysia, using three different inputs: operating time, refrigerant tonnage and differential temperature. Chang et al. [64] also assessed post retrofit energy savings for an air conditioning system, using ANNs and an energy saving evaluation model based on a parameter named *Refrigeration Operation Energy saving Effect Ratio* (ROEER).

Support vector machine
Support vector machine (SVM) was first applied to estimate building energy baselines by Dong et al. [65]. This machine learning approach is usually preferred when the training data available is small, since it proves to be very powerful in solving problems with non-linear formulations, even with small training datasets. The training time of this technique scales cubically with the size of the dataset [66],

making SVM not ideal when dealing with large-size problems. In [40], an overview of the most recent applications of SVM to building energy consumption prediction is presented.

Random forest

Random forest is an ensemble learning algorithm that constructs several decision trees and then outputs the mean of their prediction, in order to correct for the individual trees' tendency to overfit the data. This powerful methodology has been used for several applications in the domain of building energy prediction. Ahmad et al. [67] used random forests to predict hourly HVAC energy consumption, while Araya et al. [68] proposed their use for fault detection and diagnosis. In the measurement and verification framework, the use of random forests was outlined both in [43] and [30]. Random forests prove to be very accurate in the prediction of building energy usage, although the black-box nature of this algorithm means that the computational time associated with this calculation is quite high, due to the necessity of optimizing the hyper-parameters and performing cross-validation to avoid overfitting.

Gradient boosting machine

Similar to random forest, the gradient boosting machine (GBM) is a powerful machine learning algorithm based on the concept that a “strong learner", having high prediction accuracy, can be obtained by iteratively combining several less complex models, called “weak learners". Touzani et al. [51] used this approach to build an energy consumption baseline model that can be applied for energy savings estimation. The algorithm has four hyper-parameters that were optimized using grid search with 5-fold block cross-validation. The results of the GBM method were compared to the ones obtained with a piecewise linear regression model and a random forest algorithm. This analysis showed that the GBM was able to improve both R^2 prediction accuracy and CV(RMSE) in most of the analyzed cases.

Bayesian methods

As an alternative to the more traditional frequentist approach, several researchers studied the application of the Bayesian paradigm to the measurement and verification process. In Bayesian statistics, a probability model is fit to a dataset, with the goal of obtaining a probability distribution on the model parameters and on other values, like predictions for unobserved data [69]. Then, as new data becomes available, Bayes' theorem is used to update these probability distributions. Among the advantages of Bayesian methods, authors list: the possibility of automatically and exactly quantifying the uncertainty of the models (including different sources of uncertainty, like measurement errors and weather variability), lower sensitivity to outliers, the possibility to have real-time updates of the estimates, and more [70][71].

Bayesian parameters inference

Lindelöf et al. [72] applied Bayesian inference to analyse energy invoices and climate data to estimate the impact of the installation of a model-predictive controller for a heating system in an office building in Switzerland. The approach tries to estimate the probability density function (PDF) of three parameters: the building's heat-loss coefficient, the building's balance temperature, and the stochastic variations of the heating demand, conditioned on the information contained in the utility invoices. The impact of the EEM is assessed by estimating the variations of the heat loss coefficient, through the analysis of a PDF obtained by fitting a Bayesian model to the billing data before and after the EEM application. The Bayesian approach allows to extract high amounts of information from the data and proves to be especially useful in the case of data with monthly granularity. One of the main challenges of this method is that the first probability model, called prior, is often not easy to find and justify, and can be a major source of inaccuracy.

Gaussian process

The application of Gaussian processes (GP) in the M&V process was first proposed by Heo and Zavala [73], with the goal of solving certain limitations of the linear regression method. The Gaussian process approach is non-parametric, since its aim is not finding the parameters of a given function that can best fit the data, but to look for a distribution over the functions $f(x)$ potentially consistent with the observations. GPs can capture complex building energy behaviour, such as nonlinear trends, multivariable interactions and time correlations. At the same time, since GPs belong to the framework of Bayesian statistics, this method allows the savings' uncertainties to be quantified thoroughly. Burkhart et al. [74] suggested the use of Monte Carlo expectation maximization (MCEM) to enhance GP modeling and grant more accurate predictions in case of uncertain input data. Maritz et al. [75] published a guideline to perform M&V using GPs, with special emphasis on the process of kernel selection. The approach is described step by step and then applied to adjust the baseline consumption of an academic facility. A two-stage grid search technique is used to determine the best fit coefficients for the model, which is then applied to calculate savings in two different case studies. One of the main issues associated with this method is its computational and memory complexity, that increases cubically with the size of the training dataset.

Gaussian mixture regression
Srivastav et al. [76] tested the performance of Gaussian mixture regression (GMR) for building baseline energy prediction. The approach was tested on both simulated data from the US Department of Energy and on real data from a commercial building in California, accuracy was compared with a linear regression model. The model showed an estimation accuracy comparable with the multivariate regression approach in both cases, although GMR has the key advantage of allowing the computation of confidence intervals that adapt locally for different circumstances, according to the uncertainty of training data. At the same time, GMR seems to be less sensitive to data sparsity and to regressors correlation. Similarly to GPs, the main challenges of GMR are linked to its long computational time.

2.4.3 Non-routine event detection

The issue of non-routine event detection is a known challenge in the M&V research community and is common to all the previously introduced baseline estimation methods. Non-routine events (NREs) are defined as fluctuations in the energy usage of a building that are not caused by any variation of the explanatory variables of the baseline model, and that are not attributable to the applied measure itself. In order to achieve a precise evaluation of the energy savings, non-routine events must be detected, and accounted for as 'non-routine adjustments' in the estimation of avoided energy use. This process is usually performed manually and, depending on the kind of event, it might require some engineering expertise and knowledge of what the NRE was [44]. This is a considerable issue in automated M&V, as failing to identify such events could lead to an over (or under) estimation of the savings. Recently, Touzani et al. [77] proposed an automated technique, based on statistical change point detection, to identify non-routine events and adjust the savings calculations. The preliminary results of this study, carried out on a set of synthetic data created using energy simulation software EnergyPlus, show a high identification rate for true positives, as well as for false positives, suggesting that the algorithm might still be improved to achieve better results.

2.4.4 Uncertainty estimation

In the M&V context, determining the uncertainty of the obtained results proves to be an issue of major importance. Providing a range of uncertainty, together with the point estimate result, can help establishing the amount of risk associated with a given investment, and support stakeholders in making more informed decisions [78]. Energy savings estimates usually provide results in form of a single point value, the uncertainty can then be interpreted as the interval of doubt around this estimate [52]. According to the IPMVP, when dealing with energy savings, three kinds of quantifiable uncertainties are identified: sampling uncertainty, arising from the fact that in some projects not all the devices can be monitored, hence sampling techniques are used, measurement uncertainty, related to the accuracy of the monitoring infrastructure used to measure the energy consumption, and modeling uncertainty, related to the errors of the baseline models used to estimate the savings.

Reddy and Claridge [79] argued that the uncertainty in the consumption baseline model is the key factor in determining the uncertainty in the measured savings and proposed a formula to estimate it taking into account the CV(RMSE) of the employed statistical model and the relative influence of the EEM on the baseline energy consumption. Koran et al. [80] compared four different methods to calculate the uncertainty of energy efficiency savings estimated using metering data: a formula found in the ASHRAE Guideline 14 [45], an improved version of the ASHRAE formula, an exact formula that can be used in the case of ordinary least square regression, and a bootstrapping technique. All the four methods presented provided reasonable results, although the accuracy of the methods was not evaluated. Subsequently, a work by Touzani et al. [52] compared the accuracy of two different approaches to determine the uncertainty of energy efficiency savings estimations. Four different baseline models were applied: two hourly models and two daily ones. The uncertainty of the model estimates was then analyzed using two methods: the ASHRAE Guideline 14 approach and the k-fold cross-validation approach, a method to assess model accuracy commonly utilized in the machine learning community. The study was carried out on a dataset comprising whole-building electricity consumption data, sampled every 15 minutes, from 69 commercial buildings located in Central

California, Northern California, and Washington DC. The results showed that both methods underestimated the uncertainty of all the four baseline models tested, although the underestimation proved to be stronger for hourly models, probably due to higher autocorrelation of residuals.

Among the few authors to take into account other uncertainties than the modeling one, Olinga et al. [81] proposed a method to optimally allocate budget and effort in M&V while handling both sampling and modeling uncertainties. The results of their case study show a 42 % reduction of the sampling cost and an 11 % reduction of the total M&V cost thanks to the implementation of the proposed approach.

Table 2.1: Characteristics of the analyzed baseline estimation methods for M&V

Model	Advantages	Limitations	Explanatory variables used in the referenced articles	References
Linear and nonlinear regression	Easy to interpret and explain	Sometimes too simple to capture complex relationships	Indoor air temperature, outdoor air temperature, HVAC schedule	[48, 49, 53–55]
Kernel regression	Better fitting than traditional regression	Not ideal to predict long time intervals	Indoor air temperature, outdoor air temperature, HVAC schedule	[56]
Transfer functions	Can model dynamic effects caused by thermal inertia	Requires indoor temperature data	Indoor air temperature, outdoor air temperature, solar radiation, HVAC schedule	[61]
Artificial neural networks	Performs well with non-linear timeseries	Requires large amounts of data, tends to overfit, slow to train	Outdoor air temperature, wind speed and direction, visibility, air pressure, operating time, refrigerant tonnage, running time of the system, refrigerating capacity, power rating of water pumps, differential temperature	[62–64]
Support vector machine	Performs well even with small training datasets	Long computational time for large datasets	Outdoor air temperature, relative humidity, global solar radiation	[40, 65]
Random forest	High predictive accuracy	Hyper-parameters optimization and cross-validation are needed to avoid overfitting	Outdoor air and dew point temperatures, relative humidity, hour of the day, day of the week, number of occupants booked in the hotel, energy consumption of previous hour	[67, 68]
Gradient boosting machine	Higher predictive accuracy than random forest	Hyper-parameters optimization and cross-validation are needed to avoid overfitting	Outdoor air temperature, time of the week, U.S. federal holidays	[51]
Bayesian inference	Accurate uncertainty estimation	Priors are often difficult to justify and can be a major source of inaccuracy	Cooling Degree Days	[72]
Gaussian processes	Able to capture complex (nonlinear) building energy behaviour	Computational and memory complexity	Outdoor air temperature, HVAC supply temperature, occupancy, relative humidity	[73–75]
Gaussian mixture regression	Dynamic confidence intervals	The optimization problem is not trivial to solve, long computation time	Outdoor air temperature, solar radiation, outdoor humidity	[76]

2.5 Prediction and recommendation: review of the methods

In this section, various techniques to predict the effect of energy efficiency measures and to plan energy retrofitting strategies for specific buildings or groups of buildings, are analyzed. As many methods are involving the combination of deterministic models based on simulations and data-driven approaches, this section of the review presents three different categories of methods: deterministic, hybrid, and purely data-driven. Table 2.2 shows an overview of the methods discussed this section; the type of buildings where they were applied and the categories of the analyzed retrofitting measures are also schematized.

2.5.1 Deterministic methods (Building Energy Simulation)

The approaches presented here are based on the application of building energy simulation (BES) to predict the energy performance of buildings in different scenarios.

BES models for retrofit and NZEB comparative analysis

Zangheri et al. [82] used building energy modeling software EnergyPlus [83] to identify which would be the most cost-optimal retrofit combination to reach nearly zero-energy building (NZEB) levels in different building/climate combinations. The study analyzes four different building typologies of 60s-70s and ten different climate areas within the European Union. In order to perform the study, first a "base refurbishment level" was defined, as the minimum possible level of refurbishment to which compare the deeper ones. The base refurbishment level was defined following the assumption that it is not possible to not intervene at all on a building older than 40 years, and includes the rehabilitation of the building envelope, and the substitution of the old heating or cooling systems with comparable equipment. It was found that cost-optimal and NZEB scenarios are characterized by an average increased investment cost, with respect to the base refurbishment level, of 50% and 115 % respectively. The energy efficiency potential of the cost-optimal cases proved to be substantial (between 36 % and 88% primary energy savings), with associated

30 years global costs many times lower than their respective base refurbishment levels.

Similarly, Rysanek and Choudhary [84] used TRNSYS [85], a simulation tool for transient systems, to analyse different energy retrofitting scenarios for a mid-sized office building in Cambridge (UK), while taking into consideration both technical and economic uncertainty. The authors also provide an analysis of how relevant the approach is to real-world contexts. TRNSYS was also used by Valdiserri et al. [86] to evaluate the thermal demand reduction of a tertiary building in Italy, due to an improvement of the thermal envelope and installation of high efficiency windows. An investment cost analysis was also performed, using the Net Present Value (NPV) method.

BES combined with data collected from bills and questionnaires

Another frequently applied method to predict the energy savings of specific energy efficiency measures is to use building energy simulation tools and compare the simulation with the real consumption obtained from metering or energy bills. Suastegui et al. [87] used this method to evaluate potential savings in the residential sector in Mexico due to replacement of oversized HVAC units. A sample of 300 houses was analyzed and questionnaires were used to gather data about the households size and HVAC units capacity. An energy simulation of these buildings was then run using a model based on the Transfer Function Method. The model provides the optimal HVAC sizing for the analyzed households, which is then used to calculate the kWh that could be saved in these households by replacing oversized units.

2.5.2 Hybrid methods

Hybrid methods make use of data-driven techniques to optimize the results obtained with deterministic methods. The reviewed approaches involve the use of different data-driven algorithms to scale up the results obtained to a higher amount of buildings, or to find the optimal solution, within the BES results, according to a given cost function.

BES combined with Artificial Neural Networks

This method, presented by Ascione et al. [88], proposes the use of EnergyPlus simulations and artificial neural networks to predict building energy retrofitting effects and evaluate different renovation scenarios. The approach takes advantage of the reliable and rigorous assessment of EEM impact granted by building energy simulation software and scales the results obtained to a large number of buildings, through the application of artificial neural networks. This combination grants high accuracy of results, while keeping the computational times reasonably low. The method employs two different families of ANNs, one trained with pre-retrofit building simulation data and one with post-retrofit building simulation data, the difference between the outputs is considered as the improvement due to the implemented energy retrofit. The approach was tested on office buildings built in Southern Italy in the period between 1920 and 1970, about 8800 units, representing approximately 13% of the office buildings in Italy. Three independent networks are modeled in the first family (pre-retrofit), each of them having a different output: primary energy demand for heating, primary energy demand for cooling, and percentage of annual discomfort hours. The second category of neural networks (targeting the refurbished building stock), consists of four ANNs with single output: the three networks introduced for the pre-retrofit case, plus a new network included to predict the electricity produced by photo-voltaic panels and used in the building. The accuracy of the ANNs were assessed by analysing regressions and distributions of relative error between the networks' outputs and the results obtained with EnergyPlus models. In both cases (pre and post-retrofit), the accuracy of the models showed to be quite high, with the average absolute value of relative errors ranging between 6.1% and 11%.

Multi-objective and multi-criteria optimization of BES data using Genetic Algorithms

In the framework of decision aid systems for energy retrofitting strategies, two very popular solutions are multi-objective and multi-criteria optimizations. Asadi et al. [89] wrote a detailed review on the topic, explaining also the conceptual distinction between multi criteria and multi objective models: in multi-criteria optimization, the group of possible alternatives is finite and explicitly known a priori, to be evaluated according to multiple criteria, while in multi-objective optimization models, the potential solutions are implicitly determined by the optimization variables and constraints. A very popular technique, frequently used by scientists in both these cases, is the genetic algorithm (GA). Following, different applications of genetic algorithms in the building energy retrofitting field are presented.

Siddharth et al. [90] built an IT tool that uses GAs to create several combinations of building variables correlated with energy consumption. For each of these combinations, the energy consumption of the building is simulated and a nonlinear regression model is fit between the system characteristics and the annual energy demand of the building. In this way, different system configurations are determined, allowing the evaluation of hypothetical energy efficiency measures. The tool was successfully tested in three different climate zones in India and the US. Genetic algorithms and other optimization techniques, such as particle swarm optimization and sequential search, were also applied by Bichiou and Krarti [91] to optimize the selection of building envelopes and HVAC systems for houses in five different US cities, with the goal of minimizing their operating costs. The comparative analysis showed that savings in computational effort could be as high as 70% when using genetic algorithms in place of particle swarm or sequential search.

Ascione et al. [92–94] also used GAs to analyse EnergyPlus simulation data in both multi-objective and multi-criteria analyses. The approach was successfully used first to determine the optimal renewable energy mix in a building and then to identify optimal energy retrofitting strategies in typical hospital and office reference buildings.

Multi objective optimization of BES data using NSGA-II

Chantrelle et al. [95] developed MultiOpt, a multi-criteria tool that uses NSGA-II (a non-dominated sorting genetic algorithm) [96] coupled with environmental databases and assessment software (TRNSYS), to optimize the retrofitting process of buildings across a variety of different objectives. NSGA-II was also used by Delgarm et al. [97], in combination with EnergyPlus, to analyse how different architectural parameters affect the energy consumption of a building in four different climate regions of Iran. The analysis shows that the optimization process could decrease the building's energy consumption by up to 42.2 %.

Multi-objective optimization of BES data using Genetic Algorithms and Artificial Neural Networks

This optimization methodology, that combines different approaches introduced in the previous paragraphs, was used by Magnier and Haghighat [98] to reduce the energy usage while keeping the optimal thermal comfort in a residential building. The approach features the use of NSGA-II to solve the optimization problem and a multilayer feed-forward ANN to reduce the time of computation required by the analysis.

More recently, Asadi et al. [99] used a similar technique to analyze TRNSYS data and identify optimal building energy retrofitting strategies. The set of possible retrofitting actions was summarized in five decision variables introduced as inputs for the ANN: external wall insulation materials, roof insulation materials, window types, solar collector types, HVAC system. The ANN, trained with building simulation results, had four different outputs: total percentage of discomfort hours, and energy demands for space heating, space cooling and sanitary hot water. A multi-objective GA was then applied to analyze the results of the ANN analysis and find the optimal solutions in terms of energy usage, renovation cost, and thermal discomfort hours.

Mixed-Integer Linear Programming

Iturriaga et al. [100] used a Mixed-Integer Linear Programming model to design the energy renovation of an existing building, with the goal of achieving the nearly Zero Energy Building standard. The proposed approach attempts to model the energy demand of the building through a linear model, introducing the EEMs as virtual energy sources that produce, at specific points in time, the energy that would be saved. To calculate the exact demand reduction corresponding to each EEM, dynamic TRNSYS simulations are run. The linear programming approach is then used to optimize the obtained results for the optimal cost case and the Zero Energy Building case. The method was successfully implemented to obtain the system configuration that minimizes the annual net costs for a real building located in the city of Bilbao (Spain).

2.5.3 Data-driven methods

The data-driven methods analyzed in this section have the goal of providing recommendations for building energy retrofit by drawing conclusions based on the analysis of collected data from real use-cases.

User-facing Fallen Rule List using audit data

This method, presented by Marasco and Kotokosta [101], proposes the application of a fallen rule list classifier to how different building would react to different groups of EEMs. The classifier uses binary features obtained from energy audit data for over 1000 buildings in the city of New York and has the goal of providing a tool for decision-makers with the capability of either supporting, or potentially replacing, a complete energy audit. The classifier analyzes the correlation between building specific data and the EEM recommended by energy consultants after performing a building audit. The model was trained on 764 buildings and then tested on 192 buildings, showing a good overall performance for predicting the EEMs of the following categories: cooling system, distribution system, domestic hot water, fuel switching, lighting and motors, representing collectively 62% of EEMs analyzed in this study.

Artificial Neural Networks using audit data

Beccali et al. [102] implemented artificial neural networks to create a decision aid tool able to evaluate energy performance and possible refurbishment strategies for tertiary buildings in Southern Italy. The networks were trained using audit data from 151 non-residential buildings, located in different regions of Southern Italy. The audits collected information about the buildings' geometric and equipment characteristics, as well as data about ten different proposed retrofitting actions. This data was employed to determine the ideal architecture configuration for two ANNs and for their subsequent training. One of the networks estimates the effective energy performance of any building, while the other assesses key economic indicators, allowing users to gain information about possible energy savings, payback time and investment costs per kWh saved.

Clustering techniques

This method is based on the assumption that clustering techniques can help in the development of renovation plans for groups of buildings that respond similarly to the application of EEMs. Geyer et al. [23] tested the application of clustering algorithms using performance-based indicators of the impact of applied measures. The impact of the measure on the considered building is described by a parameter equal to the quotient of the emission reduction caused by the measure, and the investment costs. To assess the impact of an applied EEM, different calculation methods are applied: simplified estimations, monthly sums, dynamic simulations or building energy simulations. Two different clustering methodologies are tested: hierarchical clustering and partitioning k-means clustering. A set of six different retrofit measures, as well as their combination, was simulated. To estimate their effect, simplified calculations using monitored energy consumption and geometric information about the buildings were realized. This method allows the evaluation of how buildings with different characteristics react to applied EEMs and to identify the clusters (groups of buildings) with highest priority for action. Salvalai et al. [103] also investigated the combination of clustering algorithms and building energy simulation, to evaluate optimal renovation strategies for a sample of school buildings in Northern Italy.

Linear regression

Walter and Sohn [104] trained a multivariate linear regression model using data contained in a large building energy database, to estimate energy savings due to the implementation of particular retrofits. The model's input parameters are both categorical and numerical variables, while the response variable is the annual source energy usage intensity (EUI). Through this method, it's possible to analyse the impact of specific building properties and installed systems on the EUI, predict for possible combinations of explanatory variables not included in the database and yield predictions that have clear and well-known statistical properties. The predictors chosen include the majority of the fields in the US Building Performance Database[105], in case of highly correlated fields only one of them is chosen. This method proves to be highly effective, as the data required to perform this type of analysis is generally affordable and easy to obtain, making this approach cheaper and faster than other methods that involve the creation of building energy simulation models.

Genetic algorithm combined with A* graph search

This method was examined by Yi-Kai et al. [106], with the goal of analysing all possible retrofitting actions, and their trade-offs, to identify optimal solutions. Six experienced building renovation stakeholders were interviewed to determine the assessment scores of different renovation actions, as well as the cost information for each action. Based on this data, a two-stage hybrid GAA* algorithm (combination of Genetic Algorithm and the best first (A*) algorithm) was used to test all the possible scenarios and identify the optimal solutions. This approach was compared to two commonly adopted methods: zero-one goal programming (ZOPG) and Genetic Algorithm (GA) proving be better than either of them alone.

Table 2.2: Characteristics of the analyzed prediction and recommendation methods

Method	Type	Test buildings	Measure categories	References
BES	Deterministic	Single family houses, offices, schools, apartment blocks	Building envelope, HVAC systems, domestic hot water, PV and solar system installation, lighting system	[82, 84, 86]
BES + data from bills and questionnaires	Deterministic	Single family houses, apartment blocks	Building envelope, cooling system	[87]
BES + ANN	Hybrid	Offices	Building envelope, solar shading, ventilation system, heating system, cooling system, PV system installation	[88]
BES + GA	Hybrid	Hospitals, offices, single family houses, apartment blocks	Heating, cooling and ventilation systems, building envelope, domestic hot water, PV system installation	[90–94]
BES + NSGA-II	Hybrid	Offices, schools	Building envelope	[95, 97]
BES + GA + ANN	Hybrid	Schools, apartment blocks	Building envelope, HVAC systems, solar collectors installation	[98, 99]
Mixed integer linear programming	Hybrid	Apartment blocks	Building envelope	[100]
Linear regression	Data-driven	Commercial buildings	Building envelope, HVAC systems, lighting system	[104]
ANN with audit data	Data-driven	Schools, sport buildings, libraries, offices	Building envelope, HVAC systems, management system, lighting system, solar thermal installation, water pumping system	[102]
Clustering	Data-driven	Schools, apartment blocks	Building envelope, heating system, PV and solar thermal installation	[23, 103]
FRL with audit data	Data-driven	Apartment blocks, offices, hotels	Building envelope, HVAC systems, DHW systems, PV and solar thermal installation, lighting system, building management system	[101]
GA + A* graph search	Data-driven	Offices	Building envelope, solar shadings, water management, HVAC systems, lighting system, building management system	[106]

2.6 Discussion

In this chapter, two fundamental processes required for the improvement of building energy performance have been studied: the measurement and verification process, and the prediction and recommendation process. After describing their goals and main challenges, different methods found in literature were reviewed. The analysis was focused mainly on data-driven approaches, although for the prediction and recommendation process, deterministic methods were also considered, since their combination with data-driven techniques is becoming of increasing interest.

In the first part of the chapter, different methods for energy baseline estimation were reviewed. For every method, advantages and limitations were examined. The reviewed articles show that more complex methods generally provide more accurate estimations, although the bias-variance trade-off should be always kept in mind: as the models' complexity increases, they can become more accurate, but also more likely to overfit (fail to properly fit additional data, as new observations are added to the dataset) [107]. This said, it was found that different models still have different specific cases where they work best, regardless of their level of complexity. Another interesting insight that emerged from the review is that, when comparing different methods, being able to accurately determine the uncertainty of the results obtained is a very valuable feature. If the main concern of the M&V practitioner is to obtain the best possible estimation of model uncertainty, Bayesian methods seem to be the most optimal choice, as they provide accurate uncertainty estimations without assuming normally distributed errors. On the other hand, statistical learning techniques seem to be favoured when the main concern is the interpretability of the model, and machine learning techniques are most frequently employed when large amounts of data are available and the practitioner is interested in optimizing the model's predictive accuracy.

In the second part of this review, several deterministic and data-driven methods to predict the effect of energy retrofitting actions on buildings were analyzed and presented. Although many of the methods reviewed use deterministic building energy simulations for this task, the analysis of the simulations' results is often performed with data-driven techniques. These approaches, classified as "hybrid" models, appear to be quite popular because of the possibility to combine the accuracy of deterministic methods and the computational efficiency of large scale optimization techniques.

In conclusion, it's important to remark that the comparison of the presented methods is no trivial work, as they were all applied in different use cases, with data of different granularity, and not using the same explanatory variables. This issue is a known problem in the building performance research community and was already pointed out by Miller [108], who proposed and worked on the creation of a public dataset from electricity meters of non-residential buildings, to test and compare prediction algorithms and feature extraction techniques [50][109].

2.7 Conclusions and future work

In order to improve the energy performance of the current building stock, it is essential to implement energy renovation programs. One of the main barriers to the widespread application of such programs is the lack of information regarding the impact of retrofitting actions. It appears clear that quantifying energy savings from implemented measures and determining the uncertainty of the obtained results, are two key steps towards the achievement of a more efficient built environment. The set of calculations performed to collect this data, is often referred to as the measurement and verification process. At the same time, another major task is to be able to find tailored effective renovation strategies for specific buildings or groups of buildings, in the chapter this process was referred to as the prediction and recommendation process.

In this review, the main methods currently utilized for these two processes were studied, with a special focus on data-driven approaches, as they are innovative techniques proving to be more effective and scalable than other traditional methods [31][35]. All of the reviewed techniques have different characteristics and have been applied in some specific cases, their characteristics were discussed in detail and then schematized in Tables 2.1 and 2.2. State-of-the-art methods to identify non-routine events and estimate uncertainty in M&V were also reviewed. Thanks to the additional analysis provided by these methods, it's possible to obtain more accurate estimates of the calculated savings and of their uncertainty.

From the review work, it was also seen that, while the M&V process seems to have a well defined structure, with different established standardization protocols and a range of published scientific articles addressing the topic. The prediction and recommendation process seems to lack such a structure and a considerable standardization effort would be needed in order to establish metrics of comparison and standardized approaches for the different methods currently in use.

Finally, it appears clear that, with more and more data being collected by automated metering infrastructure, data-driven methods are becoming a fundamental tool to plan effective strategies for the energy demand reduction of the existing building stock. For this reason, it is essential that governments and institutions quickly operate to develop policies that can facilitate the collection and analysis of building energy data.

Chapter 3

Data-driven approach for daily Measurement and Verification of energy savings

This chapter was published as a journal article:

Grillone B., Mor G., Danov S., Cipriano J., Sumper A. A data-driven methodology for enhanced measurement and verification of energy efficiency savings in commercial buildings. Applied Energy 2021; 301: 117502, https://doi.org/10.1016/j.apenergy.2021.117502

3.1 Introduction

In 2018, 36 % of final energy use and 39 % of energy and process related CO_2 emissions worldwide were attributed to the building and construction sector, a 2019 report from the IEA showed [110]. In the same document, it was reported that the global emissions of the building sector increased 2 % in 2018, reaching a record amount of 9.7 gigatonnes of carbon dioxide, and marking a 7 % increase from 2010. Different studies have analyzed the low turnover rate of the existing built environment and forecasted that up to three quarters of the existing building stock might still be standing in 2050 [8, 111]. This means that improving the energy efficiency of the existing building stock is a crucial step to lower the energy demand of the building sector, and to reach the climate goals set in Europe [112] and in the rest of the world [3]. In the last years, there's been many studies that tried to asses the energy impact of entire sectors and possible carbon reduction scenarios [113, 114]. On the other hand, it was found that when it comes to the implementation of energy retrofitting measures in individual buildings, one of the greatest challenges is measuring the savings achieved, this process is commonly referred to as the Measurement and Verification process [27].

Measurement and verification (M&V) is defined as the process of using measurements to reliably determine energy savings generated within an individual building or facility by an energy efficiency intervention. Since energy savings cannot be directly measured, as they represent the absence of energy use, they are usually determined by comparing the facility's energy consumption before and after the implementation of a retrofit measure, considering appropriate adjustments for possible changes in conditions. Different methodologies and protocols have been developed for this scope, the main being: the International Performance Measurement and Verification Protocol (IPMVP), and the ASHRAE Guideline 14 [45]. Both these methodologies are based on the use of a baseline energy model to compare the energy consumption before and after the Energy Efficiency Measure (EEM) implementation. The baseline model can be defined as the energy characterization of the starting situation, and its role is fundamental in the assessment of energy savings. In fact, the baseline model has the task of segregating the effects of a retrofit program from the effects of other simultaneous changes that can affect the energy consumption, hence improving the accuracy with which energy savings are estimated.

Recent breakthroughs in advanced metering infrastructure technologies and data analytics techniques have initiated a transition of M&V to a new stage, also known to practitioners as *advanced measurement and verification* (or M&V 2.0). The main characteristic of M&V 2.0 is the possibility of performing real-time energy efficiency savings estimations, thanks to the analysis of datasets having high sampling frequency and resolution [28]. This translates into the possibility of providing dynamic energy insights, maximizing the savings estimation accuracy, and obtaining a detailed characterization of the building's energy usage and performance [29]. In the framework of advanced M&V, great effort is being put into the development of advanced baseline models that can reach high estimation accuracy thanks to the application of state-of-the-art statistical and machine learning techniques. In addition to that, many advanced M&V tools are now offering several other energy services, such as consumption profile data mining, equipment fault detection, and building energy benchmarking [30].

Comprehensive reviews regarding the state-of-the-art of data-driven measurement and verification methodologies were carried out in the last years [43, 115].

While different methods have been identified to have different strengths and weaknesses, the reviews highlighted the following knowledge gaps:

- none of the reviewed methodologies make use of the typical consumption patterns detected in the analyzed facilities as a direct predictor variable,
- all of the reviewed methodologies utilise energy baseline models which are trained only on the data that precedes the EEM implementation. This causes the loss of valuable information regarding how energy consumption fluctuates as a consequence of the variations of outdoor climate variables, which is contained in the data that follows the EEM implementation and which can be extracted (and uncoupled from the effects induced by the measures) by means of statistical methods introduced in the present chapter.

As a consequence, the results provided by most of the reviewed methodologies are affected by a rapid degradation when less than a full year of training data is available. At the same time, most of the existing methodologies only provide an estimation of energy efficiency savings, with no additional information about the energy performance and behaviour of the analyzed facilities. The novel approach proposed in this publication addresses the presented gaps by developing an enhanced data-driven methodology able to:

- consider typical building usage profiles as predictor variables in the energy baseline model, introduced by means of daily consumption patterns detected with a clustering methodology,
- perform a detailed characterization of the building energy usage, harnessing also data from the post-measure application period,
- provide accurate results even when less than a full year of training data is available,
- provide, together with the savings estimation, a range of actionable insights into the energy performance of the analyzed facilities, such as typical load consumption profiles, change-point temperature of the building, and weather dependence analysis.

This represents a significant advance to the state-of-the-art, especially in practical applications, e.g. implementation of Energy Performance Contracts, where accurate short-term monitoring with high granularity could be feasible, but yearround data acquisition is not practical from a business point of view. This feature considerably widens the scope of application of the proposed methodology, also considering that the calculations performed are based on data that can be easily obtained, such as energy consumption and climate data, making the method easy to apply to a wide range of existing buildings.

The viability of the approach was tested in two different case studies: the first was carried out with a set of synthetic data generated using the energy modeling software EnergyPlus [116], while the second was performed using a set of real consumption data from three commercial buildings located in Barcelona (Spain), where different EEMs have been implemented. For case study 1, the operation of three commercial building typologies was simulated in different climate locations, and for each of them two EEMs were introduced, at different points of the time-series, for a total of 54 unique simulation cases. Each of the buildings was simulated with and without the energy efficiency intervention, thereby allowing a comparison between the effective energy savings, simulated with EnergyPlus, and the ones estimated with the data-driven statistical model. Such a comparison made it possible to verify that low statistical model errors are translated into accurate estimations of the energy savings achieved by the EEMs. The proposed approach is also compared to another method based on the time-of-week-and-temperature (TOWT) model [49], a commonly employed method in measurement and verification applications that was the object of different studies to estimate prediction accuracy [43, 44, 109] and uncertainty [52] in measurement and verification applications. The TOWT model is also the reference model of the CalTRACK protocol [117]. For the second case study, data from three existing buildings that implemented EEMs were analyzed. The two case studies have two complementary objectives: with the first it was possible to test the methodology and showcase its features in a controlled environment, where the true savings values are known, since they can be estimated using EnergyPlus deterministic calculations. This provides a concrete value to compare the results to, and a benchmark for model performance. Conversely, the second case study has the goal of demonstrating that the proposed methodology is able to work properly not only with simulated data, but can also provide accurate results with low error when analysing monitoring consumption data from existing buildings.

The present study provides significant research contribution to each of the needs described at the beginning of this section, demonstrating considerable accuracy improvements, compared to the industry benchmark, which are even more evident when the training data used is reduced to less than a full year. At the same time, the results yielded are easily interpretable even for non experts in statistics, providing not only energy savings quantification but also a wide range of actionable insights into the energy performance of the analyzed building.

3.2 Methodology

3.2.1 Methodology overview

The energy savings calculation methodology proposed in this chapter is based on two main concepts: i) the baseline energy use of a commercial building or facility can be represented by a statistical model; ii) when an EEM is implemented in that building, the variation in the energy use caused by this EEM can be represented by adding supplementary terms to the initial baseline model. This last concept defines one of the main differences of this chapter's approach and the M&V methodologies currently employed in industry: traditional methods are based on the theory that it's possible to detect energy behaviour changes by training a model with historical data of the period that precedes the application of the EEM, and then to predict energy consumption on the post EEM application period. In contrast, in the methodology presented in this chapter, the detection of energy behaviour changes does not happen thanks to the choice of a specific training dataset, but through the definition and selection of different additive terms in the model itself. There are two main advantages related to this revised paradigm: the first one is that the whole time-series data can be used to train the baseline model, which enables an improved description of the buildings' energy performance; the second one is that, in case of more than one EEM implementation, there is no need to fit a different model for each measure. The technical details of this procedure are further explained in Sections 3.2.3 and 3.2.4.

With regard to the method to evaluate the impact of EEMs, it is supposed that implemented measures can have a two-fold effect on the energy usage of a building: they either cause a change in the daily load profile, or they modify the way the building consumption increases (or decreases) as a consequence of the variations of outdoor climate variables. Both effects are considered in the assessment of the counterfactual consumption, which is the estimated energy consumption following an intervention, as if the intervention had not taken place. To evaluate the first of these two effects, when estimating the daily counterfactual usage, instead of using the detected load profile for that day, the profile used is the one the building would have had if the measure was not applied, calculated using a prediction algorithm based on time of the year and weather variables. On the other hand, to account for weather related changes of the energy consumption, different additive terms are considered in the baseline model, that have the goal of capturing how the building weather dependence changes after the application of a given measure. The technical details of this process are explained in detail in Sections 3.2.2, 3.2.3, and 3.2.4.

In order to estimate the energy savings due to EEMs, the proposed approach requires the following data:

- energy consumption data (with hourly or sub-hourly granularity),
- outdoor temperature,
- global horizontal radiation (GHI),
- wind speed,
- public holidays calendar,
- date of application of the analyzed EEMs.

Once these data sets are collected, the procedure to estimate the energy savings can be initiated. The methodology structure can be divided in three main phases: data acquisition and preprocessing, model computation, and energy savings quantification. Fig. 3.1 illustrates the process flow diagram of the methodology.

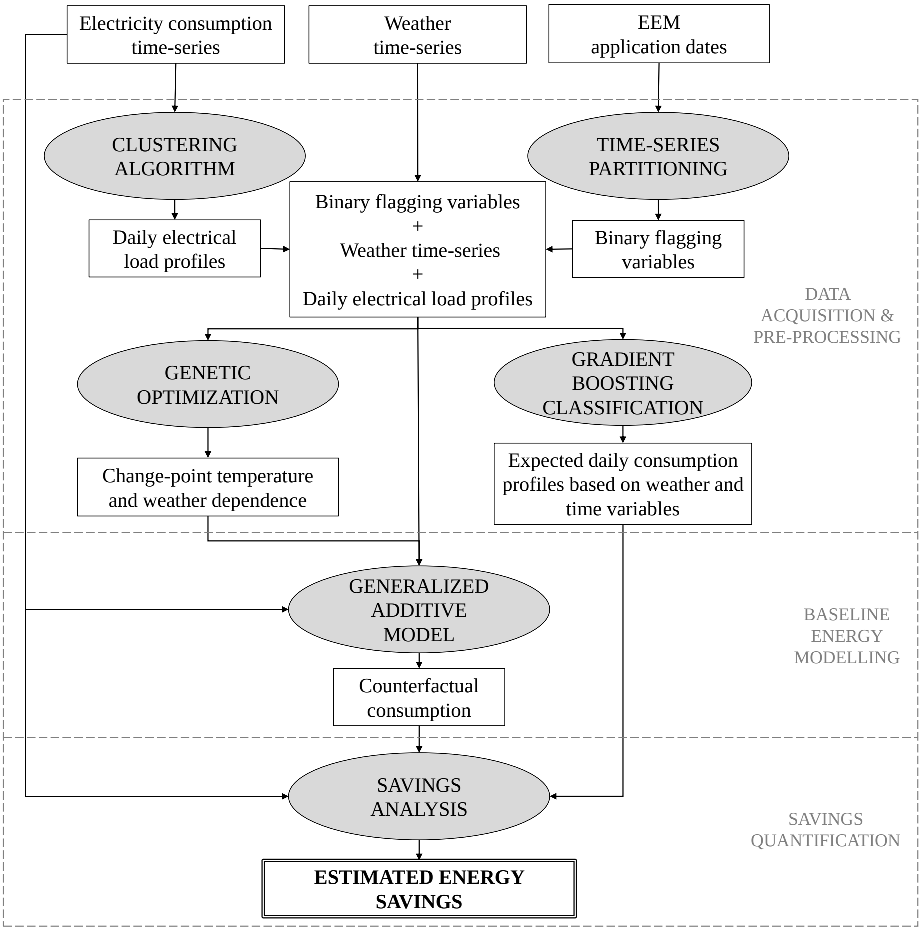


Figure 3.1: Methodology flowchart

3.2.2 Phase 1: data acquisition and preprocessing

In this first phase of the approach, different techniques are used to extract all the variables necessary for the calculation of the baseline model, including the identification of the load profile patterns, and an analysis of the weather dependence of the building.

Time-series partitioning through binary flagging variables

Binary flagging variables are introduced in this model to mark the periods of application of energy efficiency measures. For every Energy Efficiency Measure, EEM_i, that is eligible for evaluation, the sections of the time-series before and after the measure application are marked with a binary variable, $m_{i,j}$, having 0 values for the days before the measure application date, and values of 1 for the days after the measure application date. Thanks to this binary flagging variable, it's possible to separate the sections of the time series where the EEM is applied and the ones where it is not. The mathematical definition of this binary variable is expressed as:

$$m_{i,j} = \begin{cases} 0, & \text{if measure } i \text{ is not applied on day } j \\ 1, & \text{if measure } i \text{ is applied on day } j \end{cases} \tag{3.1}$$

In addition to $m_{i,j}$, another variable m is computed and assigned to each day j of the time-series, indicating the total number of EEMs applied in the building:

$$m_j = \sum_{i=1}^{n} m_{i,j} \tag{3.2}$$

where n is the total number of measures applied on the building. This means that if in a building two different measures were applied, one in January and one in September of the same year, m will have value 0 before January, 1 between January and September, and 2 after september. m and m_i mark which measures are applied (and which are not) on any given day of the time series, these variables allow for an easy identification of the consumption trends during the model training phase, and help in the selection of the additive terms that form the baseline model.

Identification of the daily electrical load profiles

To identify the daily energy usage patterns of the building, a clustering algorithm is applied. The different days of the time-series are separated into clusters of similar daily behaviour, using as input of the algorithm the building's energy consumption values, sampled every hour. Different clustering methods have been tested in literature for this purpose, with k-means clustering [118], self organizing maps (SOM) [119], and Symbolic Aggregate approXimation (SAX) [120], being popular choices. In this research, the clusters are identified using a Gaussian mixture model [121, 122], and the optimal number of clusters ak is chosen according to the Bayesian Information Criterion (BIC) of the model. The result of the clustering technique is a set of k centroid curves for 24 hours electrical load profiles, which define the typical patterns for the analyzed building. To avoid including any potential change of profile generated by an EEM application in the clusters, the profile patterns are identified using only consumption data previous to the first EEM application. Adopting the symbols introduced previously, the clustering is run on all the days that have $m = 0$, while for the days having $m > 0$, the profiles are identified by assigning each day to one of the clusters previously detected. This profile classification is achieved by calculating the cross euclidean distance matrix between the centroids of each of the clusters and the consumption profiles of each day.

Prediction of expected daily profiles based on weather and time variables

Once the daily load profiles have been classified in different clusters, an additional model predicts, for each day of the time-series, which should be the expected load profile on that day. The variables on which this model is based are: day of the week, day of the year, outdoor temperature, global horizontal radiation (GHI), and the electrical load profiles of the 7 days previous to the considered day. More specifically, the goal is to predict, based on these variables, which would have been the load profile of any day after the first EEM application, if no EEMs were applied in the building. To achieve this, the time-series is split in different sections, according to the binary flagging variables previously introduced. Then, for every section of the time-series after the first EEM implementation, the expected profile of the day is estimated using a model trained on data from the preceding time-series section. A concrete example follows, to better illustrate this process: in a building with two applied EEMs, the time series will be split in three parts: the first part, identified by $m = 0$, prior to the first measure application, the second part ($m = 1$), that includes all the days between the implementation of the first measure and the implementation of the second, and the third and last section ($m = 2$), formed by all the days from the application of the second measure until the end of the time-series. Once the split is completed, two different models are run: to predict the daily profiles during section $m = 1$, the model is trained with all the data contained in section $m = 0$, while to predict the profiles of the section that follows the application of the second measure ($m = 2$), the model is trained using the data of section $m = 1$.

To improve the accuracy of this classification, the input variables undergo a preprocessing phase: the day of the week and day of the year variables are encoded as cyclical features using Fourier's transformation [123], while the outdoor temperature and GHI are transformed using spline functions. The classification model used for this task is a gradient boosting machine (GBM), a machine learning algorithm that has been previously applied in other research works related to building energy, and has been shown to have strong predictive performance and flexibility [51, 124]. To execute the model, the XGBoost [125] library was used, in the R programming environment. The results generated here, are then used in the savings quantification phase to detect if applied EEMs caused a change in the load profile of the building and if this lead to energy efficiency savings.

Weather dependence analysis

In the weather dependence analysis, the temperature dependence of the building is identified, more specifically the change-point temperature of the building is calculated, and therefore the days of the year when the consumption is directly dependent on the outdoor temperature values. In order to perform this analysis, the temperature data is pre-processed by applying a first order low-pass filter [126]. This pre-processing is required to take into account that the energy consumption is not directly dependent on the short-term outdoor temperature variation because of the effect of the building's thermal inertia and the thermal resistance. The low-pass filter allows to ignore the fluctuations and to consider only the longer-term trend. The low-pass filtered outdoor temperature is calculated as follows:

$$T_{lp}(t) = \overline{T}(t-1)\alpha + \overline{T}(t)(1-\alpha) \tag{3.3}$$

where $\overline{T}(t)$ and $T_{lp}(t)$ are the hourly average outdoor temperature and the low-pass filtered temperature at hour t. $\overline{T}(t-1)$ is the hourly average outdoor temperature at hour $t-1$, and α is the smoothing factor of the filter (in this analysis a value of $\alpha = 0.1$ was chosen). The expression 'hourly average temperature' refers to the case of having more than one temperature measurement per each hour, in which case the average of these measurements is taken to be the reference temperature for hour t.

Once the low-pass filtered temperature has been calculated, a piece-wise linear model is fitted with these input data. The most significant points of this model are defined by the heating and cooling change-point temperatures, estimated as a change-point temperature T_{cp} plus-minus a hysteresis h which can take on positive or negative values. These two parameters define the outdoor temperatures from which a significant relationship between the building's energy consumption and the outdoor temperature conditions is detected. T_{cp} and h are calculated using a linear regression model and should be understood as a measure of how the daily aggregated electricity consumption varies with the daily average outdoor temperature values.

In order to estimate T_{cp} and h, three possible building operation modes are identified, depending on the hourly average outdoor temperature $\overline{T}$:

$$mode = \begin{cases} heating & \text{if } \overline{T} < (T_{cp} - h) \\ cooling & \text{if } \overline{T} > (T_{cp} + h) \\ equilibrium & \text{if } (T_{cp} - h) \leq \overline{T} \leq (T_{cp} + h) \end{cases} \tag{3.4}$$

From equation (3.4), it's visible that the equilibrium mode is only possible for buildings having $h > 0$, and that for buildings with $h < 0$, the heating and cooling modes can happen at the same time. Figure 3.2 conceptually shows the different consumption scenarios for a hypothetical building, depending on the possible values of h. Positive h will result in the building being characterized by an equilibrium range of temperatures around T_{cp}, where the energy consumption is not affected by the change of outdoor temperature. Negative h means that there is a range of temperatures around T_{cp}, in which the building can have heating and cooling dependence at the same time. Finally, $h = 0$ means that the building has heating dependence for $\overline{T} < T_{cp}$ and cooling dependence for $\overline{T} > T_{cp}$.

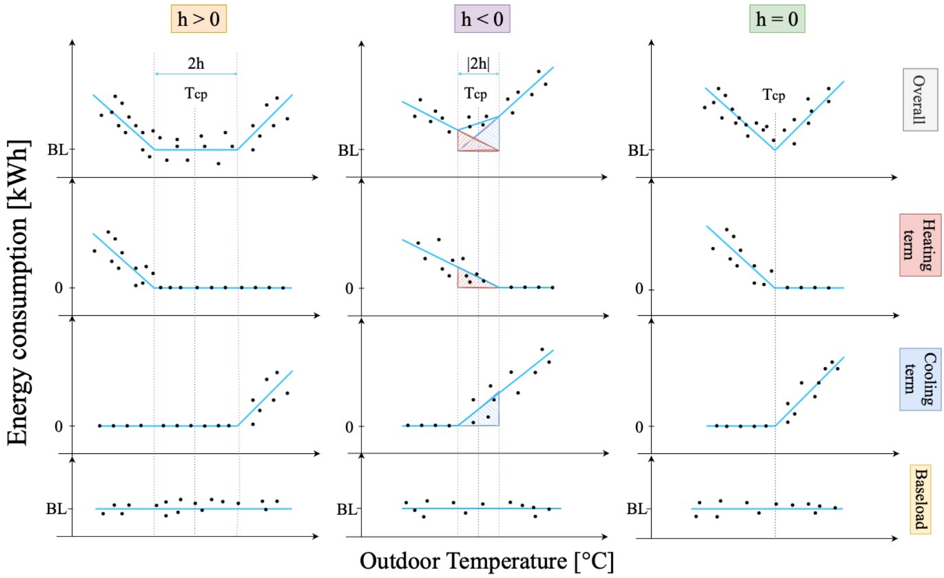


Figure 3.2: Possible consumption scenarios depending on T_{cp} and h values

Once the mode for each day of the time series has been identified, the daily mean dependent temperature T^*_{mode} is calculated, defined as:

$$T^*_{mode} = \begin{cases} (T_{cp} - h) - \overline{T} & \text{if mode} = \text{heating} \\ \overline{T} - (T_{cp} + h) & \text{if mode} = \text{cooling} \\ 0 & \text{if mode} = \text{equilibrium} \end{cases} \tag{3.5}$$

Regardless of the operation mode, a higher T^*_{mode} represents a higher deviation from the change-point temperature of the building, meaning higher energy requirements to keep the desired internal temperature. The T^*_{mode} is then used to fit the step-wise linear model that estimates the temperature dependence of the building energy usage. It is assumed that the ratio by which the building's consumption depends on the temperature T^*_{mode} varies according to the load profile and the operation mode of the day, for this reason the model is constituted of $K \times 3$ different linear terms, where K is the number of clusters identified as described in 3.2.2. The formulation of the step-wise linear temperature dependence model is the following:

$$Y_d = \sum_{mode,k} C_k\, \alpha_k + C_k\, \beta_{mode,k}\, T^*_{mode} \tag{3.6}$$

where Y_d is the daily electricity consumption of the building, α_k is the base load for cluster k, $\beta_{mode,k}$ are the linear coefficients that mark the temperature dependence according to the load profile and mode of the day (i.e. $\beta_{heating,1}$ is the coefficient that marks the temperature dependence of the building when it is in heating mode and the day's load profile is represented by cluster 1), and C_k is a variable that marks the cluster of the analyzed day and is equal to 1 if the cluster of the day is equal to k, or 0 in all the other cases. In this way, only the relevant linear coefficient is used to describe the temperature dependence of the day (i.e. when the load profile of the day corresponds to cluster 1, $C_1 = 1, C_2 = 0, C_3 = 0, ..., C_K = 0$). To identify the linear coefficients $\beta_{mode,k}$, the lasso method (least absolute shrinkage and selection operator) is used. The optimization is initiated with an arbitrary T_{cp} and h, that are then optimized with a genetic algorithm that uses as score the R^2 of the model in Equation (3.6). Note that a single T_{cp} is evaluated for the whole building, regardless of the different load profiles, meaning that effect of the change-point temperature variation between different clusters on the daily aggregated consumption is considered to be negligible. Once T_{cp} and h are optimized, the mode of the building for each day of the time-series is calculated using equation (3.4). The building mode identifies the days in which the energy usage of the building is weather-dependent, since it is considered that the weather variables will only affect the electricity consumption if the building needs active heating or cooling. Once the optimized building mode is identified, modified temperature, GHI, and wind speed vectors are calculated, having non-zero values only on the days where the building is considered weather dependent, these vectors are marked by the superscript 'dep':

$$T_h^{dep} = \begin{cases} (T_{cp} - h) - \overline{T} & \text{if mode} = \text{heating} \\ 0 & \text{if mode} = \text{cooling or mode} = \text{equilibrium} \end{cases} \tag{3.7}$$

$$T_c^{dep} = \begin{cases} \overline{T} - (T_{cp} + h) & \text{if mode} = \text{cooling} \\ 0 & \text{if mode} = \text{heating or mode} = \text{equilibrium} \end{cases} \tag{3.8}$$

$$GHI^{dep} = \begin{cases} \overline{GHI} & \text{if mode} = \text{cooling or mode} = \text{heating} \\ 0 & \text{if mode} = \text{equilibrium} \end{cases} \tag{3.9}$$

$$Ws^{dep} = \begin{cases} \overline{Ws} & \text{if mode} = \text{cooling or mode} = \text{heating} \\ 0 & \text{if mode} = \text{equilibrium} \end{cases} \tag{3.10}$$

where $\overline{GHI}$ and $\overline{Ws}$ represent the average global horizontal radiation and average wind speed of the day. By including these modified vectors T_h^{dep}, T_c^{dep}, GHI^{dep} and Ws^{dep} in the baseline estimation model, it's possible to take the weather variables into account only on the days when the building is directly dependent on them.

Finally, a last set of vectors necessary for baseline estimation is calculated, by multiplying T^{dep}, GHI^{dep}, and Ws^{dep} by the binary flagging variables m_i. The resulting vectors are used to characterize how the weather dependence of the building changed in time, as different EEMs were applied:

$$T_{h,m_i}^{dep} = T_h^{dep} * m_i \tag{3.11}$$

$$T_{c,m_i}^{dep} = T_c^{dep} * m_i \tag{3.12}$$

$$GHI_{m_i}^{dep} = GHI^{dep} * m_i \tag{3.13}$$

$$Ws_{m_i}^{dep} = Ws^{dep} * m_i \tag{3.14}$$

These vectors have a crucial role in the methodology, as they contain the information about the effect of the implemented EEMs on the building energy consumption. It is thanks to these vectors that it's possible to use the whole time-series to train the baseline model, while still excluding the energy reductions caused by the measures when estimating the savings. This process is explained in more detail in the next section, which presents the characteristics of the baseline model.

3.2.3 Phase 2: baseline energy modeling

This phase has the goal of developing a statistical model able to estimate the daily baseline energy demand of the building under analysis. Among the different possible statistical models that could be used for this task, the authors decided to work with generalized additive models (GAM),which is a specific method for supervised learning originally developed by statisticians Trevor Hastie and Robert Tibshirani [127]. GAMs are flexible statistical methods that can be used to identify and characterize nonlinear effects. In the regression setting, a generalized additive model has the form:

$$g(E(Y)) = \beta_0 + f_1(X_1) + f_2(X_2) + ... + f_p(X_p) \tag{3.15}$$

where X_1,X_2,. . . ,X_p represent the predictors, and Y is the outcome. The f_j may be functions with a specified parametric form (polynomial or un-penalized regression spline, for example), or unspecified 'smooth' functions, to be estimated by non-parametric means. This means that the model allows for rather flexible specification of the dependence of the response on the covariates. The GAM developed in the framework of this research aims to represent daily electricity consumption as a function of outdoor temperature, sun altitude, wind speed, daily load profile, and a holiday flagging variable. The model has the following form:

$$\begin{aligned} g(E(E_d)) = & \alpha T_h^{dep} + \beta T_c^{dep} + f(GHI^{dep}) + f(Ws^{dep}) \\ & + \sum_{i=1}^{n} [\alpha_i T_{h,m_i}^{dep} + \beta_i T_{c,m_i}^{dep} + f(GHI_{m_i}^{dep}) + f(Ws_{m_i}^{dep})] \\ & + \sum_{i=0}^{n} m_i(\delta_{h,i} d_h + \delta_{nh,i} d_{nh} + \sum_{k=1}^{K} \gamma_{i,k} C_k) \end{aligned} \tag{3.16}$$

where :

- $E(E_d)$ is the estimator of the daily electricity consumption,
- $\alpha T_h^{dep} + \beta T_c^{dep} + f(GHI^{dep}) + f(Ws^{dep})$ are the linear and smooth terms marking the part of consumption that depends on weather variables, evaluated on both pre and post measure application data,

- $\alpha_i\, T^{dep}_{h,m_i} + \beta_i\, T^{dep}_{c,m_i} + f(GHI^{dep}_{m_i}) + f(Ws^{dep}_{m_i})$ also identify climate-affected consumption, but only during specific sections of the time-series, since the vectors with the subscript m_i, defined in 3.2.2 are non-zero only when measure i is applied,

- n is the total number of energy efficiency measures applied in the building during the considered time-frame,

- $\sum_{k=1}^{K} \gamma_{i,k}\, C_k$ is the term representing the effect of the daily load profile on the consumption, where $\gamma_{i,k}$ is the coefficient that specifies the impact of the profile, C_k is a variable that marks the profile of the analyzed day, being equal to 1 if the profile of the day is profile k, 0 in all the other cases, and K is the total number of daily load profiles identified by the clustering algorithm,

- $\delta_{h,i}\, d_h + \delta_{nh,i}\, d_{nh}$ is the term that marks the effect of calendar holidays on the consumption: d_h is a coefficient equal to 1 if the day is a calendar holiday, and 0 if it's not, d_{nh} the opposite, while $\delta_{h,i}$ and $\delta_{nh,i}$ are the coefficients that specify the impact of the holiday variable depending on the EEM applied in a certain period of the time-series,

- m_i are the EEM binary flagging variables, equal to 1 when measure i is effective, 0 elsewhere (m_0 is considered 1 when there are no measures applied, 0 elsewhere).

All the model regressors are estimated in one single training phase, but fitting on different periods of data depending on the regressor. For each of the weather variables, two typologies of regressors are defined: baseline regressors, which are estimated using the whole time-series; and measure effect regressors, that represent the variation from the baseline and are estimated using the time periods where the measure is applied. E.g. for the GHI case, there is a baseline smooth term that represents the building dependence on solar radiation throughout the whole timeseries: $f(GHI^{dep})$. Then, additional GHI regressors $\sum_{i=0}^{n} f(GHI^{dep}_{m_i})$ allow to estimate the variations from the baseline introduced by the measures. Thanks to this fitting technique, additional data from the post EEM application date can be included in the evaluation of the baseline dependence coefficients, since the variations generated by the EEM are included in the measure effect set of coefficients. This allows an improved accuracy of the model, even when pre measure implementation time-series data available is short. The methodology used to harness this model for the measurement and verification of energy savings is explained in the next section

3.2.4 Phase 3: Energy savings quantification

In the last phase of the methodology, the baseline model is used to calculate the counterfactual energy consumption, that allows to estimate the energy savings. The accuracy metrics used to evaluate the goodness of fit of the model are also introduced.

EEM savings evaluation

The savings estimation process has different substeps, firstly the whole time-series is divided in $n+1$ sections $(P_0, P_1, \ldots, P_n)$, where n is the total number of measures applied in the building. In order to calculate the savings provided by an individual EEM during period of time P_i, the model computed in Phase 2 is used to predict what would be the energy consumption of the building during that period of time, if the behaviour of the building would still be the one observed in the previous section of the time series $P_{(i-1)}$ (counterfactual). This prediction is realized by applying the GAM model with the model coefficients fitted for section $P_{(i-1)}$, but with the exogenous variables (outdoor temperature, sun altitude, wind speed, etc.) of period P_i. Additionally, to increase the accuracy of the estimation, the daily load profiles introduced in the model (marked by the term C_k in (3.16)) are not the ones identified with the classifier described in 3.2.2. Instead, the profiles predicted by the classification model detailed in 3.2.2 are used. Thanks to this, if on a given day, the load profile changed because of the implemented EEM, the baseline consumption for that day will not be calculated based on the new load pattern, but on the one the building would have had if the measure was not applied. The estimated baseline obtained is then compared with the metered energy consumption during the same period of time P_i. If the measure had a positive effect on the energy usage of the building, the baseline consumption for the considered period will be higher than the metered one, and the energy savings achieved can be obtained by calculating the difference between both time series. The energy savings S_i for period P_i can be described as:

$$S_i = \sum_{j \in P_i} g(E(E_{d_j} | i = i-1, C_k = C_{kpred})) - \sum_{j \in P_i} E_{dm_j} \tag{3.17}$$

where:

- $\sum_{j \in P_i}$ represents the sum over the j days of period P_i being the baseline model a daily model,

- $g(E(E_{d_j}|i = i-1, C_k = C_{kpred}))$ is the estimated baseline consumption on day j, calculated with the model coefficients fit for period $P_{(i-1)}$, and substituting the load profiles identified by the clustering classifier (C_k) with the load profiles predicted by the gradient boosting machine (C_{kpred}),

- E_{dm_j} represents the metered electricity consumption on day j.

Accuracy metrics

Once the energy savings are estimated, different metrics to evaluate the goodness of fit of the model (GOF) are calculated. The GOF of the baseline model is assessed with the coefficient of variation of the root mean square error (CV(RMSE)). It is a metric frequently employed to evaluate the accuracy of energy baseline models:

$$CV(RMSE) = \frac{\sqrt{\frac{1}{n}\sum_i^n (E_i - \hat{E}_i)^2}}{\bar{E}} \times 100 \tag{3.18}$$

where n are the total days of the time-series, E_i is the measured energy consumption value on day i, $\hat{E}_i$ is the predicted energy consumption with the baseline model, on the same day, and $\bar{E}$ is the average daily energy consumption across the whole time-series. In equation (3.18), it can be appreciated that the CV(RMSE) is a normalization of the root mean square error by the mean of the measured energy consumption values $\bar{E}$, therefore, it represents the size of the model error in relation to the average energy consumption values. For the scope of this research work, two additional accuracy metrics are defined, that can be used to calculate how well different models are performing, compared to the ground truth provided by the EnergyPlus simulations. The first of these measures is the normalized percentage difference between the model estimated savings and the simulated savings calculated with EnergyPlus:

$$S_{diff} = \frac{|S_{Eplus} - S_{model}|}{E_{tot}} \times 100 \tag{3.19}$$

where S_{Eplus} are the overall savings calculated with EnergyPlus for the time period under analysis, S_{model} are the savings estimated with the data-driven model, for the same period, and E_{tot} is the total monitored energy consumption in that time-frame. S_{diff} can be seen as the deviation of the model estimated savings from the theoretical savings, in terms of percentage of the period total consumption. The other metric introduced is the baseline CV(RMSE), a parameter similar to the one presented in equation (3.18). The difference lays in the fact that the CV(RMSE) of equation (3.18) is generally calculated on data that precedes any EEM implementation, since, in real world use cases, there is no measured data about how the energy consumption of the building would have been if the measure was not implemented. Conversely, being in possession of the ground truth provided by EnergyPlus, in the setting of the present research work, it's possible to calculate the difference between the baseline consumption obtained with the data-driven models, and the one calculated using the deterministic approach:

$$CV(RMSE)_{BL}[\%] = \frac{\sqrt{\frac{1}{m}\sum_j^m (B_{E+,j} - \hat{B}_j)^2}}{\bar{B}_{E+}} \times 100 \tag{3.20}$$

where m are the total days after the measure implementation, $B_{E+,j}$ is the EnergyPlus baseline consumption value on day j, $\hat{B}_j$ is the model predicted baseline energy consumption, on the same day, and $\bar{B}_{E+}$ is the average daily EnergyPlus baseline consumption. S_{diff} is a quick and intuitive metric that can be used to evaluate the accuracy of different data-driven models in the estimation of overall yearly savings, while the $CV(RMSE)_{BL}$ can be used to compare how such models are performing in the prediction of the daily counterfactual.

3.3 Case study

The enhanced M&V methodology proposed in this chapter was thoroughly tested on two different case studies: one that uses synthetic data generated with simulation software, and one that uses monitoring data from existing buildings. The two case studies are described in detail in this section.

3.3.1 Case study 1: synthetic data

For the first case study, synthetic data was generated with the building energy simulation software EnergyPlus. Three different building typologies were designed, using the 3D design software SketchUp [128] and its plugin Euclid [129]. Each of the three building typologies was simulated in three different geographical locations and six different sets of implemented energy efficiency measures were considered, meaning that 54 unique simulations were carried out, in order to test the performance of the proposed methodology. Additional white noise was also added to the final simulation time-series, with the goal of improving their resemblance to the stochastic behavior of non simulated buildings. For each of the 54 cases, three years of building operation were simulated, with two different EEMs introduced at different dates of the time-series. Three sets of simulations were run for each case: i) first, the whole three years of operation were simulated without any measure applied; ii) then, with one EEM applied; iii) and finally, with both EEMs applied. Thanks to this approach, it was possible to simulate the energy savings associated with each EEM. To estimate the accuracy of the proposed M&V approach, a comparison of the simulated savings and the ones estimated by the data-driven model, was performed. The three years of operation that were simulated are 2016, 2017, and 2018, and the EEMs were considered to be applied each at the end of one year of operation (January 1st 2017, and January 1st 2018). Historical weather data for the selected locations was obtained from the Dark Sky weather web service [130] and used in the simulations. It's important to point out that, while it's true that synthetic data is used in this case study, this is neither an attempt to build a calibrated model, nor to create a simulation-based M&V framework. The methodology presented in this chapter is purely data-driven and does not use any information derived from the simulations' generation process.

The energy consumption time series of one of the simulated building typologies is shown in Figure 3.3. The different colors represent the state of the building in terms of the applied EEMs: for the first year of operation (red) no measure is applied, for the second year (green) the first of the two measures is applied ($m = 1$), while, for the third year (blue), both EEMs are applied ($m = 2$). A quick view of this figure shows an appreciable decreasing trend of the energy consumption, which could be associated with the application of each of the EEMs.

Pilot building typologies

As mentioned before, three different building typologies were analyzed in this case study: an office, a primary care center, and a hospital. The three geographical locations considered were Stockholm (Sweden), Berlin (Germany),

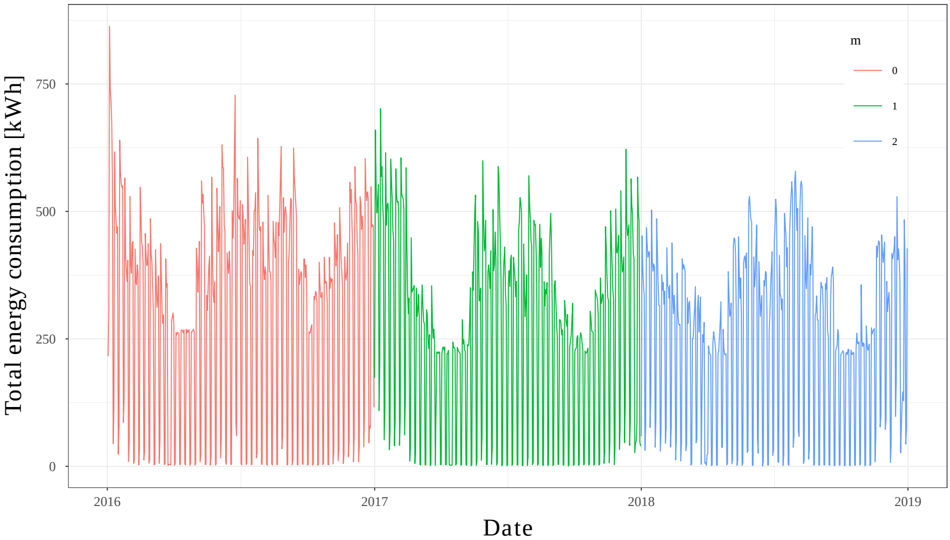


Figure 3.3: Consumption time-series for the office building, with colors marking the implemented EEMs

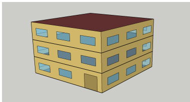
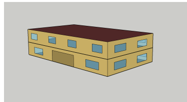
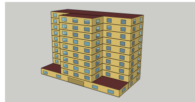
and Girona (Spain). The U-values for walls, windows, and roofs were selected using the information contained in TABULA [131, 132], a WebTool that comprises information about typical U-values of buildings of different age bands across Europe. With regards to the buildings' construction materials and equipment, the buildings were supposed to be built between the 1950s and the 1980s. In all the three building typologies, heat pumps are used for both heating and cooling, with a variable COP ranging between 2.7 and 3.5 depending on the outdoor temperature values. The construction details of the three pilot buildings are presented in Table 3.1.

EEMs details

Four different EEMs were applied in the pilot building typologies. Their details are outlined here:

- Electric equipment efficiency improvement: the peak plug load power density was reduced by an amount ranging between 2 and 6 W/m^2, depending on the building,
- Lighting efficiency improvement: the peak lighting power density was reduced by 4 W/m^2,

Table 3.1: Characteristics of the building typologies used for Case Study 1

Building Block	Pilot 1 - Office	Pilot 2 - Primary care	Pilot 3 - Hospital
Building shape			
Total floor area $[m^2]$	1680	700	9800
Average window-to-wall ratio	30 %	28 %	22 %
Number of floors	3	2	10
Floors to ceiling height $[m]$	3	3	3
Windows			
U-factor $[W/m^2K]$	3	4	3
Solar Heat Gain Coefficient	0.763	0.763	0.763
Visible Transmittance	0.812	0.812	0.812
Infiltration	1 ACH	1 ACH	1 ACH
Ventilation (peak) $[m^3/(s \cdot floor)]$	1.32	1.05	2.94
Internal loads (peak)			
Lighting power density $[W/m^2]$	13	13	8
Plug load power density $[W/m^2]$	3.55	5	18
Occupancy (peak) $[W/(m^2 \cdot floor)]$	4.8	7.2	14.4

- HVAC set-points shift: the heating temperature set-point was shifted from 21 °C to 19 °C, and the cooling one from 23 °C to 25 °C, additionally, the functioning hours were changed in order to reduce the HVAC overall usage,

- Building envelope improvement: the total air infiltration was changed from 1 ACH to 0.3 ACH, as well as the UA values for windows, walls and roof. The values depend on the specific buildings and geographic locations, and were chosen taking into account the research carried out within the TABULA project.

These four measures were then combined into six possible combinations of two measures: one at the end of the first year, and one at the end of the second year of simulation. The six combinations are shown in Table 3.2. Each building typology was simulated in the three previously mentioned geographical locations and with each of the six measure combinations, generating in this way 54 unique simulations.

3.3.2 Case study 2: monitoring data

The second case study is based on real data from two offices and a cultural building belonging to the regional authority of Catalonia (Spain), and situated

Table 3.2: Details of the EEM combinations applied in Case Study 1

Combination	First year measure	Second year measure
1	Equipment efficiency improvement	HVAC rescheduling and setpoints shift
2	HVAC rescheduling and setpoints shift	Envelope improvement
3	HVAC rescheduling and setpoints shift	Lighting efficiency improvement
4	Lighting efficiency improvement	Equipment efficiency improvement
5	Envelope improvement	Lighting efficiency improvement
6	Equipment efficiency improvement	Envelope improvement

in the province of Barcelona. The goal of this case study is to demonstrate that the proposed methodology works well not only on synthetic data coming from a controlled environment simulation, but also on monitoring consumption data coming from real-world buildings. In two of the buildings (that will be named here Building 1, Building 2, and Building 3 due to the confidentiality of the data analyzed), a substitution of all the final elements of the illumination system was realized. While for the third, a pack of different retrofit measures was realized, including the replacement of all the building lighting devices, the installation of an intelligent building energy management system, and the installation of solar screens. A summary of the information available for the three buildings, and the respective implemented measures, is available in Table 3.3.

Table 3.3: Characteristics of the three pilot buildings analyzed in Case Study 2

Building	Building 1 - Office	Building 2 - Office	Building 3 - Cultural
General information			
Total floor area [m^2]	4000+	5000+	15000+
Date start monitoring	01/08/2018	01/01/2019	01/12/2017
Date end monitoring	31/12/2019	31/12/2019	31/12/2019
EEMs applied			
Measure details	Lighting efficiency improvement	Lighting efficiency improvement	Lighting efficiency + BEMS + solar screens
EEM implementation date	01/06/2019	03/08/2019	01/11/2018

The electricity consumption time-series of Building 3 is presented in Figure 3.4, with the red and blue sections representing the consumption before and after the implementation of the measures. It can also be seen that one whole month of training data is missing from the consumption time-series, due to a technical malfunctioning of the monitoring system.

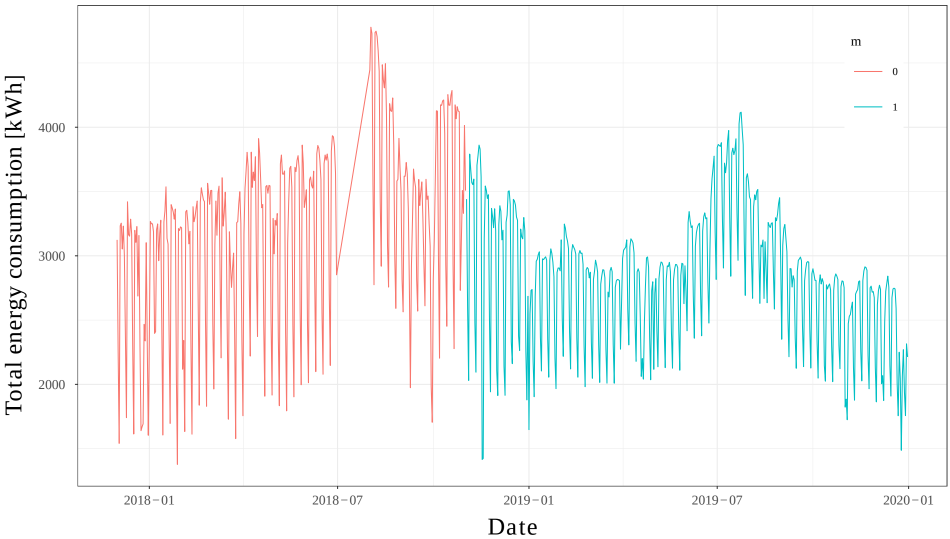


Figure 3.4: Consumption time-series for Building 3 of Case Study 2, with colors marking the implemented EEMs

3.4 Results

This section is dedicated to the presentation of the results obtained in the two case studies previously introduced.

3.4.1 Case study 1

The results for the first case study are divided into two sections: first a single simulated building is analyzed in detail, with a comprehensive analysis of all the intermediate results provided by the methodology. Then, the aggregated results obtained for all the 54 analyzed buildings are presented and discussed.

Single building analysis

This section has the goal of showcasing the details of the different steps of the methodology, and is related to the analysis of the office building, simulated with Berlin weather data, and having equipment efficiency improvement as first implemented measure and HVAC rescheduling and set-points shift as second implemented measure. This building will be referred to as the *reference building* for the remainder of this section.

The first step of the methodology is the identification of the typical daily consumption profiles. In Figure 3.5, the 8 profiles identified for the first year of operation of the office building (before any measure implementation) are displayed: the red lines represent the profile centroids, while the black lines represent the load profiles of days belonging to that cluster.

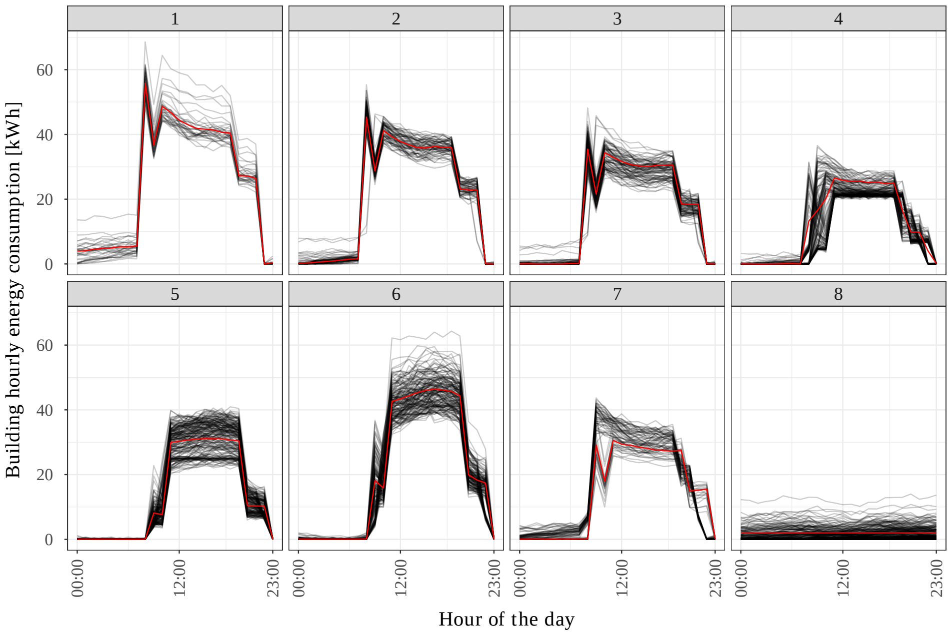


Figure 3.5: Load profiles identified for the reference building

Once the clusters have been identified, the temperature dependence of the building is estimated for each cluster of days, after obtaining the optimized change-point temperature and hysteresis (for the reference building T_{cp} = 15.2 °C, H = 2.9 °C) . Figure 3.6 shows the temperature dependence detected for each of the clusters, as the relationship between the daily energy consumption and the T^*, introduced in equation (3.5). Figure 3.7 illustrates the time distribution of the different load profiles for year 2016. An analysis of Figure 3.6 shows that most of the profiles have heating dependence, except for clusters 5 and 6, representing typical weekday summer profiles. Cluster 8 exemplifies the typical unoccupied profile (weekends and holidays), having very low consumption and slight temperature dependence for both heating and cooling. Distinction between profiles with heating dependence and profiles with cooling dependence can also be drawn by analysing the consumption peaks in Figure 3.5: winter profiles have peaks in the early morning, while summer profiles have peaks in the hours of highest solar radiation.

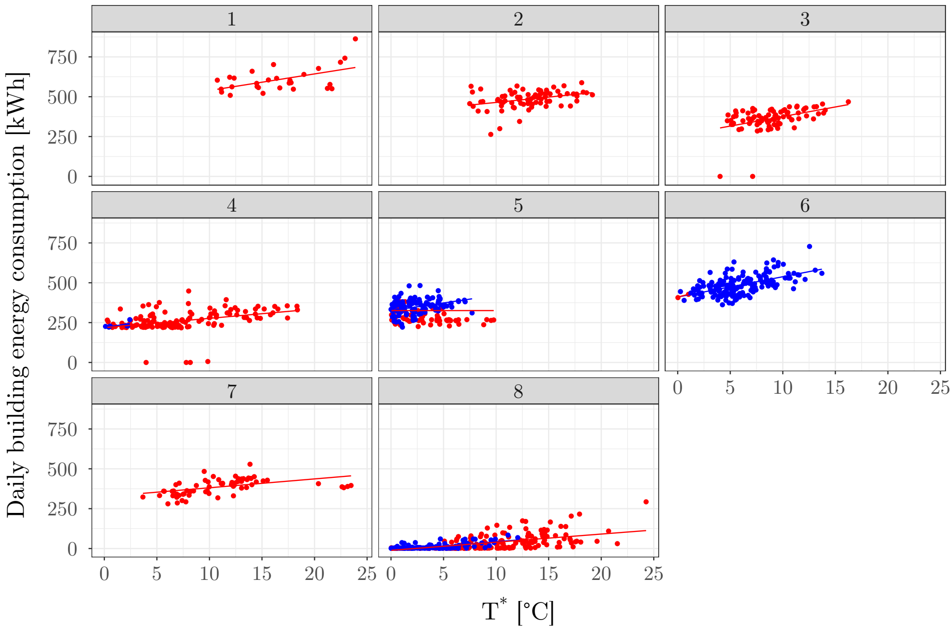


Figure 3.6: Heating (red) and cooling (blue) temperature dependence by cluster identified for the reference building (T_{cp} = 15.2 °C h = 2.9 °C)

The peak consumption values shown in figures 3.5 and 3.6 reveal how the heating load for this building is considerably higher than the cooling one. This seems realistic, since the pilot energy use was simulated using weather data from Berlin, where the climate is characterized by cold winters and moderately warm summers. A similar conclusion can be reached by studying Figure 3.8, where $T^*_{cooling}$ and $T^*_{heating}$ are shown, with $T^*_{heating}$ reaching higher values compared to $T^*_{cooling}$.

After completing the cluster and weather analyses, the GAM model is fitted to the input data. In Figure 3.9, it's possible to see the consumption predictions obtained for the office building in Phase 2 of the methodology, that is before applying the savings evaluation technique. The baseline estimated consumption (in red) is compared to the simulated consumption obtained using EnergyPlus (in black). Figure 3.9 shows a high GOF of the model, which appears to be able to capture the energy dynamics of the building, and to accurately predict its electricity consumption for the three years analyzed. Note that the statistical model developed

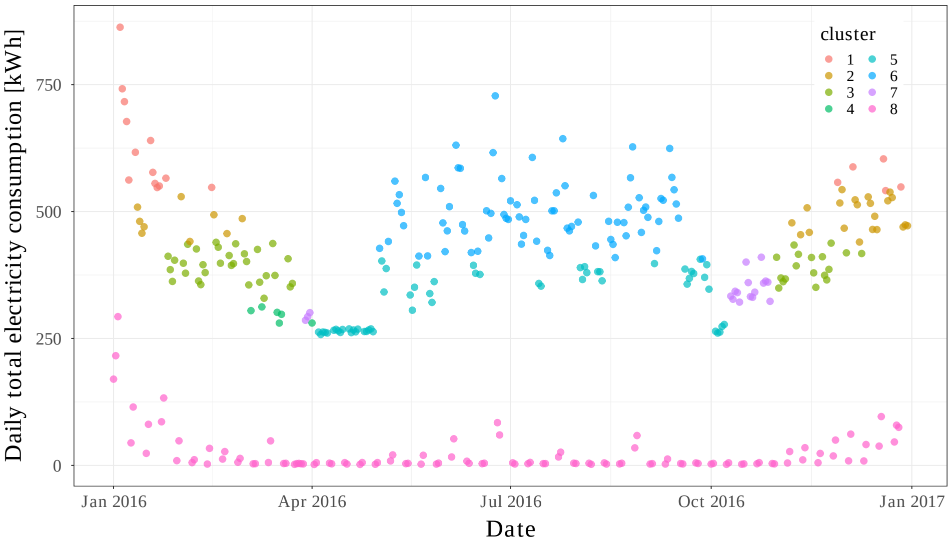


Figure 3.7: Cluster distribution identified for the reference building in the first year of operation

in this phase has the goal of simulating the normal operation of the building: no savings are estimated yet at this stage. For the case study proposed, a train/test split was performed in the GAM fitting phase: the model was trained on 80 % of the available time-series data, and then tested on the remaining 20 %. This is a common practice when dealing with predictive techniques and has the goal of making sure that the model is not over-fitting the training data and is able to provide good predictions on new observations. Ultimately, once the GAM coefficients are obtained, the savings for the implemented measures are calculated following the procedure described in Section 3.2.4.

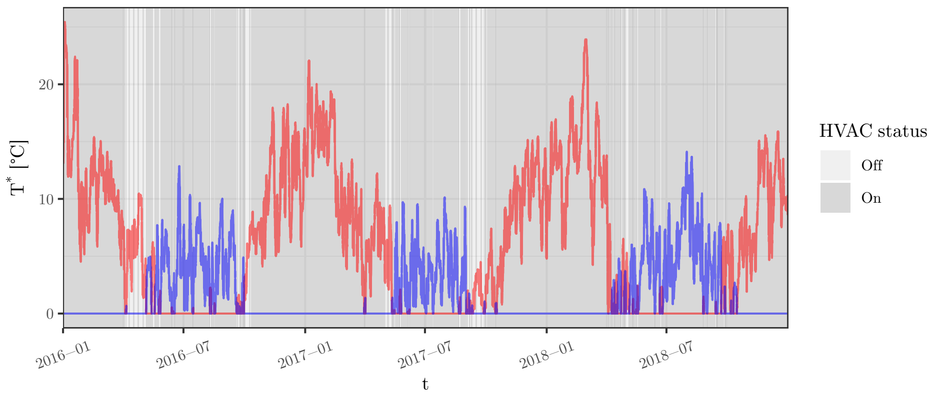


Figure 3.8: $T^*_{cooling}$ (blue) and $T^*_{heating}$ (red) profiles identified for the reference building (T_{cp} = 15.2 °C, H = 2.9 °C)

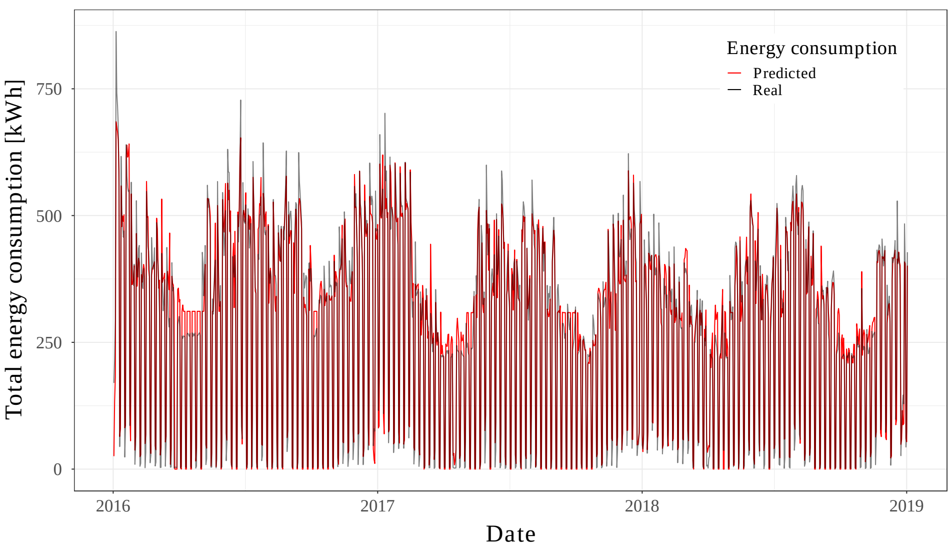


Figure 3.9: Predicted (baseline estimated) and real (simulated) consumption time-series for the reference building

Overall performance analysis

This part of case study 1 is dedicated to the analysis of the model fit and savings estimation accuracy obtained for the 54 buildings analyzed. In order to compare the GAM based methodology and the TOWT model, three different accuracy metrics were calculated:

- the CV(RMSE), calculated on a random holdout set of the training data,
- S_{diff}, the normalized percentage difference between the model estimated savings and the simulated savings calculated with EnergyPlus,
- the daily baseline $CV(RMSE)_{BL}$.

Each of these three metrics has a specific objective: the CV(RMSE) calculated on a holdout set of the training data is a commonly employed metric in the M&V setting to evaluate the goodness of fit of the model and its ability to generalize to unseen data. The normalised difference between the aggregated savings calculated with EnergyPlus and the ones obtained using the data-driven models can be used for a quick estimation of how well the models perform at estimating overall savings across a whole year of operation. While the savings difference is really intuitive in its interpretation, it does not provide information about the accuracy of savings estimations on the daily scale. For this reason, the baseline CV(RMSE) is also included in the analysis. These indicators are described in detail in Section 3.2.4.

Figures 3.10, 3.11, and 3.12 show the $CV(RMSE)$, S_{diff} and $CV(RMSE)_{BL}$ values obtained for the 54 buildings analyzed, where each box represents the distribution of the results for the 18 different simulations belonging to the noted building use (office, primary care center, hospital). The graphs contain data for the GAM based methodology, and for the CalTRACK implementation of the TOWT model, first using a full year of training data, and then reducing the training data to 9 months. The training data reduction was performed in order to test the hypothesis that one of the main advantages of the methodology proposed in this chapter is that, thanks to its ability of harnessing data from both the periods before and after an EEM is applied, it is able to accurately predict energy savings even when less than one whole year of historical data, before an EEM is applied, is available.

To estimate the CV(RMSE) for the scenarios with less training data, a 10-fold cross validation approach was used. It is a technique commonly employed within the machine learning community and was introduced in the M&V environment by Touzani et al. [52]. The section of the time-series allocated for training was divided into 10 different sub-samples, called folds. In the first iteration, the baseline model was calculated using the first 9 folds for training, while the validation error was calculated on both the held out fold of the training data *and* the rest of the time-series which was not used for the training. The procedure was then repeated 10 times, holding out a different fold every time, and the final CV(RMSE) was calculated as the mean of the results obtained in each iteration. It was decided to use this technique because it allows the CV(RMSE) to better represent how well the model is able to generalize to a whole year of conditions, when only part of that year is available as training data.

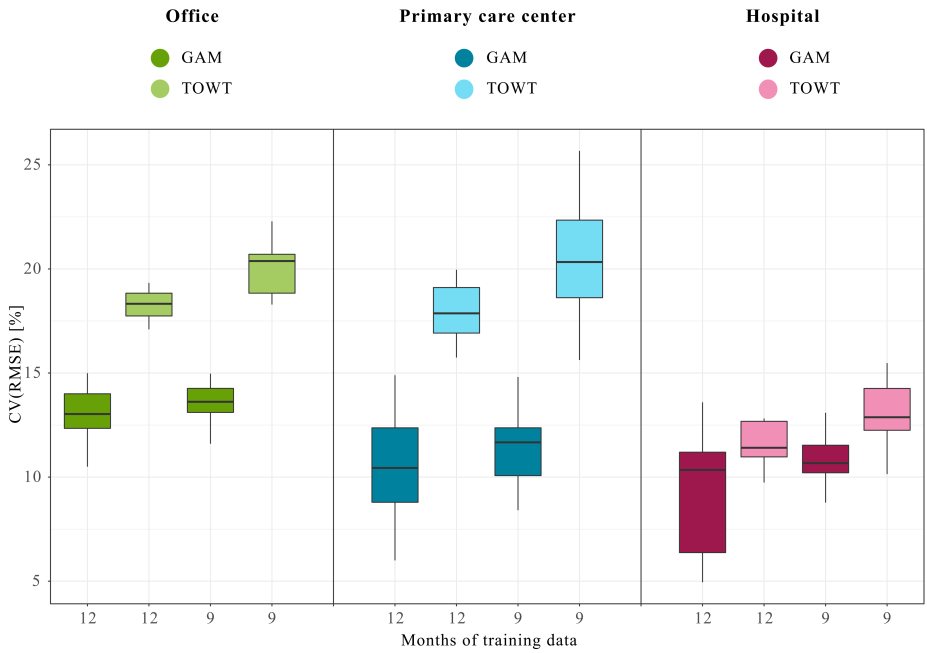


Figure 3.10: CV(RMSE) distribution for the 54 buildings analyzed in Case Study 1

3.4.2 Case study 2

The results presented for the second case study are more focused on the performance of the prediction algorithms, rather than on the intermediate results, such as profile pattern detection and temperature dependence, which have been discussed in detail for case study 1. While the first case study has the goal of showing that the proposed methodology is able to provide actionable insights and accurately estimate energy efficiency savings, the objective of case study 2 is to prove that the methodology works well not only in simulated cases, but also with monitoring time-series data coming from real-world buildings. Synthetic time-series data generated with simulation engines usually does not include stochastic variability

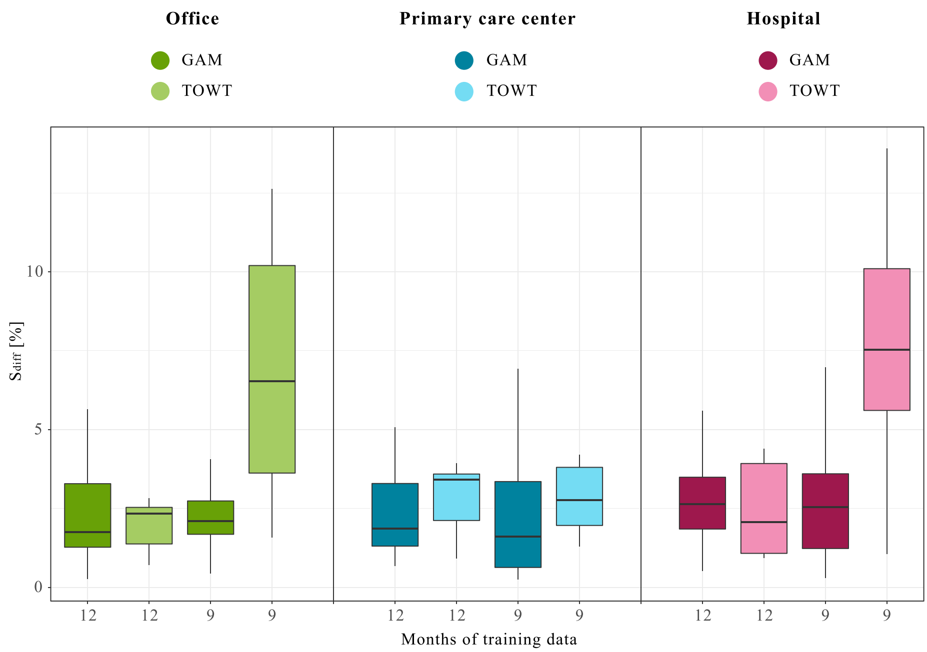


Figure 3.11: S_{diff} distribution for the 54 buildings analyzed in Case Study 1

observed in the actual energy use of existing occupied buildings, therefore it is important to test the methodology for real-world use cases and ensure that it still provides reasonable results with high goodness of fit.

For the three buildings and measures described in section 3.3.2, the methodology was applied and the results obtained were compared with the ones provided by the TOWT model, implemented according to the CalTRACK methodology. The TOWT results were calculated using the Python library *eemeter* [133]. Table 3.4 summarizes the results obtained for the three test cases. Similarly to case study 1, the table includes: the optimal number of clusters representing the load profiles, the change-point temperature and hysteresis of the building, the CV(RMSE) of the model, calculated on the test set, and the estimated savings for each measure, both in kWh and as percentage of the total reporting period consumption:

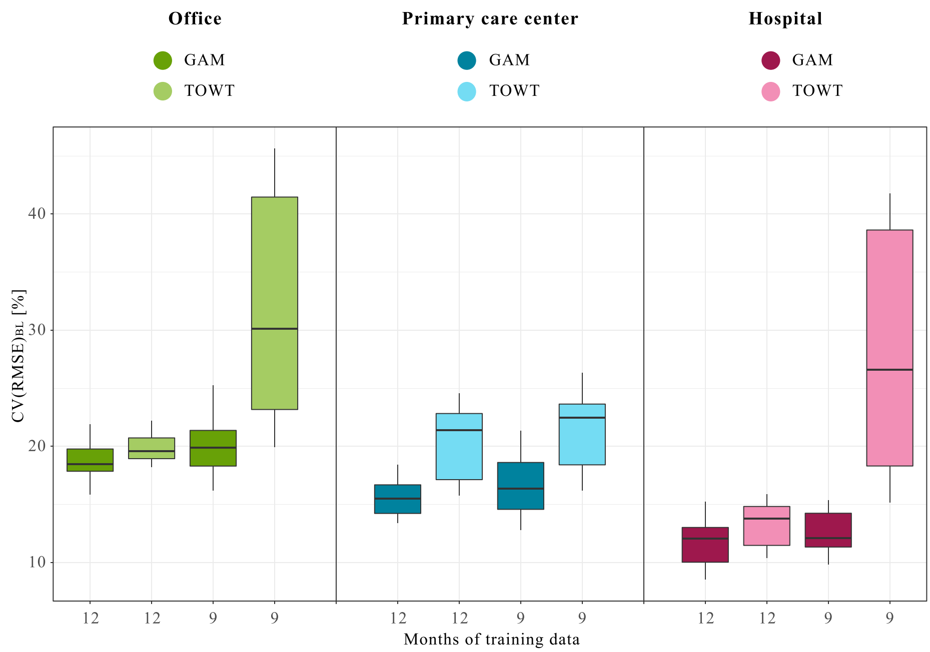


Figure 3.12: $CV(RMSE)_{BL}$ distribution for the 54 buildings analyzed in Case Study 1

Table 3.4: Results obtained for the three buildings in Case Study 2

Building Block	Building 1	Building 2	Building 3
Building information			
Number of clusters	7	7	7
Change-point temperature [°C]	20.8	21.7	21.2
Hysteresis [°C]	3.2	3.8	3
Model metrics			
CV(RMSE)	13 %	9.1 %	7.8 %
CV(RMSE) TOWT	18.1 %	11.1 %	10%
Savings			
Model estimated savings [kWh]	18592 (8.6%)	16644 (10.9 %)	108619 (8.3 %)
Savings estimated with TOWT [kWh]	12979 (6%)	26313 (17.2 %)	122280 (9,3%)

The proposed methodology provides 2-5% lower CV(RMSE) than the TOWT case for all the three analyzed buildings. In Figure 3.13, the overall fit of the model for Building 1 is also presented, showing high GOF, similar to the one seen for synthetic data in case study 1.

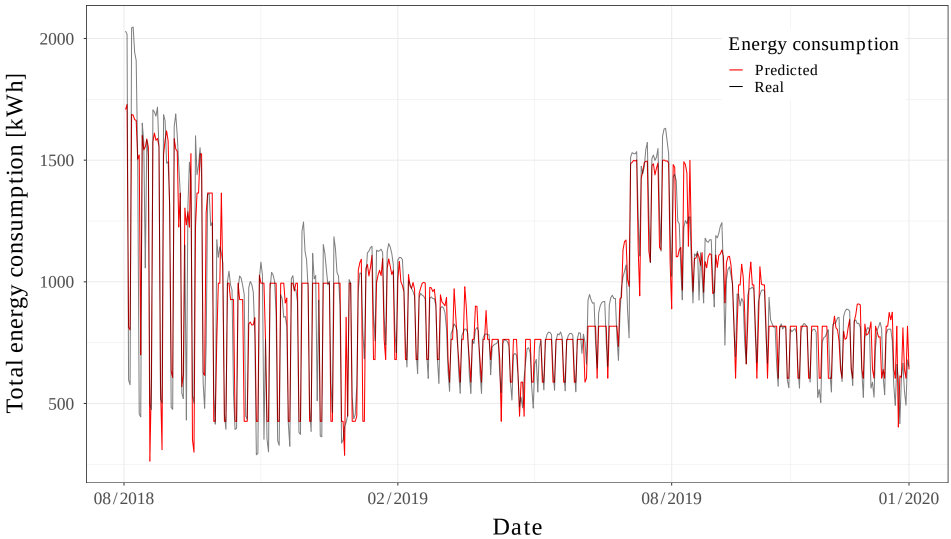


Figure 3.13: Predicted (baseline estimated) and real (metered) consumption time-series for Case Study 2, Building 1

For Case Study 2, the limited availability of monitoring data (less than 1 year for training) didn't allow to perform an analysis with reduced training data. Also, the lack of simulated data for this case prevents the comparison of the accuracy of the final savings estimations for the two tested models.

3.5 Discussion

The two analyzed case studies provide important results that showcase the proposed methodology's strengths. The detailed single building analysis of the first case study presents the additional insights that the methodology provides, such as the consumption profiles clustering and weather dependence analysis. The patterns and dependence detected are coherent with the prior knowledge about the building and the implemented measures, and represent actionable insights that energy building managers can use to improve the operational efficiency of the building. The CV(RMSE) values represented in Figure 3.10 show that the GAM based methodology represents a solid GOF improvement compared to the TOWT model, with lower median CV(RMSE) for each of the analyzed cases. It's also possible to see that the training reduction from 12 to 9 months marks a definite increase in the model error for the TOWT model, while in the GAM case the median error increase between the 12 and 9 months case is almost negligible. Only in the case of the hospital building, some isolated cases are performing better in the TOWT case than in the GAM, but in terms of median error and distribution, the GAM is still performing better. Figure 3.11 shows the savings difference S_{diff} (defined in Section 3.2.4): while the 12 month results are generally comparable between the two models, in two of the three building typologies the training data reduction results into a steep increase in the savings estimation error, with errors higher than 10% of the total building consumption in some cases. While the primary care center seems to provide more robust results in this sense, likely thanks to a reduced variance in the energy consumption values, which makes it less prone to over-fitting when removing training data, for the other two pilot buildings the TOWT model is characterized by lack of robustness in the estimations, when working with less than a full-year of training data. It's also important to note that the S_{diff} parameter shouldn't be blindly trusted, since good predictions of aggregated yearly savings do not necessarily imply accurate daily baseline estimations. This is evident since a model that underestimates the consumption during certain days of the time-series, and overestimates it during others, will have the same aggregated result as one that provides accurate daily baseline estimation for all of the analyzed days. In order to gain a deeper insight into the estimation capabilities of the proposed methodology, and taking into account the mentioned issues of the S_{diff} parameter, an additional metric, $CV(RMSE)_{BL}$ was calculated for all of the 54 analyzed cases. The $CV(RMSE)_{BL}$ distribution is shown in Figure 3.12, and illustrates again the two trends seen in Figures 3.10 and 3.11: the median of the GAM model is always lower than its TOWT counterpart and, for two of the three buildings, the training data reduction marks a steep increase in the model error. While in terms of overall aggregated yearly savings, the two analyzed models are comparable, with each seemingly performing better in certain cases, the $CV(RMSE)_{BL}$ distributions show that in terms of daily savings estimations the GAM model is more accurate than the TOWT model in the majority of the cases, with the TOWT model reaching very high error values when the training dataset is reduced.

The results of case study 2 show that the proposed methodology retains high accuracy also when used to analyze monitoring data from real-world buildings. The CV(RMSE) of the model was below 10% for two of the three buildings of case study 2, and below 15% for the third one. Even though for this case study there is no simulated ground truth to compare the estimated savings, it's still possible to see that even with missing data, and the inherent stochastic variability of real-world buildings, the proposed methodology is able to provide accurate results with high goodness of fit.

The results from both case studies show an overall superior accuracy of the proposed GAM-based methodology on the CalTRACK implementation of the TOWT model, which was taken as the industry benchmark. There might be different factors driving this performance difference: firstly, the proposed methodology implements different techniques to accurately describe the energy usage of the building before modeling the energy consumption and counterfactual. Secondly, the GAM-based methodology is able to harness part of the post EEM application data to increase the accuracy of the energy baseline model. Moreover, the improved performance of the proposed model is also achieved thanks to its ability to harness additional variables which are ignored by the TOWT model, such as global horizontal radiation and wind speed, and to aggregate and transform known variables into information valuable for the prediction, such as typical consumption profile patterns and weather dependence. Another feature that grants the proposed model increased accuracy is the preprocessing that some of the explanatory variables undergo, before being employed in the model. Specifically, time features such as day of the week and time of the year are transformed using Fourier decomposition, and temperature is treated with a low pass filter. Additionally, thanks to the weather dependence analysis module, the weather variables are only included in the model for those days when they are supposed to be affecting the building consumption. Finally, the methods described in this chapter are specifically aimed at estimating savings on the daily timescale, while the TOWT provides hourly savings which can then be aggregated on a daily level. This means that the TOWT model may have wider applicability, but lower accuracy on the daily timescale.

On a more general level, the results show that with only the use of consumption monitoring and weather data, the proposed methodology is able to achieve low model errors, with CV(RMSE) lower than 15% for all the buildings in both case studies. Regarding the availability of the necessary weather data, accurate historical GHI data can be obtained for European locations through the the Copernicus Atmosphere Monitoring Service (CAMS)[134]. Other weather services and APIs are also available both in Europe and globally, providing historical wind speed, GHI, and outdoor temperature data free of charge, some of them are: DarkSky[130], Meteonorm[135], OpenWeatherMap[136]. Given the relative ease of acquisition of the necessary weather data through different weather services and APIs, the proposed methodology is able to accurately detect savings, having as only requirement the availability of hourly electricity monitoring data. Including in the models additional variables such as occupancy levels or technical details of the different systems in use in the building is likely to increase the accuracy of the savings estimation, but at the same time would highly reduce the applicability of the methodology.

3.6 Conclusions and future work

In this chapter, an innovative methodology for the measurement and verification process of energy efficiency savings in commercial buildings was presented. The proposed approach has three main phases: first a data-preprocessing phase, where different algorithms are used to characterize the building and extract valuable information such as typical electricity load profiles, change-point temperature, and climate dependence. In the second phase, a generalized additive model is fit, using daily consumption and climate data, as well as the variables obtained in the previous phase, to estimate the baseline energy usage of the analyzed building. Finally, in the last phase, for each energy efficiency measure applied, the savings are estimated as the difference between the metered energy consumption and the estimated baseline consumption for the same period. The presented approach was tested to detect savings in two separate case-studies. The first case study analyzes three different building typologies over a three-year period. The buildings' electricity consumption time-series were simulated using the energy modeling software EnergyPlus. Each pilot had a different use (office, primary care center, hospital), and they were simulated with different climate conditions (oceanic, Mediterranean, continental), and implemented EEM combinations, for a total of 54 unique simulated cases. The second case study involves the analysis of monitoring data from three real-world buildings located in Barcelona (Spain). In both case studies, the model proved to be able to capture the dynamics of the buildings, providing CV(RMSE) below 15 %. For case study 1, the median difference between the savings estimated with the proposed methodology, and the ones obtained with the deterministic approach never exceeded 3% of the total reporting period consumption. The results were also compared to the ones obtained by applying the time-of-week and temperature (TOWT) model, following the CalTRACK methodology, with the proposed model showing overall superiority, with up to 10% lower CV(RMSE) than TOWT in both case studies.

To conclude, four main strengths were identified for the proposed methodology, linked with the ability of this approach to provide:

1. high accuracy in the estimation of savings associated with the predictions;

2. a robust savings quantification that, for the pilot buildings analyzed in case study 1, does not reduce drastically when less than a full year of training data is available;

3. wide applicability, since the data required is limited to hourly metered electricity consumption and weather data, gradually more available for most commercial buildings;

4. additional actionable information to stakeholders, such as: typical load consumption profiles, change-point temperature of the building, and weather dependence evaluation.

Although the preliminary results obtained in this study seem promising, it is clear that more work is needed in order to further validate the approach, and accurately define its strengths and weaknesses. Future work might include the use of automated feature selection techniques to include or omit certain weather variables depending on the specific building analyzed. An additional level of insight could also be added by estimating an independent change-point temperature for each detected load profile cluster in the building. At the same time, it appears clear that all the listed benefits come at the expense of adding more complexity to other commonly used approaches. Although the proposed methodology provided high accuracy of estimation in both the synthetic and real data use-cases, it would be optimal to test the presented techniques on a large cluster of real-world buildings, in order to assess how well the model generalizes to new buildings, and if the predictive accuracy is similar to the one obtained with simulated data in this study.

Chapter 4

Bayesian approach for hourly baseline estimation and consumption characterization

This chapter was published as a journal article:

Grillone B., Mor G., Danov S., Cipriano J., Lazzari F., Sumper A. Baseline energy use modeling and characterization in tertiary buildings using an interpretable Bayesian linear regression methodology. Energies 2021; 14, 5556, https://doi.org/10.3390/en14175556

4.1 Introduction

In 2019, CO_2 emissions related with the operation of buildings have reached a historical peak of 10 $GtCO_2$, representing 28% of total global carbon dioxide emissions, as shown in the latest reports of the Global Alliance for Buildings and Construction [137]. While the energy intensity of the building sector has been steadily decreasing since 2010, an average annual floor area growth rate of around 2.5% has been enough to offset this trend [138]. This highlights a vast energy efficiency potential, linked with building energy codes lacking behind in emerging economies and renovation rates remaining low in developed nations. In Europe, only about 1% of the existing building stock is renovated each year, but energy retrofitting projects might have the potential of lowering the EU's total CO_2 emissions by 5% [10].

The lack of methods to accurately quantify the savings generated by energy efficiency projects represents a significant barrier towards the attraction of financial investments in this field [139]. In recent years, investments in renewable energy worldwide have been higher than those in energy efficiency by approximately 20 % [140]. One of the reasons is the possibility of directly metering the energy generated (and therefore the return on investment), as opposed to rough estimations of savings in the case of energy renovation projects. Afroz et al. highlighted how the use of different baseline models can provide very different savings results, and how it is becoming a usual practice in the industry to prefer practical and intuitive models to more sophisticated ones with higher accuracy but poor interpretability [141]. Additionally, different studies have shown that accurately quantifying the uncertainty of obtained savings estimates is of utter importance in energy efficiency projects, but that the commonly employed methods have a tendency to underestimate said uncertainty[142].

Alongside traditional energy efficiency programmes, in recent years new business schemes are arising from the intersection between energy efficiency and demand side flexibility[143]. Projects with these characteristics require advanced baseline estimation techniques, able to accurately and dynamically estimate both energy savings and dynamic uncertainty bands at hourly or sub-hourly level. Two schemes worth mentioning in this field are Pay for Performance (P4P) [12] and Pay for Load Shape (P4LS) [13], in which customers are dynamically compensated for changing their consumption load shape in order to match the evolving conditions of the grid. These new approaches that combine energy efficiency and demand side flexibility arise from a necessity of not only lowering the overall energy usage intensity of the built environment, but also creating a permanent shift of the consumption curve in order to match the hours of highest energy production from renewable sources [14]. These findings highlight that novel measurement and verification (M&V) methodologies able to accurately estimate achieved hourly savings and uncertainty bands, as well as displaced loads, have the potential of unlocking considerable investments in energy efficiency [144].

4.1.1 M&V background

One of the main applications building energy baseline models are used for is the measurement and verification of energy efficiency savings. The International Performance Measurement and Verification Protocol (IPMVP) defines measurement and verification as the practice of using measurements to reliably determine actual savings generated thanks to the implementation of an energy management program within an individual building or facility [145]. Being energy savings a result of not consuming energy, they can't be directly quantified. The usual approach for estimating savings achieved by energy efficiency initiatives is to compare the energy usage of the facility before and after the application of the intervention, while also implementing the required adjustments to account for possible changes in conditions. In M&V, there are two main guidelines which are globally recognized: the International Performance Measurement and Verification Protocol (IPMVP), and the ASHRAE Guideline 14 [45]. Both these protocols are based on the adoption of a baseline energy model to compare the energy performance of the investigated facility before and after the implementation of an energy efficiency measure. The baseline model has the goal of characterizing the starting situation of the facility and it is used to separate the impact of a retrofit program from other factors that might be simultaneously impacting the energy consumption.

4.1.2 Bayesian paradigm in M&V

Bayesian methodologies have been proved to hold great potential for M&V, by providing richer and more useful information using intuitive mathematics [146]. Being this article mainly focused on Bayesian applications in the M&V setting, a full explanation of the theory behind the Bayesian paradigm is out of its scope. Readers are instead referred to the following texts: [69, 147, 148]. The added value provided by Bayesian inferential methods is linked to their coherent and intuitive approach to the M&V questions, as well as to their probabilistic nature, that enables automatic and accurate uncertainty quantification when estimating energy savings [70]. Shonder and Im [71] showed the Bayesian approach to be a coherent and consistent methodology that can be used to estimate savings and savings uncertainty when performing measurement and verification calculations. Lindelöf et al. [72] tested the Bayesian approach to estimate the savings coming from an ECM implementation, and stressed the clear interpretability and communicability of the results obtained with the Bayesian method. The advantages of communicating to decision makers not only point estimate results, but whole probability distributions with the additional information included. Grillone et al. [115] gathered together the latest Bayesian specifications for M&V. Among the different reviewed techniques, Gaussian mixture regression [76] and Gaussian processes [73] stand out because of their ability to capture nonlinear relationships between variables.

4.1.3 Multilevel models

One of the model specifications implemented in this work is multilevel regression. Multilevel (also known by the names of hierarchical, partial pooling, or mixed effects) models are interesting model specifications that prove useful when the analyzed data can be grouped into clusters of similar behavior. More specifically, multilevel models allow the *pooling* of information between the different clusters present in the data, meaning that the model learns simultaneously about each cluster while learning about the population of clusters [147]. In recent years, also thanks to the increase in computing power, multilevel models have gained great interest and have been applied in many different fields of science and technology [149]. In the energy field, they have found application in the calibration of building energy models [150, 151], forecasting of electricity demand [152–154], estimation of overhead lines failure rate [155], estimation of photovoltaic potential [156], and to analyze the degradation of electric vehicle batteries [157]. In the M&V setting, Booth et al. [158] used a hierarchical framework to generate energy intensity estimates for various dwelling types. These estimates were then used to calibrate the parameters of an engineering based bottom-up energy model of the housing stock.

4.1.4 Present study

Although some research has been carried out on the use of Bayesian methods for M&V, to the best of the authors' knowledge, no paper introduced a Bayesian linear regression model, with clear interpretability of the parameters and providing high granularity predictions. Some published works proposed a Bayesian approach with physical interpretability of the coefficients [72], but only providing predictions with monthly granularity, while others focused on extending the Bayesian approach to high frequency time-series predictions, at the cost of reducing the model interpretability [73, 76]. Our scope in the present work is to introduce a Bayesian linear regression model that can be easily interpreted and explained to stakeholders, but that at the same time can be applied to time-series with high granularity and can provide accurate and dynamic hour by hour consumption and uncertainty estimates. The model proposed was tested on one of the largest publicly available datasets of non-residential building energy consumption, creating in this way a benchmark for future studies.

The model proposed in this chapter uses time features and coefficients marking the temperature dependence of the building, while also including information about its typical consumption profile patterns, detected using a clustering algorithm. All the parameters of the model have a clearly interpretable meaning, and the Bayesian approach is able to automatically estimate what are the heating and cooling change-point temperatures of the building, that is the outdoor temperatures below/above which a significant relationship between the building's energy consumption and the outdoor temperature conditions is detected. The methodology comes with several advantages:

1. The model is explainable and its coefficients have a clearly interpretable meaning, a feature which is highly valued by investors in building energy renovation and other stakeholders of the industry.

2. Being the model probabilistic, uncertainty is estimated in an accurate, automatic, elegant and intuitive way.

3. The model coefficients and target variable are expressed in terms of probability distributions instead of point estimates, providing a range of additional actionable information to stakeholders.

4. The provided uncertainty ranges are characterized by dynamic locally adaptive intervals, reflecting how the uncertainty is not symmetrically distributed around the mean of the predictions, and that the values can vary depending on the distribution of the explanatory variables.

5. The methodology provides extra useful information for stakeholders such as the typical consumption profile patterns of the building, as well as the probability distributions of the heating and cooling change-point temperatures and of the heating and cooling linear coefficients.

6. Since Bayesian models are fit to provide reasonable results even when the training data available is scarce, the methodology is well-fit to solve M&V problems, where it is not always feasible to obtain long training time-series with high granularity. Furthermore, the model also has high applicability since the data required is limited to hourly electricity consumption and outdoor temperature values, which are fairly easy to obtain.

The mentioned advantages represent a significant advance to the state-of-the-art, widening the scope of M&V applications and providing the necessary risk mitigation required by financial institutions, thanks to an automatic, dynamic and accurate estimation of uncertainty intervals. Furthermore, the possibility of computing locally adaptive uncertainty ranges can be of fundamental value in those projects where the time-allocation of the savings is as important as the savings themselves. The approach was tested on an open dataset containing electricity meter readings at an hourly frequency for 1578 non-residential buildings. The dataset was collected within the framework of the Building Data Genome Project 2 [159] and was partly used in the Great Energy Predictor III competition hosted in 2019 by ASHRAE [160]. Different model specifications were tested on the buildings contained in the dataset, in an attempt to identify their effect on the final accuracy of predictions. This model comparison process was structured in four consecutive phases, which are described in detail in Section 4.2.2. The chapter is structured as follows. First the methodology is described in detail, together with the model comparison process. Then the case study is presented and the results obtained are displayed and discussed. Finally, conclusions and future work recommendations are drawn.

4.2 Methodology

In the present chapter, a Bayesian linear regression methodology capable of providing detailed characterization of building energy consumption, as well as high granularity baseline energy use predictions, is introduced. Different Bayesian model specifications were tested and mapped to changes in the accuracy of hourly energy consumption predictions for a dataset of more than 1500 buildings. Various model variables, prior distributions, and posterior estimation techniques were compared, as well as the effect of implementing a multilevel regression in place of a classical single-level model. The different model specifications were tested on the same dataset, and the results were compared in terms of CV(RMSE) and coverage of the uncertainty intervals. In the following paragraphs, first, the characteristics of the proposed Bayesian approach are presented in detail. Then, the the procedure and metrics used to perform the model comparison are discussed.

4.2.1 Bayesian methodology

The main aim of the Bayesian inference methodology proposed in this research work is to characterize the energy consumption dynamics of individual buildings by means of a linear regression model that uses time and weather features as predictors. The development of such a model enables a deeper understanding of the analyzed buildings' energy performance and the generation of baseline energy use predictions. In the framework of Bayesian inference, this model is estimated by applying Bayes' theorem [69] as follows:

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)} \tag{4.1}$$

where M are the model parameters and D are the measured data. $P(M|D)$ is the conditional probability of the model parameters' values, given the measured data from the building, this is also called *posterior distribution*. $P(D|M)$ is the probability of measuring the observed data, conditional on the model parameter values, it can also be called *likelihood*. $P(M)$ represents the prior knowledge of the modeler about the plausible distribution of the model parameters, this is referred to as the *prior probability distribution*. $P(D)$ represents the probability of observing the data and is usually called *marginal probability* or *marginal likelihood*. In practical terms: we define a regression model, aimed at estimating building energy consumption values based on time and weather features, then we provide to this model a set of measurement data, a prior probability distribution of the model parameters (based on our previous knowledge of building physics) and a likelihood function for the observed data. Through the formula in (4.1), the Bayesian inference model is then able to provide a posterior probability distribution of the model parameters and of the target variable (electricity consumption). The model and target variable posteriors can then be used to characterize the energy consumption of the analyzed building and to generate baseline energy consumption predictions and uncertainty bands. The flowchart in Figure 4.1 represents a the structure of the proposed methodology, including the initial phase of data pre-processing in which a clustering algorithm is used in order to detect recurrent daily load profiles in the building, which then are used as data for the regression model. Following, the technical details of the presented methodology are presented and discussed.

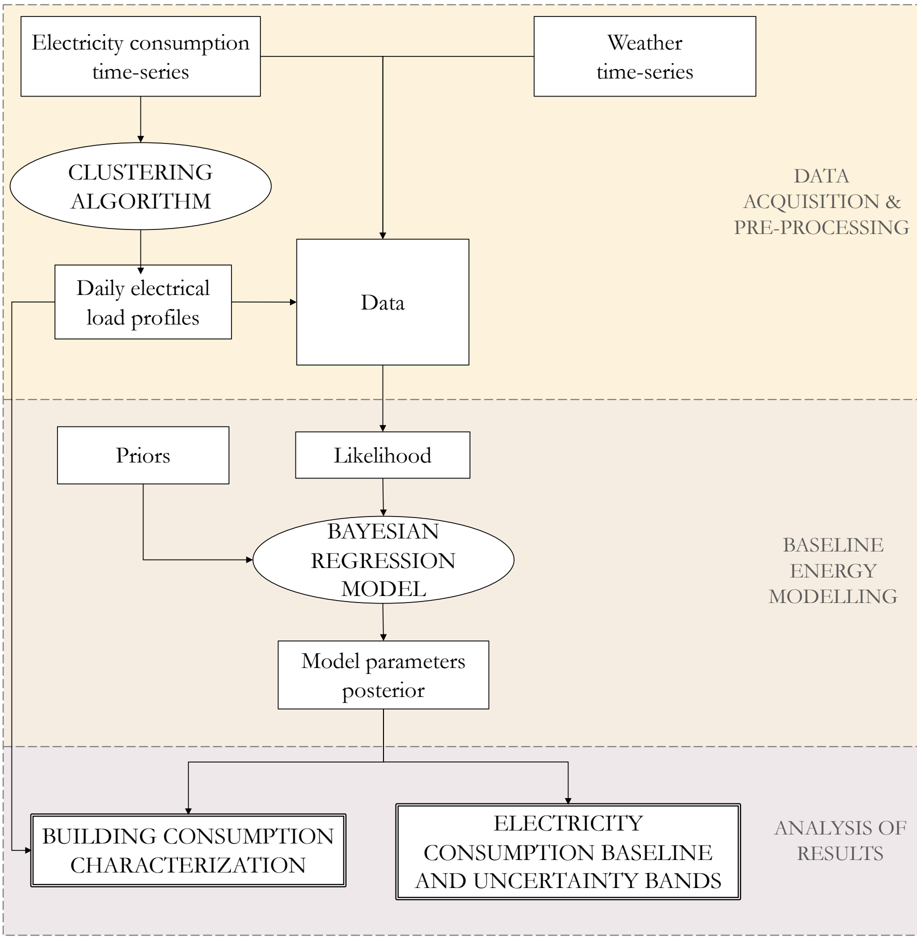


Figure 4.1: Methodology flowchart

Data pre-processing

The first step of the proposed Bayesian approach is a data pre-processing phase that is critical in order to build the dataset that is used in the analysis. The pre-processing workflow consists in the detection of clusters of days that have similar electricity usage patterns. In order to implement the pattern recognition algorithm, the data is first transformed through the following steps:

1. The original frequency of the consumption data is resampled (aggregated) to have one value per 3 hours (8 values per day).

2. For each day in the time-series, the consumption values Q^{abs} are transformed into daily relative values Q^{rel}. $Q_t^{rel} = \frac{Q_t^{abs}}{\sum_{t \epsilon day} Q_t^{abs}}$

3. A matrix of relative consumption values is generated, having as rows the days of the time-series and as columns the 8 parts of the day defined in point 1.

4. The values in the matrix are transformed with a normalization between 0 and 1. This enables more accurate predictions with the clustering algorithm. $Q_{day,dh}^{norm,rel} = \frac{Q_{day,dh}^{rel} - min(Q_{dh}^{rel})}{max(Q_{dh}^{rel}) - min(Q_{dh}^{rel})}$

To detect the profile patterns, a spectral clustering methodology is implemented [161]. This specific clustering technique is performed by embedding the data into the subspace of the eigenvectors of an affinity matrix. This is done through the use of a kernel function, specifically, the one used in this application was a radial basis kernel, also known as Gaussian kernel. Using this kernel, an affinity matrix $A_{i,j}$ that is positive, symmetric, and depends on the Euclidean distance between the data points is defined:

$$A_{i,j} \simeq exp(-\alpha \|x_i - x_j\|^2) \tag{4.2}$$

Then we define the degree matrix $D = \sum_j^n a_{i,j}$, a diagonal matrix that summarizes the affinity of each element of A with all the other elements of the matrix. Using the affinity and the degree matrices we can calculate the unnormalized graph Laplacian:

$$L = D - A \tag{4.3}$$

If the analyzed data is actually spaced so that there are different clusters, the Laplacian L will then be approximately block-diagonal, and each block will define a cluster. From the analysis of the eigenvalues and eigenvectors of L, it's possible to estimate what is the optimal number of clusters.

Once the recurrent daily profiles have been detected for the training year, a classification algorithm calculates the cross euclidean distance matrix between the detected cluster centroids and the load curve of each day of the test year, assigning them to one of the identified clusters.

Regression model

The linear regression model developed in this work is based on several coefficients aimed at capturing the dynamics that drive the energy consumption of the analyzed buildings. The hourly energy consumption of the building is supposed to be partially dependent on time-features such as the hour of the day, and partially on thermal dynamics driven by outdoor temperature variations. In the estimation of the temperature-dependent term, a heating change-point temperature and a cooling change-point temperature are defined: the outdoor temperatures below/above which a significant relationship between the building's energy consumption and the outdoor temperature conditions is detected. This model structure is based on the concept of linear change-point models, first introduced in literature with the PRISM method [162]. The coefficients of the model, with the exception of the change-point temperatures, are supposed to be varying depending on the load profile of the day (detected with the clustering algorithm previously introduced), or on the day-part, a fictitious variable generated after dividing the day in 6 day-parts of 4 hours each. The mathematical description of the Bayesian linear regression model used follows, where the square brackets notation has been used to represent the variables that have a different value for each profile cluster k or day-part j. The *likelihood* assigned to the observed variable (hourly electricity consumption) was a normal distribution with mean μ_i and standard deviation σ:

$$y_i \sim Normal(\mu_i, \sigma) \tag{4.4}$$

$$\mu_i = \alpha_{k[i]} + f_{dh,i} + \beta_{c,j[i]} \cdot (T_{o,i} - T_{cp_c}) \cdot d_c \cdot dep_{c,k} + \beta_{h,j[i]} \cdot (T_{cp_h} - T_{o,i}) \cdot d_h \cdot dep_{h,k} \tag{4.5}$$

$$\sigma \sim Exponential(1) \tag{4.6}$$

where:

- $\alpha_{k[i]}$ is the intercept of the model, one for each profile cluster k detected,
- $f_{dh,i} = \sum_{p=1}^{n} \delta_{k,p[i]} sin(2\pi p \frac{h_{d,i}}{24}) + \gamma_{k,p[i]} cos(2\pi p \frac{h_{d,i}}{24})$ represents the effect of the hour of the day h_d, following a Fourier decomposition with n harmonics.

$\delta_{k,p[i]}$ and $\gamma_{k,p[i]}$ are the linear coefficients that mark the weight of each hour on the final electricity consumption, one for each profile cluster k detected and for each harmonic p.

- $\beta_{c,j[i]}$ and $\beta_{h,j[i]}$ are the coefficients that represent the piece-wise linear temperature dependence of the model, one for each daypart j previously defined.
- $(T_{cp_h} - T_{o,i})$ and $(T_{o,i} - T_{cp_c})$ are the difference between the outdoor temperature and the change-point temperatures detected by the model for heating and cooling.
- $d_{h,k}$ and $d_{c,k}$ are logical variables making sure that the temperature-dependent term is only evaluated when the outdoor temperature is above/below the cooling/heating change-point temperature

$$d_h = \begin{cases} 1 & \text{if } (T_{cp_h} - T_{o,i}) > 0 \\ 0 & \text{if } (T_{cp_h} - T_{o,i}) \leq 0 \end{cases} \qquad d_c = \begin{cases} 1 & \text{if } (T_{o,i} - T_{cp_c}) > 0 \\ 0 & \text{if } (T_{o,i} - T_{cp_c}) \leq 0 \end{cases}$$

- $dep_{h,k}$ and $dep_{c,k}$ are two logical coefficients that are automatically optimized by the model and that mark whether or not a certain profile cluster k has heating or cooling dependence.

The predictors used in the regression model were one of the model specifications that were evaluated in the model comparison stage. More specifically, the inclusion of a wind speed predictor was also tested when comparing different model specifications. The mathematical formulation of the second model tested in the comparison stage (with the addition of the wind speed term) follows:

$$\begin{aligned} \mu_i = &\alpha_{k[i]} + f_{dh,i} + \beta_{c,j[i]} \cdot (T_{o,i} - T_{cp_c}) \cdot d_c \cdot dep_{c,k} \\ &+ \beta_{h,j[i]} \cdot (T_{cp_h} - T_{o,i}) \cdot d_h \cdot dep_{h,k} + W_{s,i} \cdot \beta_{ws,j[i]} \end{aligned} \tag{4.7}$$

Where linear dependence between the energy consumption and the wind speed W_s was supposed, with coefficient β_{ws}, depending on the daypart j. Finally, it's important to note that for both regression models a log-transformation of the target variable was performed, in order to improve the prediction accuracy.

Prior probability distributions

An important step of Bayesian modeling is to choose the prior probability distributions of the model parameters. These values express the belief the modeler has about the probability distribution of the parameters before analysing the data. If no previous information is known about the probability distribution of the parameters, then a uniform prior with a lower and an upper bound is a common choice, meaning that all the values contained between the two boundaries have the same prior probability. In order to test the effect of more or less informative priors on the accuracy of the model predictions, the model was first run with uniform priors on most of the parameters. The uninformative priors set were the uniform distributions with different lower and upper bounds:

$$\alpha_k \sim Uniform(-100, 100) \tag{4.8}$$

$$\delta_{k,p}\,, \gamma_{k,p}\,, \beta_{c,j}\,, \beta_{h,j} \sim Uniform(-5, 5) \tag{4.9}$$

$$dep_{c,k}\,, dep_{h,k} \sim Uniform(0, 1) \tag{4.10}$$

$$T_{cp_h}\,, T_{cp_c} \sim Uniform(-50, 50) \tag{4.11}$$

In the model comparison stage, a set of regularizing prior distributions was also implemented, in order to test their impact on the accuracy of the model predictions. A presentation and discussion of the regularizing priors used follow:

$$\alpha_k\,, \delta_{k,p}\,, \gamma_{k,p} \sim Normal(0, 1) \tag{4.12}$$

$$\beta_{c,j}\,, \beta_{h,j}\,, \beta_{ws,j} \sim HalfNormal(1) \tag{4.13}$$

$$dep_{h,k}\,, dep_{c,k} \sim Bernoulli(0.5) \tag{4.14}$$

$$T_{cp_h}, T_{cp_c} \sim Uniform(T_{min}, T_{max}) \tag{4.15}$$

For the model intercept α_k and the time features coefficients $\delta_{k,p}$ and $\gamma_{k,p}$, a weakly regularizing normal prior distribution, centered in 0 and with standard deviation equal to 1 was set. For the heating and cooling coefficients $\beta_{c,j}$ and $\beta_{h,j}$, a half-normal prior with standard deviation equal to 1 was used. This prior distribution assigns zero probability to negative values of the temperature coefficients, according to the intuition that outdoor temperature deviations above the cooling change-point temperature or below the heating change-point temperature, will result in the electricity consumption of the building either increasing ($\beta > 0$) or staying unchanged ($\beta = 0$). For the change-point temperatures, a uniform prior was used, ranging between the minimum and the maximum outdoor temperatures observed in the training data T_{min} and T_{max}. The priors of the dependence coefficients $dep_{h,k}$ and $dep_{c,k}$ were modeled as a Bernoulli distribution with probability $p = 0.5$, giving them an equal prior probability of being either 0 or 1.

Pooling techniques

Previously, a description was provided of how, in the pre-processing phase, the days of the time-series are clustered according to the shape of the load curves. In the construction of a model for observations which are grouped together on a higher level, there are two conventional alternatives: either modeling all the observations together independently of the clusters they belong to, or calculating independent coefficients for each of the clusters. Advantages and disadvantages of these two approaches have been widely analyzed in the framework of the bias-variance trade-off discussion [107]. In this chapter, we decided to explore a third option as well, that is referred here as *partial pooling*, and that goes also by the name of multilevel, hierarchical, or mixed effects regression. The term partial pooling refers to the action of *pooling* information between different clusters during the modeling phase. In our case, a *complete pooling* approach would be equivalent to the first of the previously mentioned alternatives: to assume that there is no variation between days having different load shapes, and to produce a single model estimate for the model parameters, independently of the clusters. The *no pooling* approach, on the other hand, would be to assume that the variation between the clusters is infinite, therefore nothing learned for days with a certain load shape can help predict days belonging to a different cluster. The partial pooling approach produces estimates by including in the model individual coefficients for each of the detected clusters (similarly to the no pooling case), but with the additional assumption that these same coefficients have a common prior distribution that is adaptively learned by the model [147]. In other words, in multilevel regression, although individual model parameters are estimated for each cluster, the information provided by each cluster can be used to improve the estimates for all the other clusters.

When translating this into mathematical formulation, the only difference from a traditional Bayesian regression model without multilevel structure is that, instead of the usual prior distributions, an adaptive prior is defined. The adaptive prior is a function of two additional parameters, often referred as hyperparameters, and that in turn have a prior distribution, called a hyperprior. In the context of this chapter, we decided to test whether pooling information among different clusters when estimating the intercept and time features coefficients could improve the model performance. The following adaptive priors were tested in the model comparison stage, in place of the regularizing priors defined in the previous section:

$$\alpha_k \sim Normal(\bar{\alpha}, \sigma) \tag{4.16}$$

$$\delta_{k,p} \sim Normal(\bar{\delta}, \sigma) \tag{4.17}$$

$$\gamma_{k,p} \sim Normal(\bar{\gamma}, \sigma) \tag{4.18}$$

$$\bar{\alpha}\,, \bar{\delta}\,, \bar{\gamma} \sim Normal(0, 1) \tag{4.19}$$

The advantage of multilevel regression is that it's still possible to include in the model the variability caused by the higher level variables (load shape clusters in this case), but at the same time overfitting is avoided, by stating that the priors of the model coefficients related to those variables are drawn from a common distribution. The partial pooling estimates will be less underfit than the average coefficient from the complete pooling approach, and less overfit than the no-pooling estimates. This turns out to be particularly strong when there is not so much data available for some of the categories, because then the no pooling estimates for those clusters will be especially overfit [147]. On the the other hand, when there is plenty of data for each of the groups, the effect of implementing multi-level regression turns out to be less influential on the final result, this is why we decided to test whether or not implementing this technique might have a positive effect on the accuracy of the model predictions.

Posterior estimation methods

According to the theory of Bayesian statistics, after defining the required variables, Bayesian models update the prior distributions previously set to obtain the posterior distribution. A unique posterior distribution exists for each combination of data, likelihood, parameters, and prior. This distribution represents the relative plausibility of the possible parameter values, conditional on the data and the model. Historically, being able to effectively estimate the posterior has always been one of the main practical issues of Bayesian modeling. Starting from the 1990s, thanks to the access to cheap computational power, Markov chain Monte Carlo (MCMC) [163] methods started being the prevailing technique used for posterior estimation purposes in Bayesian modeling. MCMC approaches follow the idea that, instead of computing a direct approximation of the posterior, which is many times unfeasible from a computational point of view, such an approximation can be obtained by drawing samples from the posterior. The samples drawn provide a set of possible parameter values, and their frequency represents the posterior plausibilities. One of the most popular MCMC algorithms used in practical applications is the Hamiltonian Monte Carlo (HMC). Originally proposed in 1987 by Duane et al. , this algorithm is able to draw samples more efficiently by reducing the correlation between them thanks to the simulation of Hamiltonian dynamics evolution [164]. In the present research, HMC is one of the techniques used for posterior estimation. More specifically, the No-U-Turn Sampler (NUTS) algorithm was used: an extension of the HMC algorithm that provides accurate and efficient posterior samples without needing user intervention or tuning runs [165].

One of the drawbacks of MCMC methods is that when models start having a very high number of variables and data points, their computational cost becomes extremely high. Automatic Differentiation Variational Inference (ADVI) is an alternative technique that resolves the problem of the computational bottleneck by turning the task of computing the posterior into an optimization problem [166]. First, for a given set of model parameters θ and observations x, a family of distributions $q(\theta)$, parametrized by a vector $\phi \in \Phi$ is hypothesized. The optimization is then aimed at finding the member of that family that minimizes the Kullback-Leibler (KL) divergence to the exact posterior:

$$\phi^* = \arg\min KL(q(\theta;\phi) \parallel p(\theta|x)) \tag{4.20}$$

Since the KL formulation involves the posterior, and therefore does not have an analytic form, rather than minimizing the KL, the analysis is aimed at maximizing the evidence lower bound (ELBO):

$$\mathscr{L}(\phi) = \mathbb{E}_{q(\theta)}[\log p(x,\theta)] - \mathbb{E}_{q(\theta)}[\log q(\theta,\phi)] \tag{4.21}$$

$\mathscr{L}(\phi)$ is equivalent to the opposite of the KL divergence, up to the constant $\log p(x)$, hence maximizing the ELBO is equal to minimizing the KL divergence [167].

When comparing MCMC methods to ADVI, the first are often more computationally intensive, but they have the advantage of providing (asymptotically) exact samples from the posterior. Variational inference, on the other hand, is aimed at estimating a density which is only close to the target and tends to be faster than MCMC. This makes ADVI the best choice when the objective of the analysis is testing many different models on large datasets. MCMC is often used with smaller datasets, when computational time is not an issue, or in specific situations where the higher computational cost is not considered a problem compared to the added value of obtaining more precise posterior estimations [168].

In this chapter, the computational cost of MCMC methods was recognized as a possible issue, therefore ADVI was used as the main posterior estimation technique. Only in the last phase of model comparison, considering the fact that ADVI is an optimization algorithm that does not guarantee to provide exact samples from the target density, the posterior was estimated with both ADVI and MCMC sampling techniques. This allowed to evaluate whether or not the ADVI approximation was producing models that performed significantly worse for our use-case.

Uncertainty intervals

One important difference between Bayesian statistics and traditional frequentist methods is related to how they quantify uncertainty. More specifically, when defining uncertainty intervals for model predictions, Bayesian methods give rise to *credible intervals* while frequentist methods generate *confidence intervals*. Although having similar names, the meaning and interpretation of these statistical concepts are profoundly different. This is connected to the inference problems these two approaches to statistics are seeking to answer. The Bayesian inference problem is the following: given a set of model parameters θ and observed data D, what values of θ are reasonable given the observed data? On the other hand, the inference question asked by the frequentist approach is: are the observed data D reasonable, given the hypothesised model parameter θ? This results into two different approaches to the modeling process: while frequentists consider θ to be fixed and D to be random, Bayesians consider θ to be random and D to be fixed [169]. This gives rise to the different interpretation the credible and confidence intervals have: computing a 95 % credible interval is equivalent to stating "given the observed data D, there is 95% probability that the model parameter θ will fall within the credible interval". On the other hand, since in the frequentist approach the parameters are fixed, while the data is random, computing a 95% confidence interval means stating "If the data-generating process that produced D is repeated many times, the computed confidence interval will include the true parameter θ 95% of the times". Therefore, while the Bayesian solution represents a probability statement about the parameter values, given fixed interval bounds, the frequentist solution concerns the probability of the bounds, given a fixed parameter value. Credible intervals capture the uncertainty in the parameter values and can therefore be interpreted as probabilistic statements about the parameter. Conversely, confidence intervals capture the uncertainty about the computed interval (i.e., whether the interval contains the true value or not). Thus, it is not possible to interpret confidence intervals as probabilistic statements about the true parameter values. The frequentist method is not wrong, it's just answering a different question, that usually tells us nothing about the specific dataset we have observed. In fact, when dealing with one specific dataset (which is usually the case in the M&V setting), all that a confidence interval can say is: "for this specific dataset, the true parameter values are either contained or not contained in the confidence interval".

In order to better clarify these concepts, a practical example for the energy baseline modeling case will be presented. We suppose that two energy baseline models have been estimated, one Bayesian and one frequentist, with the corresponding uncertainty intervals. In the Bayesian case we're able to state the following: given the training data observed (energy consumption and outdoor temperature time-series), there is 95% probability that the estimated model parameters (and therefore the predictions) will fall within the calculated credible interval. On the other hand, the corresponding frequentist statement would be: if we could repeat many times the data-generating process that lead to the energy consumption and outdoor temperature values that we observed, 95% of the times the estimated model parameters would be included in the calculated confidence interval. It is evident that, for the case in analysis, the Bayesian interpretation is the one which is the most reasonable and coherent. In spite of this, it is fair to state that for many common problems, such as linear regression, the Bayesian and frequentist intervals can coincide. This is also the reason why, in many contemporary scientific studies, Bayesian interpretation is (erroneously) applied to frequentist confidence intervals. Unfortunately, this match many times ceases when the models start becoming more complex than standard linear regression [170]. An additional feature of Bayesian credible intervals is that, while confidence intervals are delimited by two (random) numbers, credible intervals are represented by probability density functions, which are visibly more suited to risk assessment or uncertainty quantification problems [146].

In many Bayesian studies, including this research work, credible intervals are represented as highest posterior density intervals (HDI or HPDI), defined as the narrowest interval containing the specified probability mass. This means that a 95% HDI will be represented by the narrowest interval containing 95% of the probability mass. Frequentist confidence intervals are equal-tailed, since they are generally computed by adding or substracting a fixed number from the mean. While this can work for symmetrical distributions, real data is often asymmetrical and frequentist equal-tailed intervals will lead to including unlikely values on one side, while excluding more likely values on the opposite. HDIs solve this problem by providing local adaptive uncertainty bands with equally likely (but not symmetrical) lower and upper bounds [70].

4.2.2 Model comparison

In this section, the structure of the model comparison that was performed is discussed. The comparison was carried out in four consecutive test phases, that are summarized in Table 4.1. The specifications which are tested in each phase are highlighted in the table. The first model run was the benchmark model, to which the others were compared. In the second phase the goal was to test the use of regularizing priors, while in the third phase the inclusion of a wind speed feature in the regression model was assessed. Finally, in the fourth phase, two different techniques for the posterior estimation were tested, and a comparison was performed between regular regression and multilevel regression. In each consecutive phase, the best performing specification previously tested was then implemented.

Table 4.1: Model comparison overview.

Phase	Prior	Wind speed feature	Posterior estimation	Pooling
1	Uninformative	No	ADVI	No pooling
2	**Regularizing**	No	ADVI	No pooling
3	Best according to previous results	**Yes**	ADVI	No pooling
4	Best according to previous results	Best according to previous results	**ADVI, NUTS**	**Partial pooling, No pooling, Complete pooling**

Model comparison phases

Phase 1

The first test was run using the benchmark regression model with time and temperature features and uniform priors. The coefficients were supposed to have different values according to the detected load profile cluster, but no pooling was performed between the different clusters. The posterior distributions were estimated using the ADVI technique. This is the benchmark model specification to which the following versions were compared.

Phase 2

In the second test, the same regression model of Phase 1 was used, but this time with regularizing priors in place of the flat priors previously used. Again no pooling was performed, and the posterior distributions were estimated using

ADVI. This phase was aimed at understanding whether the use regularizing priors, selected using domain knowledge about building energy consumption modeling, could improve the prediction accuracy.

Phase 3

In the third set of model tests, a wind speed predictor was added to the linear regression model, as described in Section 4.2.1. Regarding the prior choice, it was decided to use whichever priors were granting higher accuracy between the ones tested in Phase 1 and 2. The pooling and posterior estimation techniques used were the same used in the previous phases. This model comparison phase was aimed at understanding whether adding a wind speed feature could improve the prediction accuracy.

Phase 4

In the last model comparison phase, the regression model providing the best accuracy between the one including and excluding the wind speed predictor was used. The best performing priors, already used in Phase 3, were selected again. Different pooling techniques were now compared (partial pooling, no pooling, complete pooling), as well as two different methods to estimate the posterior distributions: ADVI and a HMC sampling through the NUTS algorithm. This phase had the goal of testing whether implementing multilevel regression in place of regular regression could bring added value in terms of model accuracy or uncertainty bands estimation for our use case. At the same time, it was possible to evaluate the goodness of the ADVI posterior approximation by comparing it to the one obtained with the sampling approach.

Comparison metrics

Three main metrics were used in order to compare the different models tested. The CV(RMSE) was used to assess the prediction accuracy of the model, while the coverage and adjusted coverage parameters were analyzed to evaluate the uncertainty bands estimated.

CV(RMSE)

The coefficient of variation of the root mean square error (CV(RMSE)) is a metric frequently employed to evaluate the accuracy of energy baseline models:

$$CV(RMSE) = \frac{\sqrt{\frac{1}{N}\sum_i^N (y_i - \hat{y}_i)^2}}{\bar{y}} \times 100 \tag{4.22}$$

with N being the total number of hours of the time-series, y_i the measured energy consumption value on hour i, $\hat{y}_i$ the energy consumption predicted using the regression model, on the same hour, and $\bar{y}$ the average hourly energy consumption across the whole time-series. Equation (3.18) shows how the CV(RMSE) is nothing but a normalization of the root mean square error by the mean of the measured energy consumption values $\bar{y}$, representing a comparison between the model error and the average energy consumption values for the selected building or facility.

Coverage

In order to compare the accuracy of the uncertainty intervals estimated with the tested model specifications, the concept of coverage probability was used. When evaluating uncertainty intervals, the coverage probability represents the proportion of times that the real observed value is contained within the estimated interval.

$$Coverage = \frac{1}{N}\sum_{i=0}^{N} \psi_i \times 100 \tag{4.23}$$

where:

$$\psi_i = \begin{cases} 1 & \text{if } \hat{y}_i^{low} < y_i < \hat{y}_i^{high} \\ 0 & \text{else} \end{cases} \tag{4.24}$$

with N being the total number of hours of the time-series, y_i the measured energy consumption value on hour i, $\hat{y}_i^{high}$ and $\hat{y}_i^{low}$ the higher and lower bounds of the predicted uncertainty bands for hour i. One problem of this metric is that, in its evaluation, it doesn't take into account the size of the estimated uncertainty intervals. A disproportionately large interval would always have 100% coverage, but this can be an indication that the model that generated such interval is misspecified. In order to solve this issue a new metric, called adjusted coverage, was defined. This newly defined metric modifies the traditional coverage concept with additional terms that have the objective of penalizing larger intervals.

$$Adjusted\,coverage = \frac{1}{N}\sum_{i=0}^{N}\phi_i \frac{log(\hat{y}_i^{high})}{log(\hat{y}_i^{low})} \tag{4.25}$$

where:

$$\phi_i = \begin{cases} \frac{log(y_i - \hat{y}_i^{high})}{log(y_i)} + 1 & \text{if } y_i > \hat{y}_i^{high} \\ \frac{log(\hat{y}_i^{low} - y_i)}{log(y_i)} + 1 & \text{if } \hat{y}_i^{low} > y_i \\ 1 & \text{else} \end{cases} \tag{4.26}$$

This additional metric enables a more accurate comparison of the different Bayesian inference models tested. The traditional coverage concept, which simply checks whether the real value lies within the predicted uncertainty interval, does not provide enough information to correctly compare the bands. The adjusted coverage solves the issues of the coverage metric by adding two penalization terms. The first of these terms is included in ϕ_i, a variable representing whether or not the metered value y_i is contained in the interval. If y_i is within the interval, then ϕ_i is equal to 1, otherwise it's equal to 1 plus a term that is proportional to the distance between the closest bound and the metered value. In this way 'bigger errors' of the model are penalized more. The second penalization term is $\frac{log(\hat{y}_i^{high})}{log(\hat{y}_i^{low})}$, which is proportional to the size of the predicted uncertainty interval. Both terms are expressed as a ratio of logarithms, in order to prevent the adjusted coverage from reaching very high values when the metered y_i or the $\hat{y}_i^{low}$ are close to zero. This metric, although not suitable for a comparison between different buildings, serves well the purpose of comparing uncertainty intervals predicted by different models for the same building. Among models with equal coverage, those having smaller adjusted coverage are to be preferred, since this represents a lower overall model uncertainty.

4.3 Case Study

The presented approach was tested on an open dataset containing electricity meter readings at an hourly frequency for 1578 non-residential buildings. It was decided to test the approach on this dataset because of the generally recognized need, in the building performance research community, of testing novel techniques on common datasets [108]. This helps providing meaningful comparisons of accuracy, applicability, and added value between methodologies. The dataset is part of the Building Data Genome Project 2, a wider set that contains readings from 3,053 energy (electricity, heating and cooling water, steam, and irrigation) meters from 1,636 buildings. Of the original 1636 buildings, only the 1578 containing electricity meter readings were selected, since the proposed methodology has the goal of modeling hourly electricity consumption. A thorough description and exploratory data analysis of this dataset is available in Miller et al. [159]. A summary of the remarks that might be of interest follows:

- the buildings belong to 19 different sites across North America and Europe, with energy meter readings spanning two full years (2016 and 2017),
- there are 5 main primary use categories: education, office, entertainment/public assembly, lodging/residential, and public services,
- the weather data provided includes information about cloud coverage, outdoor air temperature, dew temperature, precipitation depth in 1 and 6 hours, pressure, wind speed and direction,
- for most of the buildings, additional metadata such as total floor area and year of construction is available.

Table 4.2 contains an overview of the sites present in the dataset and the number of buildings having electricity meter readings. Each site is assigned an animal-like site code name and each building is characterized by a *Unique Site Identifier* consisting of the site code name, an abbreviation of the building primary space usage, and a human-like name unique for each building. An example of a building's unique site identifier is: *Rat_health_ Gaye*.

In this case study, the year of 2016 was used to train the model. Validation of the model prediction accuracy was then performed on data from 2017. The

Table 4.2: Overview of the sites from which the meter data was collected and number of buildings having electricity meter readings for each site.

Site	**Actual Site Name**	**Location**	**Buildings**
Panther	Univ. of Central Florida (UCF)	Orlando, FL	105
Robin	Univ. College London (UCL)	London, UK	52
Fox	Arizona State University (ASU)	Tempe, AZ	137
Rat	Washington DC - City Buildings	Washington DC	305
Bear	Univ. of California Berkeley	Berkeley, CA	92
Lamb	Cardiff - City Buildings	Cardiff, UK	146
Eagle	Anonymous	N/A	106
Moose	Ottawa - City Buildings	Ottawa, Ontario	13
Gator	Anonymous	N/A	74
Bull	Univ. of Texas - Austin	Austin, TX	123
Bobcat	Anonymous	N/A	35
Crow	Carleton Univ.	Ottawa, Ontario	5
Wolf	Univ. College Dublin (UCD)	Dublin, Ireland	36
Hog	Anonymous	N/A	152
Peacock	Princeton University	Princeton, NJ	45
Cockatoo	Cornell University	Cornell, NY	117
Shrew	UK Parliament	London, UK	9
Swan	Anonymous	N/A	19
Mouse	Ormand Street Hospital	London, UK	7

CV(RMSE), coverage and adjusted coverage were calculated between 2017 model predictions and observed meter readings.

4.4 Results

The Results section is divided in two parts. First, the model comparison results are analyzed and the model specifications that produced the highest prediction accuracy are pointed out. Then, the results obtained for two individual buildings are presented in detail, highlighting the outputs of the methodology, such as the clustering of the daily profiles, the change-point temperatures and coefficients detected, as well as predictions for the full test year. All the coverage and adjusted coverage estimations provided in this section are referred to 95% HDIs. Regarding the software used for the analysis, the consumption profile clustering was performed with the kernlab library [171] in the R programming environment, while the Bayesian models were calculated using the python library PyMC3 [172].

4.4.1 Model comparison

The results obtained for each of the model comparison phases are presented here in form of boxplots for the CV(RMSE), coverage and adjusted coverage variables. It's important to note that the dataset that was selected for the case study contains various buildings for which the test year data is very different from the training year data. Examples of this are buildings with flat consumption in the whole test year or in large parts of it, as well as buildings having completely different energy consumption trends in the training and test year, meaning that no baseline model could provide accurate predictions. In Figure 4.2, the electricity consumption time-series of one of these 'outlier' buildings is shown. In order to not manually exclude any building from the analysis, while the results for all the buildings were calculated, in the boxplots the outlying values were hidden, meaning values 1.5 IQR above the upper quartile or below the lower quartile are not shown. The analysis performed in this section is based on median values and boxplot analysis that are not affected in any way by the exclusion of the outlier buildings from the plots.

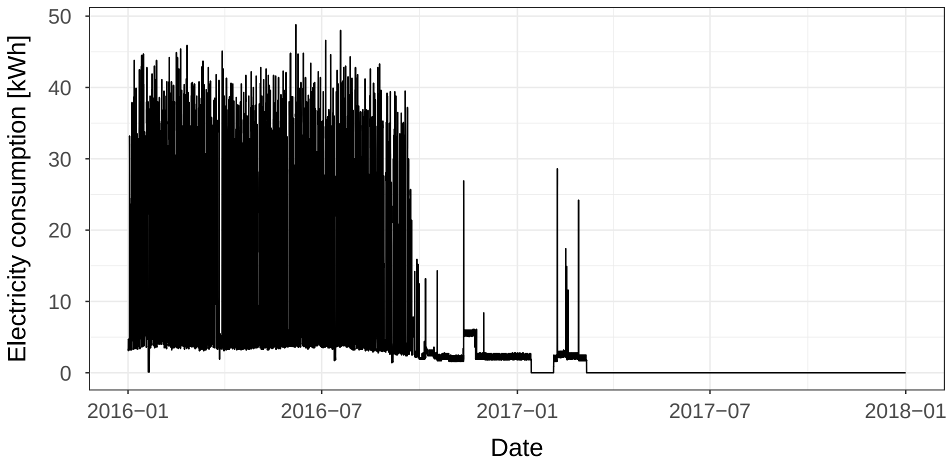


Figure 4.2: Electricity consumption time-series for the building *Lamb_ assembly_ Delilah*.

Figures 4.3 and 4.4 show the results in terms of CV(RMSE) and coverage for the first three phases of model comparison. In the second phase of the analysis, the benchmark model of Phase 1 was updated with regularizing priors in place of the uninformative ones previously used. In the third phase, the model from Phase 2, which proved to be better than the benchmark model both in terms of CV(RMSE) and adjusted coverage, was complemented with an additional term marking the effect of wind speed. From Figures 4.3 and 4.4, we can see that while the use of regularizing priors caused only a slight decrease of CV(RMSE), the uncertainty bands were highly affected by this change. A quick analysis of Figure 4.4 shows that, while the model with uninformative priors had the highest coverage, this was mainly due to the estimation of disproportionately large uncertainty bands. The addition of the wind speed term, on the other hand, seems to have no effect on neither the CV(RMSE) nor the coverage of the model, meaning that for this dataset using wind speed as a predictor variable for electricity consumption does not provide any benefit.

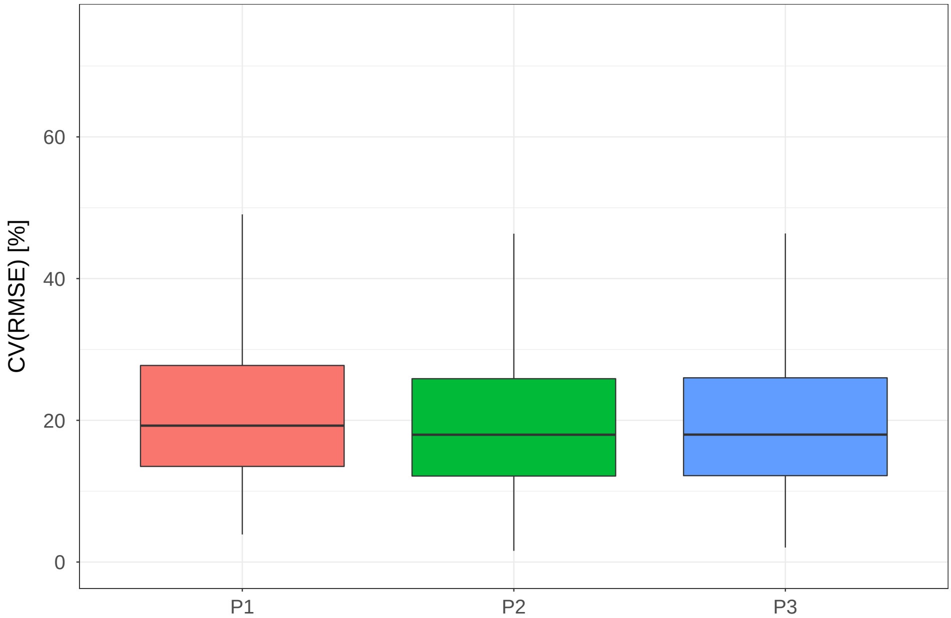


Figure 4.3: CV(RMSE) results obtained in Phases 1,2 and 3 of model comparison.

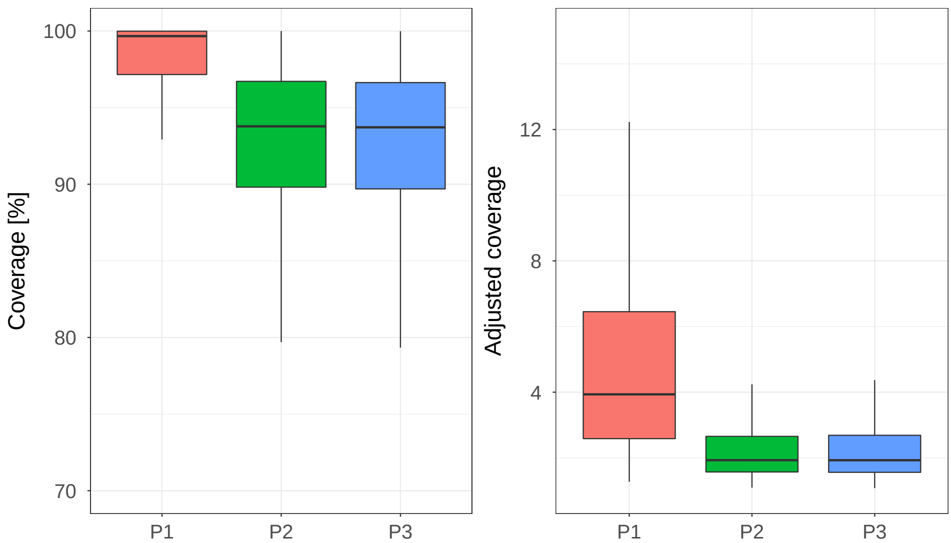


Figure 4.4: Coverage and adjusted coverage for the 95% HDIs obtained in Phases 1,2 and 3 of model comparison.

The first three phases of model comparison helped prove that the regression model and prior distributions used in Phase 2 were the ones providing the highest accuracy with the smallest number of predictor variables. Phase 4, on the other hand, was aimed at testing the performance of multilevel regression and different posterior estimation techniques. In Phase 4, the model of Phase 2, originally characterized by the use of a *no pooling* approach and ADVI estimation, was compared with specifications that used the same regression model and prior distributions of Phase 2, but implementing complete and partial pooling as well. The use of the NUTS (MCMC sampling) algorithm for the estimation of the posterior distribution was also tested, using 2000 tuning steps and then sampling 4 chains with 5000 samples each. Figures 4.5, 4.6 and 4.7 show the boxplots obtained for the final phase of model comparison, while Figure 4.8 allows to simultaneously analyze the CV(RMSE) and adjusted coverage metrics. The first clear conclusion that can be drawn is that complete pooling is the least performing of the three pooling approaches, which is coherent with the method used to build the regression model (giving high value to the clustering approach to detect different consumption trends). At the same time, the results obtained show that the use of partial pooling marks a definite coverage improvement over no pooling for the NUTS estimation case, while when using ADVI the two models are practically equivalent. When comparing the posterior estimation techniques, the use of MCMC sampling yields comparable CV(RMSE) and slightly better adjusted coverage over the variational inference approach: Figure 4.8 provides a helpful visualisation in this regard.

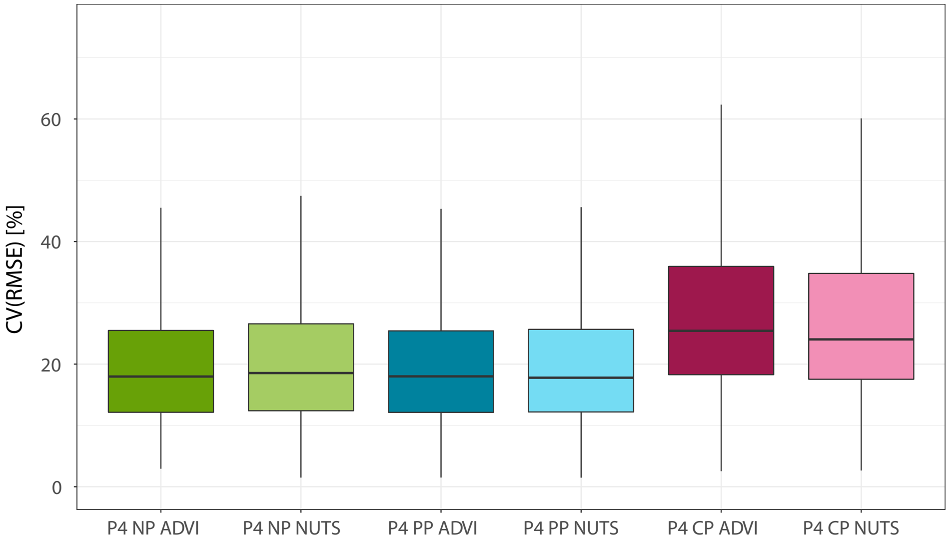


Figure 4.5: CV(RMSE) results obtained in Phase 4 of model comparison.

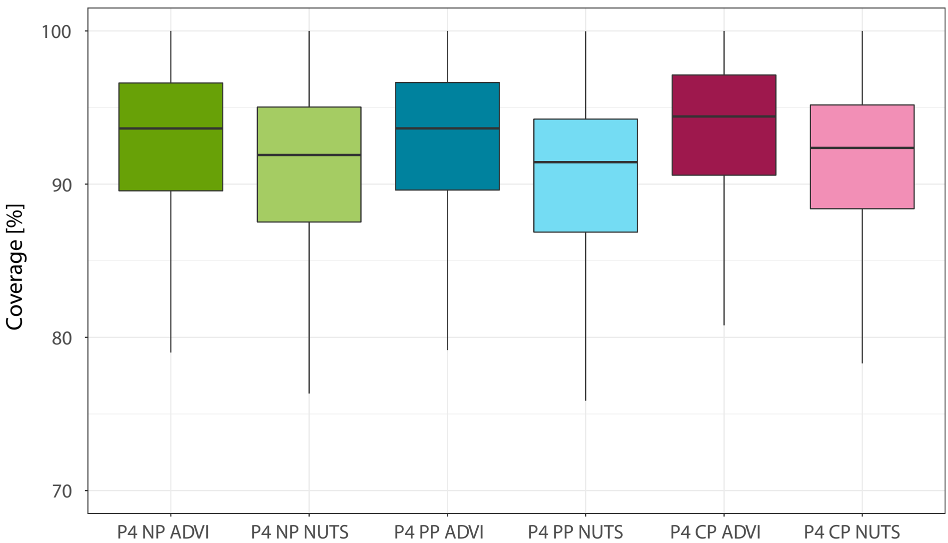


Figure 4.6: Coverage for the 95% HDIs obtained in Phase 4 of model comparison.

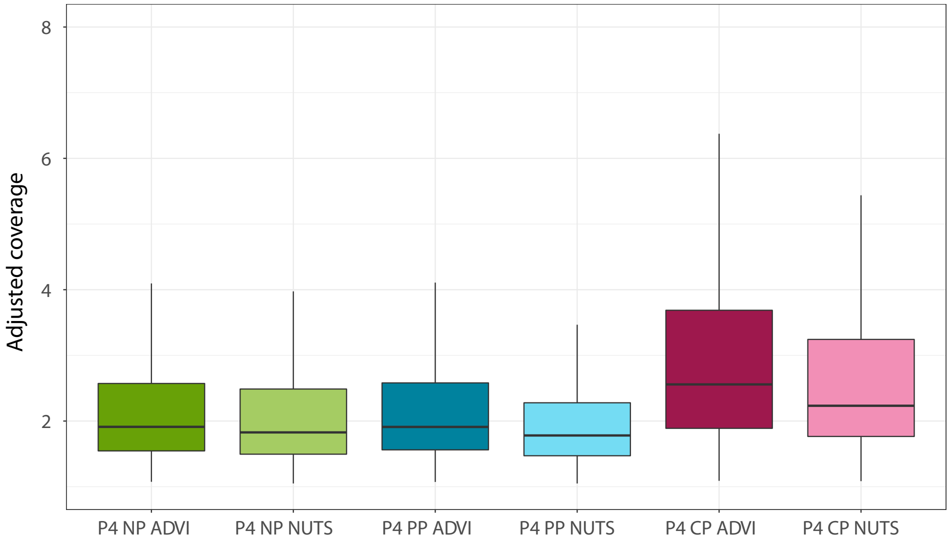


Figure 4.7: Adjusted coverage for the 95% HDIs obtained in Phase 4 of model comparison.

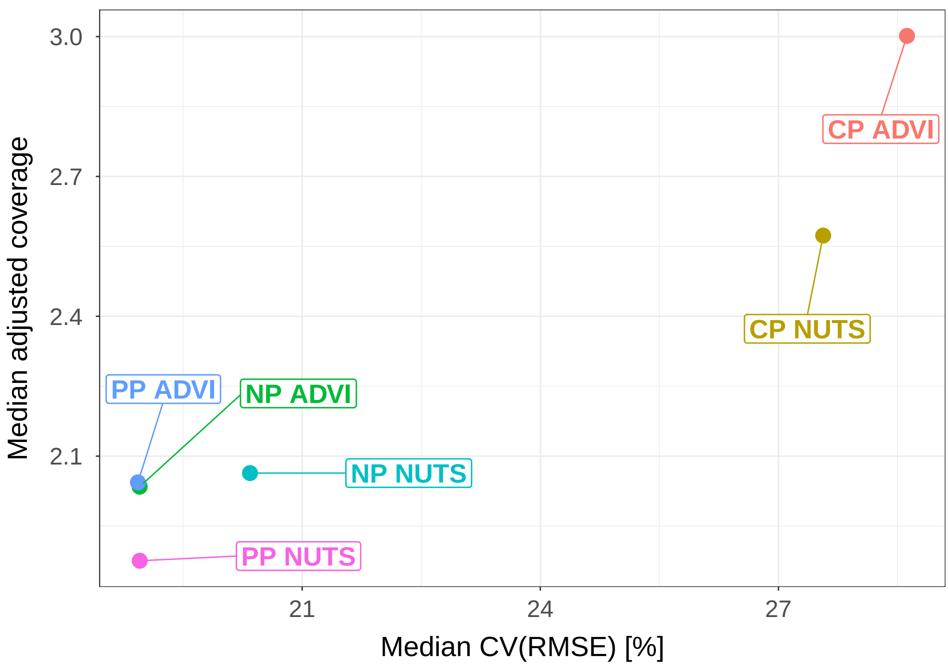


Figure 4.8: Median CV(RMSE) vs median adjusted coverage obtained in Phase 4 of model comparison.

Regarding the computational time difference between ADVI and NUTS, the analysis was run on a remote server with 125 GB of RAM and an Intel Xeon Processor with 2.60GHz base frequency, using 12 of the 32 available cores. The average computational times required for the analysis are summarized in Table 4.3, which presents also the median CV(RMSE), coverage, and adjusted coverage obtained with each model of phase 4. It appears that NUTS estimations require around 35 times more computational power than the ADVI case, while providing only modest improvements in terms of adjusted coverage. It's worth mentioning that, for the no pooling and partial pooling cases, the computational time required to model one building is directly proportional to the number of different load profile patterns detected by the clustering algorithm, since each additional load profile increases the number of variables that need to be estimated in the model.

Table 4.3: Results obtained in Phase 4 of model comparison. Median CV(RMSE), coverage and adjusted coverage are shown, as well as the average computational time required to estimate the model for a single building of the dataset.

Model	**CV(RMSE) (%)**	**Coverage (%)**	**Adjusted coverage**	**Computational time**
NP ADVI	18.95	92.54	2.03	15 seconds
PP ADVI	18.93	92.61	2.04	16 seconds
CP ADVI	28.62	93.25	3	9 seconds
NP NUTS	20.34	90.82	2.06	8 minutes
PP NUTS	18.95	89.94	1.87	9 minutes
CP NUTS	27.56	91.18	2.57	1.5 minutes

4.4.2 Individual buildings

In this section, the model results obtained for two individual buildings are analyzed, in terms of consumption predictions, weather dependence and posterior distributions identified for the model parameters. The results shown are the ones obtained using partial pooling and ADVI posterior estimation.

The first building analyzed is *Rat_health_Gaye*, a healthcare facility located in Washington DC with a total floor area of 2220 square meters. The five recurrent load profiles identified by the clustering algorithm for this building are represented in Figure 4.9, with the red lines representing the profile centroids and the black lines the load profiles of days belonging to that cluster. In Figure 4.10, the electricity consumption time-series for both the training and test years (2016 and 2017) is shown.

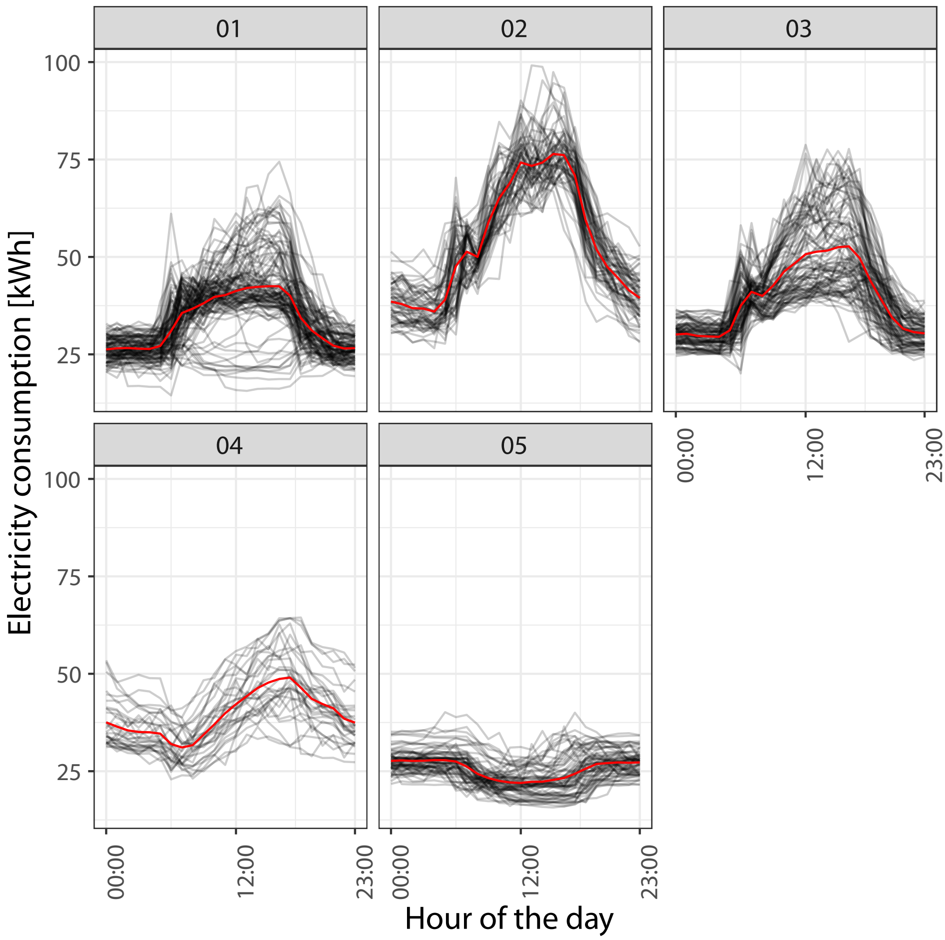


Figure 4.9: Recurrent daily load profiles identified for the building *Rat_ health_ Gaye*.

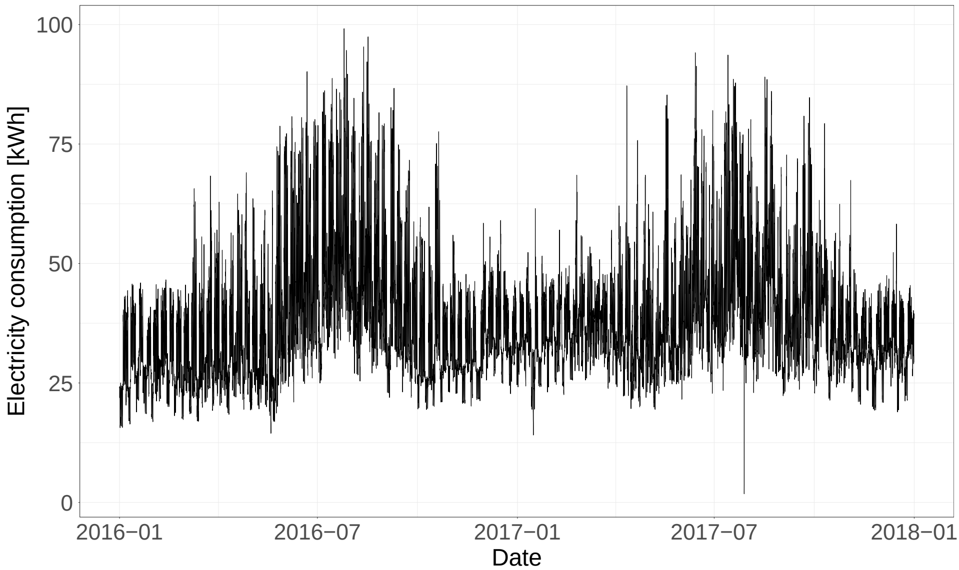


Figure 4.10: Electricity consumption time-series for the building *Rat_ health_ Gaye*.

At first glance, it appears that this building is using electricity for the cooling system, while a different energy source is used for heating, since the consumption is quite constant during winter months and peaks in the summer period. An analysis of the change-point temperatures estimated by the Bayesian regression model can help confirm this hypothesis. Figure 4.11 shows the posterior distributions estimated for the heating and cooling change-point temperatures, both the mean value and the 94% HDIs are marked.

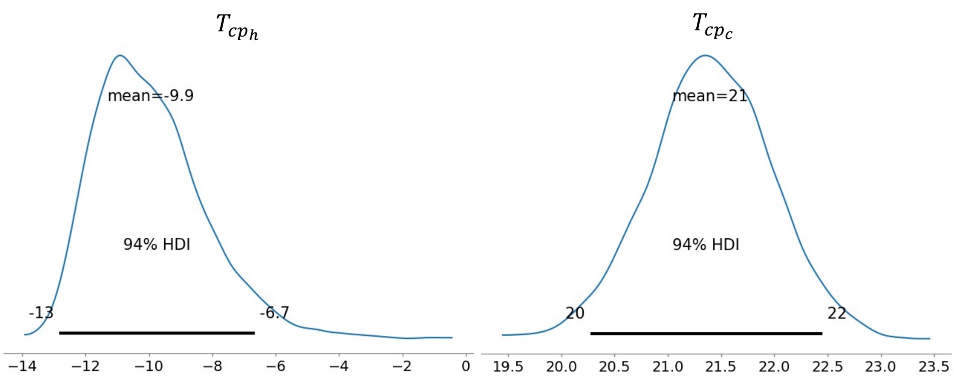


Figure 4.11: Posterior distributions estimated for the heating and cooling change-point temperatures for the building *Rat_ health_ Gaye*.

In order to better understand these results, it can be useful to visualize the electricity consumption and outdoor temperature time-series, together with the heating and cooling change-point temperatures estimated by the model. Figure 4.12 shows this relationship for the training year, with the estimated mean heating and cooling change-point temperatures marked as a dotted line, and their 94% HDIs represented by the shaded area. This plot confirms the initial hypothesis and validates the results of the model. The building has cooling dependence: it can be seen how the electricity consumption starts increasing once the outdoor temperatures exceed the estimated cooling change-point temperature. At the same time, the model detected no heating dependence, showing that the heating change-point temperature would correspond to the lowest temperature observed in the training year. This conclusion is confirmed by studying how the electricity consumption does not react to changes in temperature below T_{cp_c}.

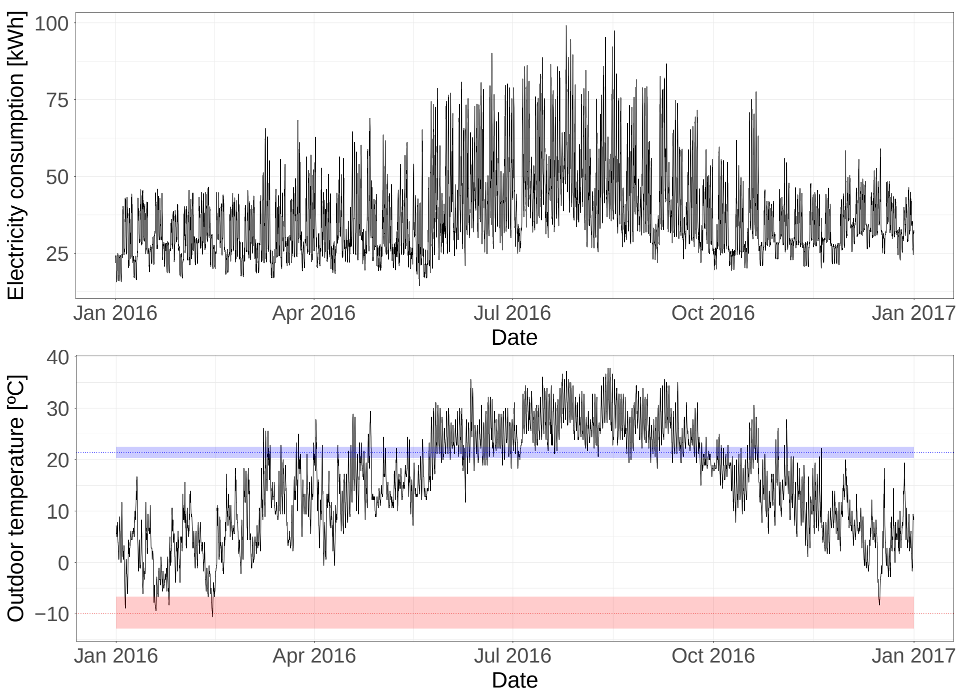


Figure 4.12: Electricity time-series vs outdoor temperature values for the building *Rat_health_Gaye*. The heating and cooling change-point temperatures estimated by the model are shown in red and blue with the corresponding 94% HDIs.

In Figure 4.13 the electricity consumption time-series of the test year is shown (in black), together with the model predictions and 95% uncertainty bands (in red). In order to better visualize the uncertainty bands and predictions, Figure 4.14 shows a segment of this same plot that includes only predictions for the month of July 2017, where the building energy consumption is being affected by the outdoor temperature values. Overall, the model is able to accurately predict the consumption time-series of the test year and to capture the temperature dependence dynamics of the building. This model was characterized by a CV(RMSE) of 15.2%, a coverage of 96.7% (for the 95% HDI) and an adjusted coverage of 1.94.

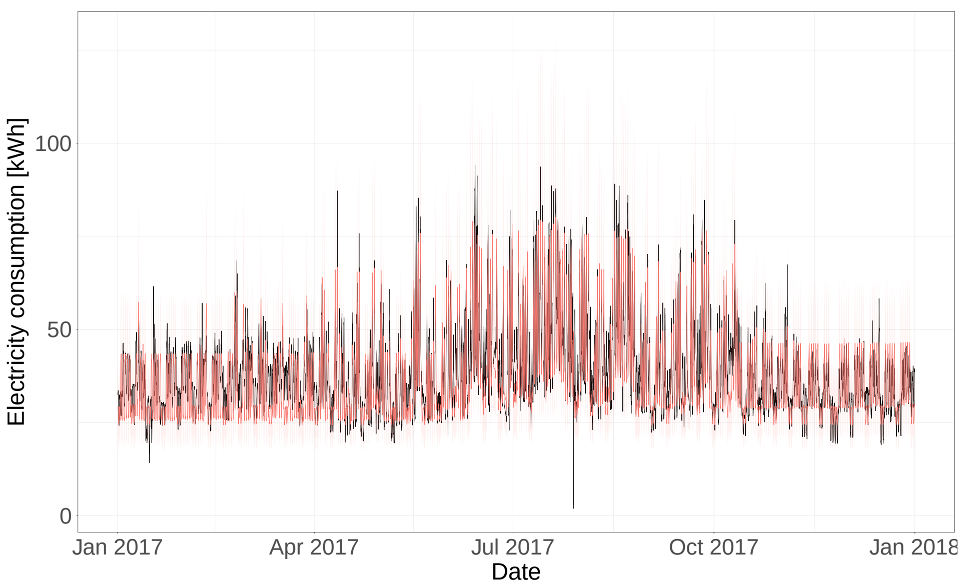


Figure 4.13: Test year metered electricity time-series (in black) and model predictions (in red) for *Rat_health_Gaye*.

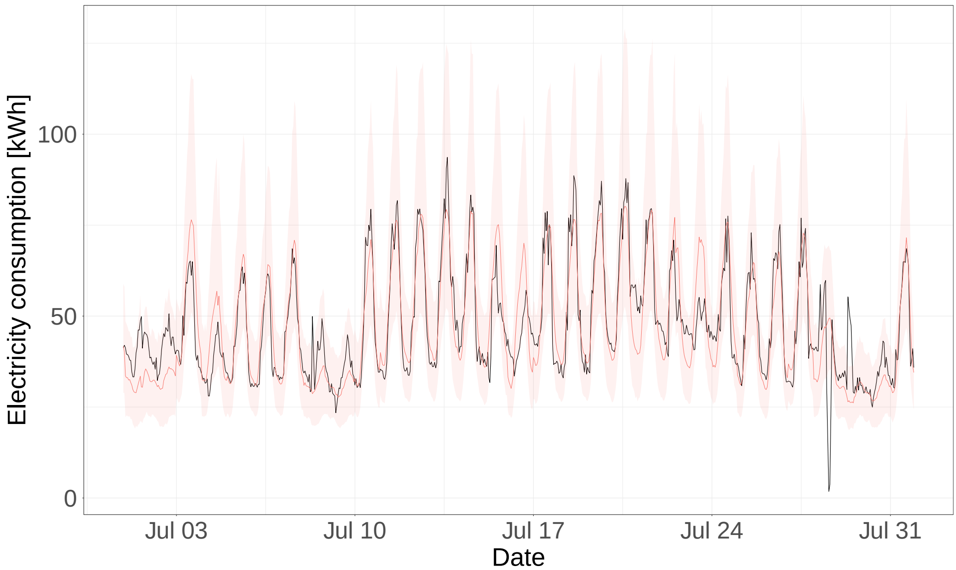


Figure 4.14: July 2017 metered electricity time-series (in black) and model predictions (in red) for *Rat_ health_ Gaye*.

The second case presented is that of a building which has both heating and cooling consumption dependence. The building analyzed is *Rat_ education_ Royal*, a school in Washington DC with a total floor area of 7218.6 square meters. The building electricity consumption time-series for the training and test year is shown in Figure 4.15, where it's possible to observe that the consumption is peaking both in summer and winter months, a tendency which is more pronounced in the training year than in the test year.

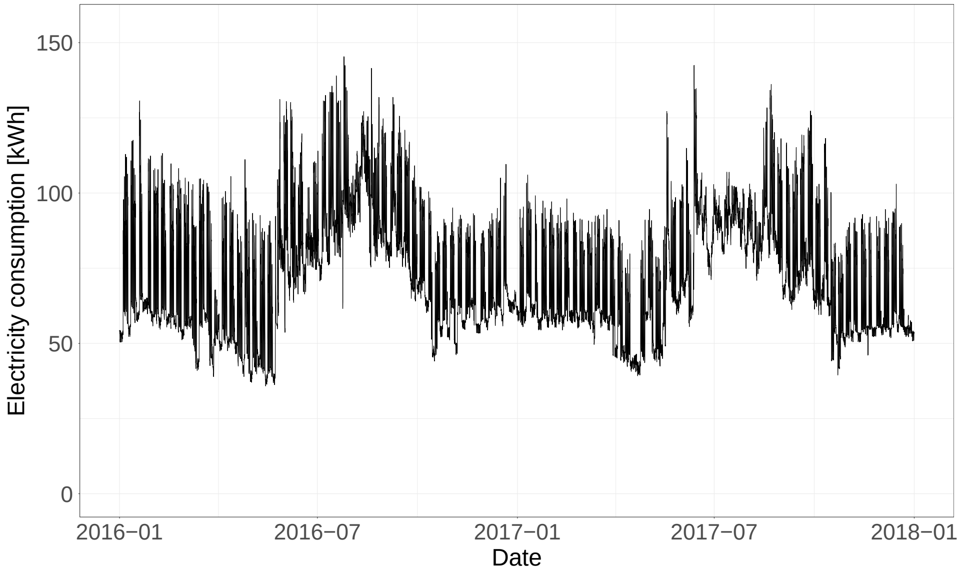


Figure 4.15: Electricity consumption time-series for the building *Rat_ education_ Royal*

Figure 4.16 shows the posterior distributions estimated by the model for the heating and cooling change-point temperatures T_{cp_h}, T_{cp_c} and for the linear coefficients β_h, β_c. The estimated heating impact on the consumption is lower than the cooling one, with a mean β_h of 0.0066 vs a mean β_c of 0.021. Another interesting point is that the cooling dependence posterior has a multimodal distribution, while the heating dependence posterior is unimodal. Since the β_h and β_c were supposed to be changing according to the part of the day, as explained in Section 4.2.1, this characteristic of the posteriors means that while the heating dependence was estimated to be very similar for each day-part, the impact of cooling on the overall consumption was detected to be stronger in certain parts of the day rather than others.

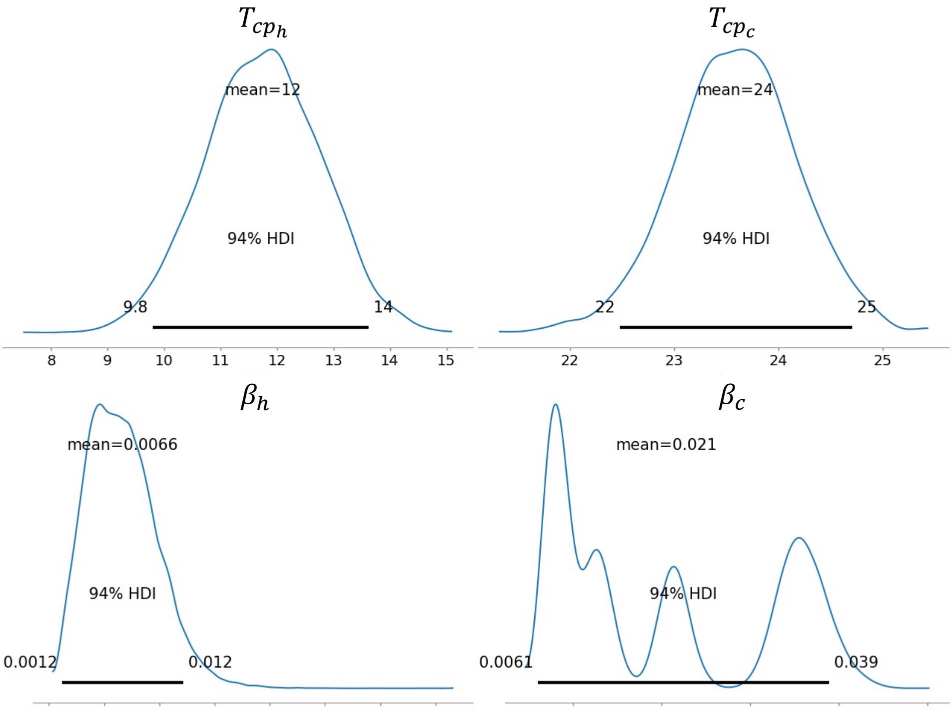


Figure 4.16: Posterior distributions estimated for the heating and cooling change-point temperatures (upper panels) and for the heating and cooling linear dependence coefficients (lower panels) in *Rat_ education_ Royal*.

In Figure 4.17, the electricity consumption time-series is represented, together with the outdoor temperature values and the T_{cp}. From this graph, the heating and cooling dependence of the building appears evident, and it's also possible to notice the increased effect of cooling over heating depicted by the estimated β_h and β_c, by comparing the consumption peaks with the corresponding temperature differences.

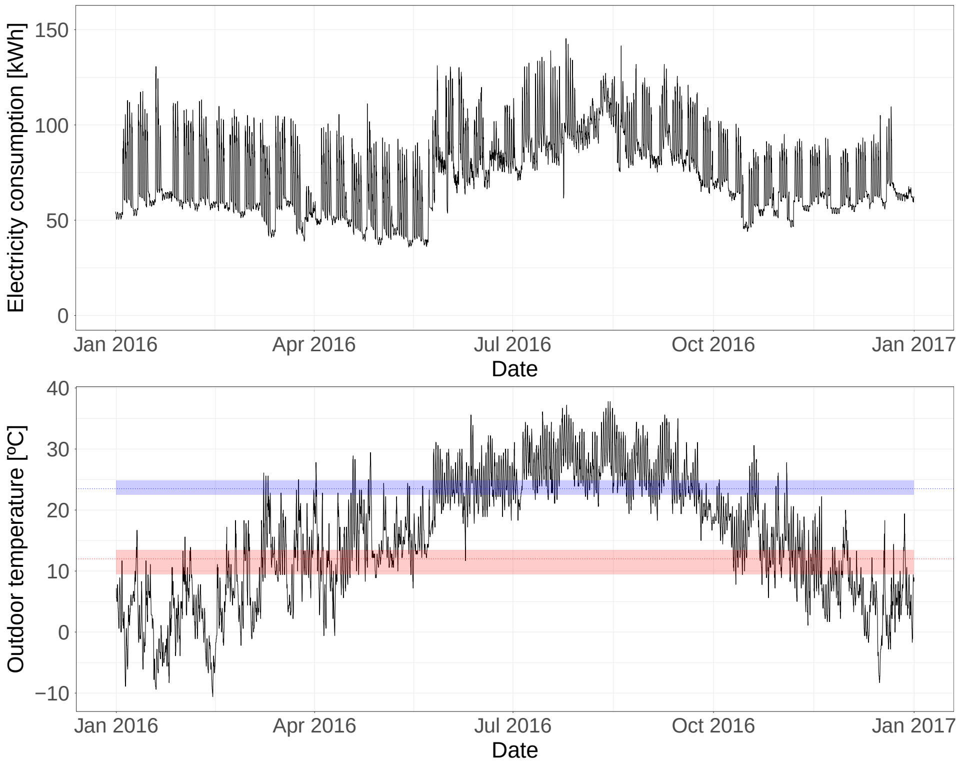


Figure 4.17: Electricity time-series vs outdoor temperature values for the building *Rat_education_Royal*. The heating and cooling change-point temperatures estimated by the model are shown in red and blue with the corresponding 94% HDIs.

In Figure 4.18, the electricity consumption time-series of the test year is shown (in black), together with the model predictions and 95% HDI (in red). Overall, the model provides a good fit to the metered electricity consumption time-series, with most of the data points being included within the 95% HDI. The model was characterized by a CV(RMSE) of 10.9%, a coverage of 95.2% (for the 95% HDI) and an adjusted coverage of 1.55.

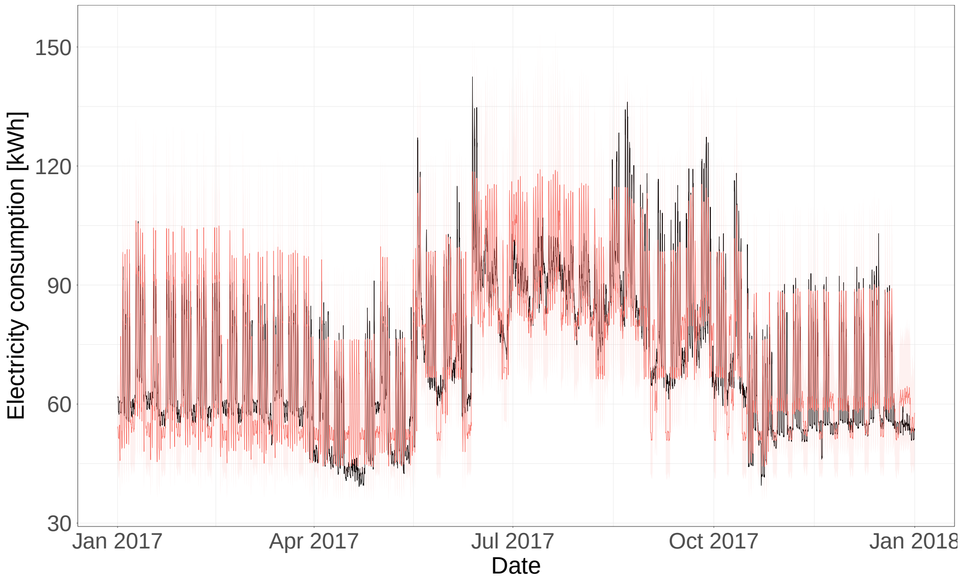


Figure 4.18: Test year metered electricity time-series (in black) and model predictions (in red) for *Rat_ education_ Royal*.

4.5 Discussion

The results obtained in the four model comparison phases provided important insights into the predictive capabilities of various model specifications. The first three phases enabled an analysis of the effects of different prior distributions and of the potential improvement generated by the inclusion of wind speed as a predictor variable in the model. The last phase of comparison used the best performing priors and regression model identified in the previous phases in order to test the effect of multilevel regression and different posterior estimation techniques. The main conclusions that can be drawn from the results obtained are the following:

- The use of regularizing priors based on building physics knowledge improves the model both in terms of CV(RMSE) and coverage.
- The addition of a regression term to take into account the wind speed did not improve the model predictive capabilities.

- The use of MCMC sampling techniques to estimate the posterior distribution yields comparable results to the variational inference method, despite being characterized by a more than 30-fold increase in computational time.

- When looking at the ADVI case, partial pooling has an almost negligible effect on the prediction accuracy, providing only a very modest median CV(RMSE) and coverage improvement, and a slightly worse adjusted coverage.

- In the NUTS case, the partial pooling regression seems to have a stronger impact, improving the CV(RMSE), but at the same time reducing the coverage of the model.

- The computational requirements of the partial pooling regression are comparable to the ones of the no pooling case.

The individual building analysis also provided interesting insights that showcase the strengths of the proposed methodology. The results of the profile clustering algorithm and of the Bayesian regression model unlock insights that can be used to depict a clear image of the consumption habits of the building in analysis. The results allow for a detailed characterization of the energy consumption trends, including the typical daily load profile patterns, the heating and cooling change-point temperatures, as well as the impact of the heating and cooling terms on the total electricity usage. For the buildings analyzed, the posterior distributions of the model coefficients obtained, shown in Figure 4.11 and 4.16, accurately represented the trends depicted by the consumption and outdoor temperature time-series. These posteriors allow to gain insight on the weather dependence of the analyzed buildings, unlocking actionable information in terms of energy performance. In Figures 4.12 and 4.17, it's possible to see how the posteriors identified for T_{cp_c}, T_{cp_h}, β_c, β_h are actually an accurate representation of the relationship between outdoor temperature and electricity consumption for the two showcased buildings. Such results open very interesting possibilities in terms of energy management improvement, as well as recommendations for energy retrofit plans. Comparing change-point temperatures and heating and cooling coefficients from a portfolio of similar buildings can disclose actionable insights about the analyzed buildings, enable targeting and prioritization of energy renovation strategies, and help energy managers make decisions backed by data.

The model also proved to be very accurate at predicting the electricity consumption baseline for the showcased buildings, while also providing small uncertainty bands, as seen in Figure 4.13 and 4.18. The accuracy of this baseline means that this model can be implemented in several practical applications, such as anomaly detection, energy performance analysis, or dynamic measurement and verification of energy efficiency savings. In fact, valid estimations of the effects of implemented energy conservation measures are impossible without a model able to provide accurate baseline model predictions. Nevertheless, it's important to highlight that, because of the daily profile classification used, the proposed methodology is only able, in the form presented in this chapter, to detect hourly changes in consumption within days that kept an overall similar load shape compared to the one they had before the implementation of the measure. In the case of being interested in evaluating energy retrofit actions that completely altered the consumption profile of the building energy use, the present methodology should be coupled with an algorithm able to predict, depending on time and weather features, the consumption profile that a building would have had, on a certain day, if the energy retrofit measure was not applied. Such an algorithm was devised and presented in [173], and could be seamlessly coupled with the proposed methodology.

4.6 Conclusions and future work

In this research work, a Bayesian methodology to model electricity baseline use in tertiary buildings was presented. The methodology has the feature of being based on Bayesian linear regression with interpretable terms, while at the same time being able to accurately predict electricity consumption time-series with high granularity. In order for the proposed approach to work, the only required data are historical electricity consumption and outdoor temperature values. The approach is based on a data pre-processing phase, in which the training data is analyzed in order to identify recurrent electricity load profiles, a modeling phase, in which the Bayesian linear regression model is run, and a results analysis phase, in which the posterior distributions estimated by the model are used to obtain a characterization of the building energy consumption and to generate baseline predictions.

When performing Bayesian regression, the quality of the results can depend on many different factors, such as the prior distributions specified, the covariates included in the regression model, or the technique used to estimate the posterior distribution. In order to compare different possible model specifications, a model comparison strategy, structured in four consecutive phases, was devised. The methodology was tested on the Building Data Genome Project 2, an open dataset containing 3,053 energy meters from 1,636 non-residential buildings. Within this dataset, 1578 buildings having electricity meter readings at hourly frequency for the years 2016 and 2017 were selected. For each building of the dataset, the Bayesian regression model was trained on the first year of data and then validated on the second year. The model comparison stage provided valuable results: regularizing priors performed better than uninformative ones, while the wind speed predictor did not have any effect on the model. Regarding the posterior estimation, the use of an MCMC sampling technique provided comparable results to the variational inference case, despite the almost 30-fold increase in computational time required, while multilevel regression provided a slight improvement in terms of CV(RMSE) and adjusted coverage, with differences that are more evident when using the MCMC sampling technique to estimate the posterior. At the same time, a more in-depth analysis of the results presented for the two showcased individual buildings demonstrated that the proposed methodology is able to provide a detailed characterization of the analyzed buildings' energy use, as well as accurate baseline predictions, characterized by effective uncertainty intervals. The possibility of associating Bayesian credible intervals to the estimated posterior distributions represents a fundamental feature that allows to implement the results obtained from this methodology in risk assessments for energy renovation projects.

The results presented highlight several possible applications for the proposed methodology, including energy performance improvement, energy use intensity characterization, quantification of energy conservation measures and risk mitigation in energy retrofit projects. An accurate, non-intrusive and scalable methodology, such as the one presented in this research work, can help driving down measurement and verification costs for energy efficiency projects, hence increasing their feasibility and profitability. Furthermore, reliable real-time measurement and verification is a requirement for many innovative energy efficiency models in which payments are handed out only when the savings are demonstrated and verified. To conclude, four main strengths were identified for the presented approach:

1. the explainability of the model and the interpretability of its coefficients even for non-technical audiences;

2. an elegant, efficient, dynamic and coherent estimation of uncertainty, that makes it apt to be used in financial risk assessments of retrofit strategies;

3. the ability to provide a detailed building energy use characterization that can employed by energy managers to improve the performance of their facilities;

4. high scalability to big data problems, because of the low computational complexity and the limited data requirements.

The features presented for the proposed methodology make it appealing for many different real-world applications in the field of energy efficiency, but at the same time it is also evident that this research is still in an initial stage and more work is needed in order to refine the approach. Future work might involve the testing of different likelihoods, such as the Student-t likelihood, which might be helpful in specific cases where high resiliency to outliers is required. The intersection of this methodology with Bayesian Additive Regression Trees (BART) might also be tested, as well as the effect of different predictors, such as solar radiation, or the implementation of alternative clustering techniques. The posterior estimation techniques could also be an object of further study: understanding whether the quality of predictions can improve by increasing the number of MCMC samples would be of great interest, as well as a quantification of the computational power that would be required for such estimations. Finally, it would be valuable to see this methodology in action in a real-world measurement and verification protocol, in order to evaluate how it ranks compared to other similar methodologies built for this purpose.

Chapter 5

Concept methodology for recommendation of energy retrofitting strategies

5.1 Introduction

In this chapter, a concept methodology is presented, with the goal of recommending and prioritizing energy efficiency strategies in tertiary buildings by combining many of the techniques discussed in this thesis. This methodology was designed to be implemented in big data building repositories that are currently in development in the framework of the EN-TRACK and BIGG projects [174, 175]. As such, the methodology is intended as a recommendation for future work more than a completely developed study. The availability of organized, clean, and trustworthy building and consumption data is crucial to the correct implementation of the techniques presented. Therefore the present methodology is also meant to be a guideline to enable a fruitful data collection process for any platform aimed at building energy management and efficiency improvement.

Given the fundamental importance of data availability, the methodology was conceived as a scalable process in which the recommendations presented change depending on the amount of data available in the platform. The main goal was to create a conceptual methodology that can provide some preliminary results even in the initial stages of the data collection process, but that, as more and more data is fed into the repository, will improve the quality and the accuracy of the provided recommendations. The system is aimed at providing actionable information for decision-makers in various ways:

1. provide a rough estimate of savings potential for the analyzed buildings,

2. help understanding which measures are the best fit for a certain building and will yield the highest savings,

3. enable prioritization for a portfolio of buildings, identifying the buildings with the highest savings potential and for which the renovation strategies will produce the best results.

Recommendation systems have been the object of a wide range of research studies during the last three decades. Their application has also been extended to the field of smart sustainable cities, and a detailed review of these systems can be found in Quijano Sánchez et al. [176]. When it comes to energy efficiency, the main use-cases that have been explored are related to smart homes and user behavior recommender systems [177, 178]. While the general purpose and definition of recommendation systems is to address the problem of estimating ratings for items that have not been seen by a user [179], in the present research, we extend this concept to the problem of the energy efficiency management of a portfolio of buildings. The recommendation system that we describe and propose in this chapter is a tool that analyzes the results obtained by implementing different energy conservation strategies in a group of buildings or facilities, in order to predict their effect in new buildings and possibly recommend their implementation.

In the building energy modeling research community, several articles have addressed the problem of predicting the effect of energy retrofitting strategies in buildings by using deterministic techniques. A detailed review of these approaches is presented in Chapter 2 of this thesis. Unfortunately, a considerable drawback of such methodologies is their poor scalability, linked to the complexity of the building energy modeling process and their problematic generalizability. Data-driven techniques offer a different and innovative approach to this problem, but research in this field is still in an initial stage. A first interesting data-driven solution to this problem was provided by Walter and Sohn, who used the data contained in the United States Building Performance Database to build a multivariate regression model able to estimate changes in energy usage intensity (EUI) due to retrofit implementations and depending on building characteristics [104]. More recently, Miller and Meggers have studied the possibility of leveraging smart meter data and building characteristics to predict potential energy savings and characterize buildings, producing very interesting results and demonstrating that there is highly valuable information that can be brought to light by analyzing big data repositories of tertiary buildings [109, 180]. Finally, a significant contribution to this new research field has been provided by Xu et al., that implemented a tree-based machine learning algorithm to study a portfolio of commercial buildings located in the US, with the goal of predicting the effect of different retrofit options based on building characteristics and weather data [181].

In the rest of this chapter, we first describe the characteristics of the concept methodology proposed, then provide a mock-up of the recommendation system, to finally present some conclusions and future work guidelines.

5.2 Methodology

The recommendation system described in this chapter is based on the insights that can be inferred by analyzing energy conservation strategies implemented in peer buildings. It is built on different modules that can be implemented depending on the amount of comparison data available from other buildings. A flowchart describing the presented methodology is shown in Figure 5.1, where the models and calculations are highlighted in yellow and the provided results in green.

To begin, the building for which the recommendations are built is analyzed in order to exclude measures that are not compatible with its characteristics; this process is named compatibility check. Then, according to a list of similarity criteria specified by the user, the group of comparison buildings is defined in a process called similarity check. These buildings are then analyzed, together with the list of measures identified by the compatibility check, in order to understand which is the most appropriate recommendation module to be implemented.

Depending on the amount of data available from similar buildings, three independent recommendation modules can be implemented, that provide results with different levels of detail:

1. Building energy benchmarking → savings potential and area of focus
2. Retrofit effect prediction → estimation of expected energy savings by measure
3. Multi-label classification → list of relevant measures for the analyzed building

The first module implements benchmarking techniques in order to estimate the potential savings that can be achieved and the measure categories that would yield the highest results. The second module aims at quantifying, in the most accurate possible way, the impact that an energy renovation strategy will have on

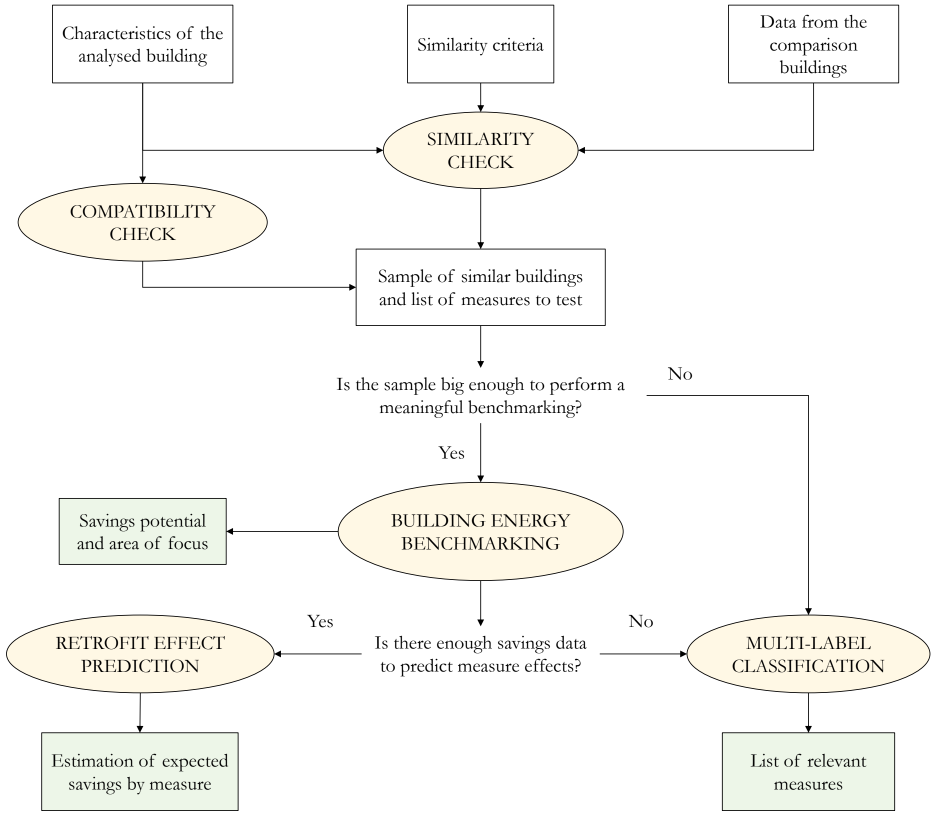


Figure 5.1: Recommendation methodology flowchart

the analyzed building. The third module has the objective of providing a list of popular implemented measures in similar buildings and is implemented in case of not having a big enough sample of similar buildings to implement the first or the second recommendation modules. For the rest of this section, the building energy benchmarking module will be referred to as Module 1, the retrofit effect prediction as Module 2, and the multi-label classification as Module 3.

5.2.1 Data requirements

This section briefly presents the data requirements for the three different recommendation modules previously introduced.

Building energy benchmarking

In order to successfully implement the first module of recommendation, the energy usage intensity (EUI) of all the buildings available in the platform needs to be calculated. This is represented by the ratio of the total energy load of the building and a normalization parameter of choice, usually the building's total floor area. The first module also depends on the results of a load segmentation and a characterization algorithm. The first separates the base-load of the building from the portions that depend on the heating and cooling operations, while the second is aimed at determining the change-point temperatures, the heating and cooling coefficients, as well as the load profile patterns of the buildings. Different load segmentation methodologies are present in literature [182], while an example of a data-driven characterization model that can be used for such purpose is the one introduced in Chapter 4.

Retrofit effect prediction and multi-label classification

The second and third recommendation modules are similar in their goal, but while the retrofit effect prediction represents a quantitative analysis aiming at predicting the potential savings that can be obtained by implementing a certain energy renovation strategy, multi-label classification is a simple qualitative analysis, usually carried out when there is not enough data available to perform the calculations of Module 1 and 2. The features used for these two recommendation modules are various, but they can be mainly divided into building features, usually information about the building that can be introduced by the building owner or manager, and temporal features, which can be extracted by analysing the energy consumption and weather data available. This feature characterization was first introduced by Miller and Meggers in [109]. Tables 5.1 and 5.2 present a list of the possible building and temporal features that might be used for retrofit effect prediction and multi-label classification. The features related to the measure effect in Table 5.2 are exclusively used in Module 2, and their availability marks the main distinction between the recommendations provided in the second and third recommendation modules. A short description of the temporal features presented in Table 5.2 follows:

Mean, max, min EUI: energy usage intensity, corresponding to the average, maximum, and minimum energy load of the building, divided by a normalization parameter of choice, usually the total floor area of the building. Can be calculated at hourly, daily, or monthly levels.

EUI variance: variance of daily minimum and maximum consumption values. It can be calculated across the whole time-series and separately for the heating and cooling seasons.

Load ratio: the ratio of the minimum and maximum load. It can be calculated at hourly, daily, or monthly levels.

Heating and cooling change-point temperatures: the outdoor temperatures from which a significant relationship between the building's energy consumption and the outdoor temperature conditions is detected.

Heating and cooling dependence: the impact of temperature differences above and below the change-point temperatures on the total energy load of the building.

Minimum and maximum consumption hours: hours at which the minimum and maximum consumption occur, can be calculated across the whole time-series and separately for the heating and cooling season.

Pattern volatility: a metric that analyzes how consistent are the load profile patterns of the building are over the course of a whole year. A building that frequently changes daily load shapes during the year is considered more volatile.

Load shape category: the building load profile patterns can be assigned a different category depending on their shapes, then these categories can be used to detect buildings having similar patterns.

Average annual number of days with temperature within different ranges: different ranges of temperatures can be defined (i.e. less than -10 ºC, between -10 and 0 ºC, between 0 and 10 ºC, between 10 and 25 ºC, between 25 and 35 ºC, more than 35 ºC), the number of days per year belonging to each temperature range will help recognize regions with similar climate.

Average monthly precipitation, cloud cover, and solar radiation: more weather variables that can help classify climate regions.

Measure effect: the data entries in this category are aimed at estimating the effect of a certain implemented measure on the buildings where it was implemented. For each measure implemented, the size of the investment is registered, together with the building EUI before and after the measure implementation, and the total energy savings achieved. Savings directly calculated with measurement and verification techniques, such as the one presented in Chapter 3 are given a higher weight than deemed or estimated savings manually introduced.

Table 5.1: Potential building features used in the recommendation system

General data	Construction data	Insulation	Heating and cooling	Lighting	Generation	Schedule
Primary use category	Type of construction	Window U-factor	Has heating system	Lighting type	PV capacity installed	Daily opening hours
Primary use subcategory	Historic building	Wall insulation	Has cooling system	Intelligent lighting system	Area available for PV installation	Weekly days off
City	Year of construction or full renovation	Roof insulation	Heating system energy source			Yearly holidays
Climate region	Total floor area	Solar shading installed	DHW energy source			
Energy efficiency certificate	Heated floor area					

Table 5.2: Potential temporal features used in the recommendation system

Consumption	Heating and cooling	Load shape	Weather variables	Measure effect
Mean, max, min EUI	Heating change-point temperature	Minimum and maximum consumption hours	Average annual number of days with temperature within different ranges	Implemented measure
EUI variance	Cooling change-point temperature	Pattern volatility	Average monthly precipitation	Investment
Load ratio	Heating and cooling dependence	Load shape category	Average monthly cloud cover	Energy savings obtained
			Average monthly solar radiation	Pre and post measure EUI

5.2.2 Checks

Similarity and compatibility checks are two pre-processing calculations that have the goal of increasing the robustness of the provided results. While the compatibility check is aimed at discarding potential recommendations that are not compatible with the characteristics of the building, the similarity check has the goal of selecting the most accurate similar building sample from the buildings in the repository, according to user-specified similarity criteria.

Compatibility check

The compatibility check is based on the concept that, when providing recommendations within a platform, robustness should be ensured at all times. Giving unrealistic recommendations that are not compatible with the characteristics of the building is likely to reduce user engagement and cause skepticism towards the results provided by the platform. For this reason, before performing any kind of modeling or calculation, the compatibility of the list of analyzed measures with the characteristics of the building is checked. The list of compatibility decisions includes, but should not be limited to, the following statements:

1. Exclude measures related to fuel-based heating systems if the building uses electricity for heating.
2. Exclude heating or cooling measures if the building has no heating or cooling dependence.
3. Exclude lighting measures if energy-efficient lighting is already implemented.
4. Exclude insulation measures depending on the last date of full refurbishment of the building.
5. Exclude PV installation recommendation if there is no available area for it.
6. Exclude scheduling changes if the schedules have already been optimized.

Similarity check

The similarity check is aimed at the selection of the most similar sample of buildings among the ones available in the repository. Users should be able to change the similarity criteria that they want to use, as well as the minimum number of buildings selected for comparison. Similarity criteria that can be used include:

1. Building category and subcategory
2. Type of construction
3. Year of construction
4. Total floor area
5. City
6. Climate region and other weather-related variables

Similarity criteria can be used individually or combined into groups. As more and more buildings are included in the repository, it will be possible for users to use more stringent criteria without having to sacrifice the number of buildings included in the comparison.

5.2.3 Savings potential and area of focus through building energy benchmarking

The first recommendation module is based on the energy consumption characterization of the analyzed buildings and is aimed at providing areas of possible improvement rather than specific measures. This module is only implemented if a big enough selection of similar buildings is identified in the platform, according to the criteria discussed in 5.2.2. Initially, the building is compared to its peers, and a range of potential savings is estimated by comparing its EUI with the ones of the average and best performing buildings in the selection. This enables a rough estimation of whether the building is a good target for potential retrofit strategies. This calculation will be more accurate if more stringent criteria are selected for the similarity check: i.e. the potential savings that can be obtained will be more credible if we select all offices having a certain construction typology, same approximate schedule, and same geographic location, rather than just all the offices in the repository. It is evident that by selecting more stringent comparison criteria, fewer buildings will be available for the comparison, which could result in biased conclusions. A good trade-off must be found for the criteria, which will depend on the total amount of buildings present in the platform: as more and more buildings are included, it will be possible to select more stringent criteria while not sacrificing the size of the building sample.

After the first initial stage, where the general savings potential is assessed, an analysis of the categories of measures that can yield the highest savings is performed. This is obtained by taking advantage of the results of load segmentation and characterization algorithms introduced in Section 5.2.1. The load segmentation model allows to differentiate the base-load of the building from the consumption that depends on heating and on cooling activities. On the other hand, thanks to the characterization techniques, it's possible to determine the change-point temperatures, the heating and cooling coefficients, as well as the load profile patterns of the buildings. The results obtained for the selected building are compared with the ones obtained for its peers. By comparing the load segmentation, the change-point temperatures, and the heating and cooling coefficients, it's possible to determine if the highest savings are achievable by targeting the base, heating, or cooling load of the building. A joint analysis of the load segmentation and EUI can help to determine the potential savings that could be achieved by implementing different categories of measures. At the same time, a study of the load profile patterns of the selected and peer buildings can unveil possible scheduling and management measures that can be implemented.

5.2.4 Retrofit effect prediction

The retrofit effect prediction module enables predictions of the potential savings that might be obtained by implementing certain energy conservation strategies in the buildings or facilities in analysis. The goal of this module is to identify the relationship between the energy efficiency savings obtained with certain measures and the characteristics of the buildings where the measures were implemented. While the algorithm can work even when little data is available, the accuracy of this module is connected to the level of detail of the set of characteristics available for the buildings. The possible features that can be used for this recommendation algorithm are displayed in Tables 5.1 and 5.2. It is evident the list mentioned goes in a deep level of detail and that not all of this information might be known for the buildings in the platform. Nevertheless, as mentioned before, the proposed methodology is scalable, in that if the data is available, it will be possible to harness it to obtain more accurate results, while if some of the features are missing, it will still be possible to obtain results, just to a lesser degree of accuracy. The algorithms that have been shown to work best for this kind of tasks are tree-based model [181], therefore a recommendation for a first implementation of this module can be to use causal forests, random forests, or a gradient boosting methodology.

5.2.5 Selection of relevant measures with multi-label classification

This module of recommendations has the goal of providing a list of possible relevant measures for the selected building without including a range of potential savings associated with these measures. This calculation is carried out only in the case of not having a big enough sample of similar buildings to perform the benchmarking-based recommendation or if there is not enough data about achieved energy savings to perform a prediction of potential retrofit impact. Recommendations from this module are provided with a system similar to collaborative filtering. A collaborative filtering system suggests to users items preferred by ‘similar’ people. In the present application, the system suggests retrofit actions implemented by similar buildings. Hence, the task of this module is two-fold: selecting buildings similar to the one in analysis and selecting the most relevant measures implemented among those buildings.

The approach chosen for this recommendation module is to treat the problem as a multi-label classification problem. Each building in the platform is labeled with m classes, equal to the total number of available measures present in the platform that could potentially be applied in the analyzed building. A binary output is assigned to each class for each sample: if the measure was implemented in the building, this is represented by a value of one. Otherwise, a zero is assigned. The model then learns from all the other buildings in the platform and tries to predict, based on their characteristics and the characteristics of the building in analysis, the most relevant measures to recommend. The result obtained for the target building is, for each possible measure that could potentially be implemented, a rank between zero and one, with the measures that are a better fit reaching a rank closer to one. This could be comparable to running m separate binary classification tasks, one for each measure considered, but multi-label classifiers have the advantage of allowing the treatment of multiple classes (measures) simultaneously, accounting for correlated behavior among them as well. There are several algorithms that can be used for multi-label classification. Some of the most renowned are: decision trees, k-nearest neighbors, neural networks, and random forests [183].

5.3 Recommendation concept designs

In this section, some illustrations are shown of how the results from the concept methodology presented in this chapter could be displayed in a platform containing enough data to perform all the calculations described in the previous paragraphs. In Figure 5.2, the results of the building energy benchmarking module are shown: the EUI of the building is compared to the one of buildings of similar use and geography, and then to all the buildings of similar use in the entire database. A rough estimation of the savings potential for this building is performed by comparing its EUI to the average EUI of the group of buildings selected for comparison.

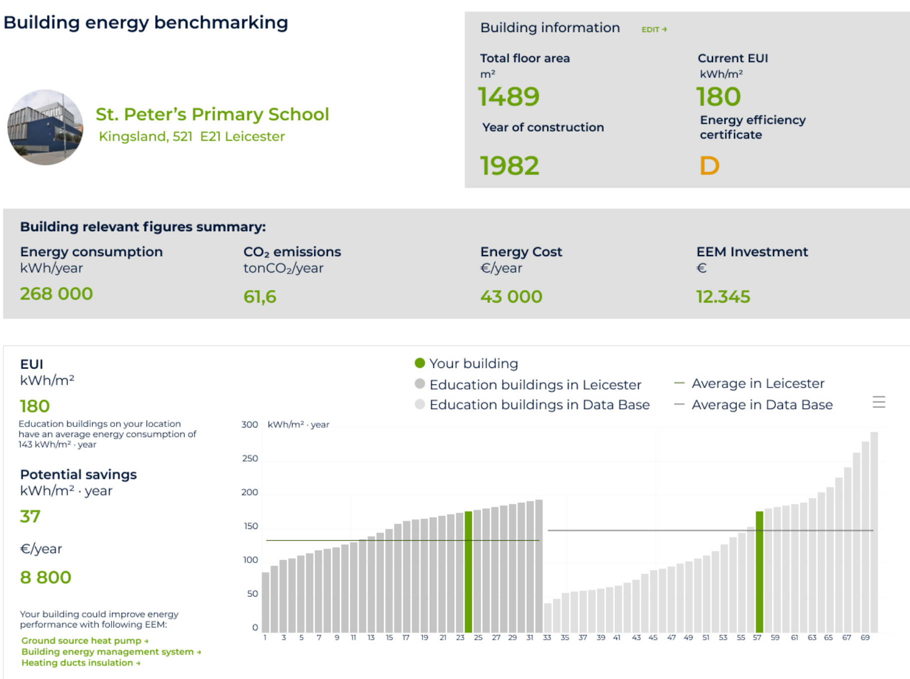


Figure 5.2: Energy benchmarking for a hypothetical educational building located in the UK

In Figure 5.3, a heating and cooling benchmarking is performed, supposing that by using load segmentation and characterization techniques it was possible to estimate the heating and cooling load of all the analysed buildings, as well as their change-point temperatures. From the analysis, it appears that the heating system is a potential candidate to be targeted by energy conservation measures, since the heating EUI of the analysed building is above the average EUI for the comparison group.

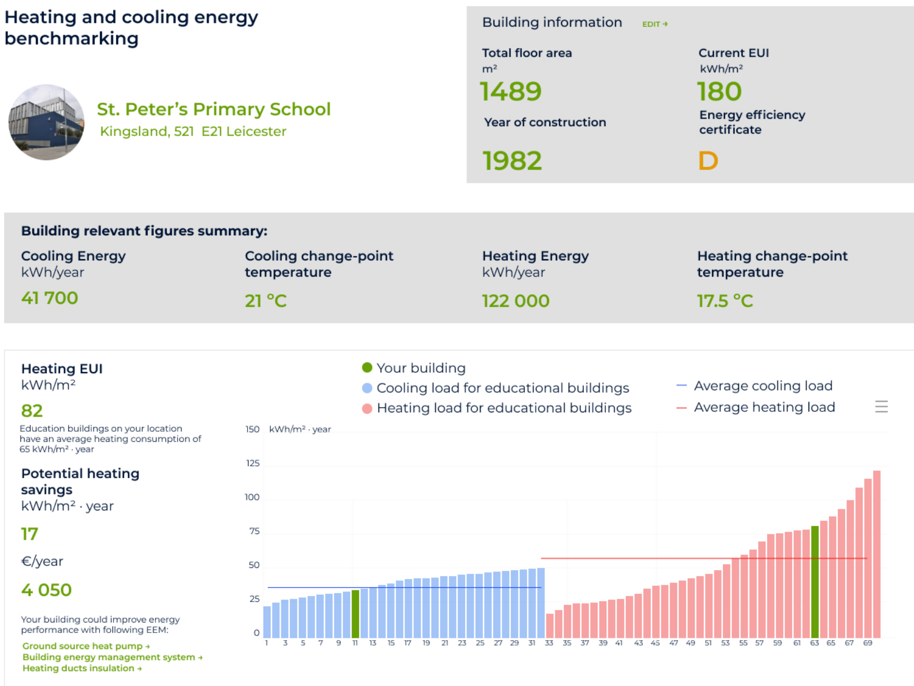


Figure 5.3: Heating and cooling benchmarking for a hypothetical educational building located in the UK

Finally, in Figure 5.4, the retrofit effect prediction calculations are shown, where a scenario is given for the installation of a ground source heat pump in the building. Information about expected energy, CO_2, and economic savings is presented, as well as some basic data about the selected building. Together with the point estimates, uncertainty ranges for the expected savings are provided. These uncertainty bands would depend on the amount of data used for comparison and the affinity between the target building and the buildings where this measure was also implemented.

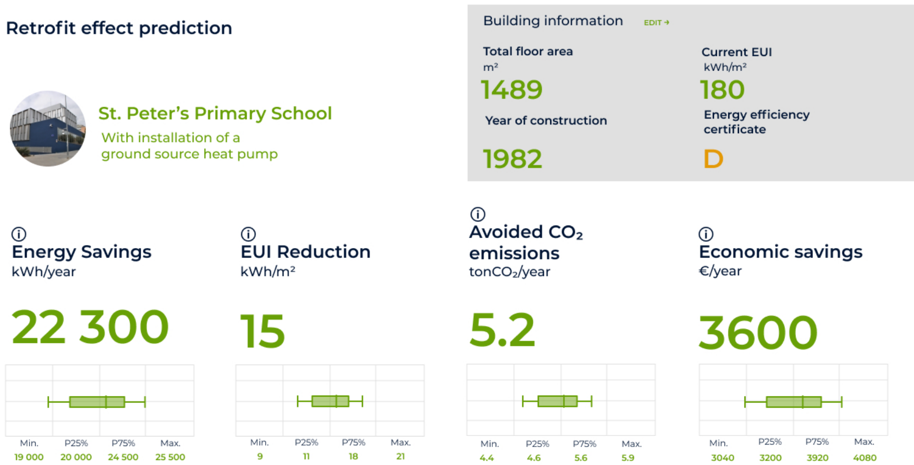


Figure 5.4: Retrofit effect prediction scenario for the installation of a ground source heat pump in a hypothetical educational building located in the UK

5.4 Conclusions

The problem of recommending energy efficiency measures tailored to specific buildings and predicting the effect of energy renovation strategies has been the object of a wide range of studies. Most of these studies employ deterministic approaches based on building energy simulation models. In recent years, the emergence of several platforms containing smart meter energy data and building characteristics has increased the interest of the research community towards the possibility of using data-driven approaches to solve this kind of problems.

The methodology presented in this chapter analyzes big data contained in building energy management platforms to determine savings potential for specific buildings and recommend the implementation of suitable energy renovation strategies. The process is based on the analysis of groups of similar buildings: through the implementation of different algorithms, the savings potential of the analyzed buildings is evaluated, and the relationship between the effect of energy retrofit strategies and the characteristics of the buildings in which they were implemented is estimated. The proposed methodology is highly dependent on the amount and quality of the data available in the platform but has been designed in a way that ensures scalability and robustness. When the available data is limited, preliminary results are provided indicating the most relevant retrofit strategies for the building in analysis, while increasingly accurate recommendations become available as more and more data from other buildings is included in the analysis. The robustness of the presented recommendations is also ensured, by excluding from the analysis the list of measures that are not compatible with the characteristics of the target building.

The presented methodology goes in the direction of enabling the analysis of large quantities of data from building energy management platforms and drawing conclusions based on insights presented by the data. The proposed recommendation methodology also enables the process of energy renovation prioritization within a certain portfolio of buildings. Building managers and institutions frequently face the challenge of having a financial budget for energy renovation programs but without a clear way of determining which measures, and in which buildings, would yield the highest savings and return on investment. The techniques discussed in this chapter can help decision-makers with targeting the buildings which could benefit the most from an energy renovation program and help to perform a preliminary screening and financial analysis of the project.

While the research interest towards the topics discussed in this chapter is high, the proposed methodology is still in a very initial stage. The presented analysis is at a conceptual level, and in order to validate it a considerable work of data collection is required. Nevertheless, knowing the potential results of the techniques presented and the data required for their implementation is an important first step, meaning that the methodology can serve as a guideline for the data collection process when setting up a building energy management platform.

Chapter 6

Conclusions

6.1 Summary and concluding remarks

This thesis is a study of how data-driven techniques can be used to reduce the carbon footprint of the existing built environment and help the global transition towards a sustainable future. Latest research trends have shown that digital tools are becoming a fundamental asset to harness the energy savings potential of buildings and facilities. This research aims to analyze state-of-the-art techniques in this field and propose novel methodologies that can be used to evaluate and recommend energy conservation strategies.

The lack of information about the impact of previously implemented and planned retrofitting actions has been found as one of the main barriers to the widespread implementation of energy renovation programs in buildings. Data-driven approaches present a quick and cost-effective way to perform this kind of analysis without the need for costly ad-hoc engineering studies or intrusive monitoring. As more and more data is being collected in the buildings ecosystem, statistical and machine learning techniques have the potential of harnessing this data to provide actionable insights that can aid stakeholders and policymakers.

When it comes to performing measurement and verification calculations of achieved savings in energy renovation projects, baseline energy models cover a pivotal role. Thanks to these models, the counterfactual energy consumption can be calculated, which represents the estimated energy consumption of the building if the renovation actions did not take place. The accuracy of baseline model calculations is representative of the accuracy of estimated energy savings, meaning that a more detailed and precise model can translate into increased confidence in the results obtained and make energy efficiency more attractive to investors.

In this thesis, two novel data-driven methodologies are introduced, to estimate energy consumption baselines in non-residential buildings. Both these approaches are based on the results of a range of other statistical learning techniques that allow to use different data sources to estimate typical consumption patterns, evaluate weather dependence and characterize the consumption of the analyzed buildings. The first of these two techniques is aimed at producing an accurate and robust estimation of daily energy efficiency savings harnessing the information provided by the recurrent daily consumption profiles identified in the buildings. The second approach utilizes a Bayesian framework to generate hourly baseline predictions with a special focus on accurately quantifying the uncertainty of the obtained results.

In the final part of this work, the possibility of combining the different data-driven techniques introduced in the thesis to provide recommendations for the implementation of energy conservation strategies in buildings is also studied. A theoretical description is provided of how a recommendation and prioritization system could be implemented in a big data building repository to support stakeholders in the planning of energy retrofitting projects.

6.2 Contributions

The main contribution of this thesis is a study of the value of data-driven methodologies to reduce the energy consumption of the built environment, and more specifically to evaluate and recommend energy efficiency strategies in buildings and facilities. Each chapter of this work contributed to the development of this goal.

Chapter 2 presents a detailed review of state-of-the-art techniques currently used to verify energy efficiency savings and predict retrofitting scenarios in buildings. Special attention is given to data-driven techniques, as they prove of high interest because of their cost-efficiency and easy scalability. The data-driven baseline estimation methods analyzed are classified in statistical learning techniques, machine learning techniques, and Bayesian methods. The strengths and weaknesses of each of these approaches are also studied and schematized.

The third chapter presents a novel approach for the measurement and verification of energy efficiency savings at daily scale in non-residential buildings. This methodology is tested in different case studies with simulated and real data and is compared to one of the main state-of-the-art methodologies used for measurement and verification, showing comparable or better accuracy and higher robustness to missing data. The approach is also characterized by low minimum data requirements and the ability to provide additional actionable insights to users together with the savings estimations.

In Chapter 4 another data-driven methodology is introduced. This novel approach can be used to estimate hourly energy baseline predictions and characterize the consumption of the analyzed buildings. The methodology is based on a Bayesian linear model that uses clearly interpretable terms, making it easily explainable to technical as well as non-technical stakeholders. The Bayesian framework also offers a coherent and robust estimation of uncertainty, meaning that this approach could potentially be employed to perform de-risking assessments in the energy efficiency field. The techniques presented in this chapter were tested on one of the largest open repositories of non-residential building energy data.

Chapter 5 presents a concept methodology to recommend and prioritize energy efficiency strategies in buildings. This kind of analysis is usually performed with deterministic methods, tailored to the specific buildings in study. The approach proposed in Chapter 5, on the other hand, is data-driven and its core principle is to harness the information contained in big data building repositories. The process is based on the analysis of groups of similar buildings and the possibility to map achieved savings from certain retrofit programs to the characteristics of the buildings where the programs were implemented. The potential data requirements for implementing this recommendation and prioritization module in a building energy management platform are also outlined, meaning that they can be used as a guideline for the data collection process when building a platform of this sort.

6.3 Future research

The research carried out in this thesis opens different lines of research work for the future.

1. The measurement and verification methodologies presented in Chapter 3 and 4 could be enhanced by including a module that assesses the extra benefits

associated with the achieved savings, depending on the grid conditions at the day and hour in which the savings occurred.

2. An improvement of the Bayesian methodology introduced in Chapter 4 could be attempted by using the obtained change-point temperatures to perform load segmentation calculations in the analyzed buildings. Techniques to aggregate the uncertainty estimated at hourly level to the one associated with whole energy efficiency programs could also be investigated.

3. The methodology presented in Chapter 5 for recommendation and prioritization of energy retrofitting projects should be validated with collected data from buildings. This would mainly consist of a big dataset that includes building energy consumption data, information about implemented retrofit measures with relative savings, and characteristics of the buildings.

4. Finally, all of the techniques introduced are based on data and, as such, could benefit from further validation and testing with different data from the one that was analyzed in this thesis. This could be obtained by integrating the presented techniques as analytics modules within a big data building energy management platform.

Publications

Three different articles were published in peer reviewed scientific journals stemming from the research work of this doctoral thesis. The mentioned journals are Renewable and Sustainable Energy Reviews, Applied Energy and Energies. The table below shows the JCR metric characteristics for these journals, for the year 2020.

JCR journal metrics for Renewable and Sustainable Energy Reviews, Applied Energy and Energies

Journal Name	JIF 2020	Q 2020	Rank Category
Renewable and Sustainable Energy Reviews	14.982	Q1	Green & Sustainable Science & Technology
Applied Energy	9.746	Q1	Energy & Fuels
Energies	3.004	Q3	Energy & Fuels

1. Grillone B., Danov S., Sumper A., Cipriano J., Mor G. A review of deterministic and data-driven methods to quantify energy efficiency savings and to predict retrofitting scenarios in buildings. Renewable and Sustainable Energy Reviews, 2020; 131: 110027, https://doi.org/10.1016/j.rser.2020.110027

2. Grillone B., Mor G., Danov S., Cipriano J., Sumper A. A data-driven methodology for enhanced measurement and verification of energy efficiency savings in commercial buildings. Applied Energy 2021; 301: 117502, https://doi.org/10.1016/j.apenergy.2021.117502

3. Grillone B., Mor G., Danov S., Cipriano J., Lazzari F., Sumper A. Baseline energy use modeling and characterization in tertiary buildings using an interpretable Bayesian linear regression methodology. Energies 2021; 14, 5556, https://doi.org/10.3390/en14175556

One conference article related to the work developed in this thesis was also published:

Grillone B., Mor G., Danov S., Cipriano J., Carbonell J., Gabaldon E. Use of generalized additive models to assess energy efficiency savings in buildings using smart metering data. Proceedings of Sustainable Built Environment Conference, Scilla, Italy, 16-17 May 2019

Bibliography

[1] IPCC. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Tech. rep. IPCC, 2021.

[2] IPCC. Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Tech. rep. IPCC, 2021.

[3] United Nations Framework Climate Convention on Climate Change. Paris Agreement. 2016.

[4] Climate Action Tracker. Warming Projections Global Update. Tech. rep. Climate Action Tracker, 2021.

[5] United Nations, Department of Economic and Social Affairs, and Population Division. World urbanization prospects: the 2018 revision. OCLC: 1120698127. 2019.

[6] Ilaria Delponte and Corrado Schenone. RES Implementation in Urban Areas: An Updated Overview. In: *Sustainability* 12.1 (2020),

[7] Global Alliance for Buildings and Construction. 2020 Global Status Report for Buildings and Construction. Tech. rep. Global Alliance for Buildings and Construction, 2020.

[8] International Energy Agency, ed. Transition to sustainable buildings: strategies and opportunities to 2050. Paris: IEA Publ, 2013.

[9] International Energy Agency. Market Report Series: Energy Efficiency 2018. Tech. rep. International Energy Agency (IEA), 2018.

[10] COMMISSION RECOMMENDATION (EU) 2019/ 786 - of 8 May 2019 - on building renovation - (notified under document C(2019) 3352). 2019.

[11] Electric Power Research Institute. State Level Electric Energy Efficiency Potential Estimates. Tech. rep. Electric Power Research Institute, 2017.

[12] Julia Szinai. Putting your money where your meter is. A study of pay-for-performance energy efficiency programs in the united states. Tech. rep. Natural Resources Defense Council, 2017.

[13] Brian F. Gerke, Giulia Gallo, Sarah J. Smith, Jingjing Liu, Peter Alstone, Shuba Raghavan, Peter Schwartz, Mary Ann Piette, Rongxin Yin, and Sofia Stensson. The California Demand Response Potential Study, Phase 3: Final Report on the Shift Resource through 2030. In: (2020). Publisher: LBNL.

[14] Final report of the California public utilities commission's working group on load shift. Tech. rep. California Public Utilities Commission, 2019.

[15] International Energy Agency. Better energy efficiency policy with digital tools. Tech. rep. International Energy Agency (IEA), 2021.

[16] European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions-A policy framework for climate and energy in the period from 2020 to 2030, COM (2014) 15 final, Brussels. Tech. rep. 2014.

[17] Pieter de Wilde. The gap between predicted and measured energy performance of buildings: A framework for investigation. In: *Automation in Construction* 41 (2014),

[18] Ray Galvin. 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'. In: *Energy and Buildings* 69 (2014),

[19] Pekka Tuominen, Krzysztof Klobut, Anne Tolman, Afi Adjei, and Marjolein de Best-Waldhober. Energy savings potential in buildings and overcoming market barriers in member states of the European Union. In: *Energy and Buildings* 51 (2012),

[20] Constantine E. Kontokosta. Modeling the energy retrofit decision in commercial office buildings. In: *Energy and Buildings* 131 (2016),

[21] Zhenjun Ma, Paul Cooper, Daniel Daly, and Laia Ledo. Existing building retrofits: Methodology and state-of-the-art. In: *Energy and Buildings* 55 (2012),

[22] Paul A. Mathew, Laurel N. Dunn, Michael D. Sohn, Andrea Mercado, Claudine Custudio, and Travis Walter. Big-data for building energy performance: Lessons from assembling a very large national database of building energy use. In: *Applied Energy* 140 (2015),

[23] Philipp Geyer, Arno Schlüter, and Sasha Cisar. Application of clustering for the development of retrofit strategies for large building stocks. In: *Advanced Engineering Informatics* 31 (2017),

[24] Wenzhuo Li, Choongwan Koo, Taehoon Hong, Jeongyoon Oh, Seung Hyun Cha, and Shengwei Wang. A novel operation approach for the energy efficiency improvement of the HVAC system in office spaces through real-time big data analytics. In: *Renewable and Sustainable Energy Reviews* 127 (2020),

[25] E. Iddio, L. Wang, Y. Thomas, G. McMorrow, and A. Denzer. Energy efficient operation and modeling for greenhouses: A literature review. In: *Renewable and Sustainable Energy Reviews* 117 (2020),

[26] Fabiano Pallonetto, Mattia De Rosa, Francesco D'Ettorre, and Donal P. Finn. On the assessment and control optimisation of demand response programs in residential buildings. In: *Renewable and Sustainable Energy Reviews* 127 (2020),

[27] EVO. International Performance Measurement and Verification Protocol. Tech. rep. 2017.

[28] Ellen Franconi, Matt Gee, Miriam Goldberg, Jessica Granderson, Tim Guiterman, Michael Li, and Brian Arthur Smith. The Status and Promise of Advanced M&V: An Overview of "M&V 2.0" Methods, Tools, and Applications. Tech. rep. LBNL–1007125, 1350974. 2017.

[29] Colm V. Gallagher, Kevin Leahy, Peter O'Donovan, Ken Bruton, and Dominic T.J. O'Sullivan. Development and application of a machine learning supported methodology for measurement and verification (M&V) 2.0. In: *Energy and Buildings* 167 (2018),

[30] Jessica Granderson and Samuel Fernandes. The state of advanced measurement and verification technology and industry application. In: *The Electricity Journal* 30.8 (2017),

[31] Aurélie Foucquier, Sylvain Robert, Frédéric Suard, Louis Stéphan, and Arnaud Jay. State of the art in building modelling and energy performances prediction: A review. In: *Renewable and Sustainable Energy Reviews* 23 (2013),

[32] Hai-xiang Zhao and Frédéric Magoulès. A review on the prediction of building energy consumption. In: *Renewable and Sustainable Energy Reviews* 16.6 (2012),

[33] A.S. Ahmad, M.Y. Hassan, M.P. Abdullah, H.A. Rahman, F. Hussin, H. Abdullah, and R. Saidur. A review on applications of ANN and SVM for building electrical energy consumption forecasting. In: *Renewable and Sustainable Energy Reviews* 33 (2014),

[34] Muhammad Qamar Raza and Abbas Khosravi. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. In: *Renewable and Sustainable Energy Reviews* 50 (2015),

[35] Zeyu Wang and Ravi S. Srinivasan. A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. In: *Renewable and Sustainable Energy Reviews* 75 (2017),

[36] B. Yildiz, J.I. Bilbao, and A.B. Sproul. A review and analysis of regression and machine learning models on commercial building electricity load forecasting. In: *Renewable and Sustainable Energy Reviews* 73 (2017),

[37] Kadir Amasyali and Nora M. El-Gohary. A review of data-driven building energy consumption prediction studies. In: *Renewable and Sustainable Energy Reviews* 81 (2018),

[38] Tanveer Ahmad, Huanxin Chen, Yabin Guo, and Jiangyu Wang. A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. In: *Energy and Buildings* 165 (2018),

[39] Chirag Deb, Fan Zhang, Junjing Yang, Siew Eang Lee, and Kwok Wei Shah. A review on time series forecasting techniques for building energy consumption. In: *Renewable and Sustainable Energy Reviews* 74 (2017),

[40] Yixuan Wei, Xingxing Zhang, Yong Shi, Liang Xia, Song Pan, Jinshun Wu, Mengjie Han, and Xiaoyun Zhao. A review of data-driven approaches for prediction and classification of building energy consumption. In: *Renewable and Sustainable Energy Reviews* 82 (2018),

[41] V.S.K.V. Harish and Arun Kumar. A review on modeling and simulation of building energy systems. In: *Renewable and Sustainable Energy Reviews* 56 (2016),

[42] Sang Hoon Lee, Tianzhen Hong, Mary Ann Piette, and Sarah C. Taylor-Lange. Energy retrofit analysis toolkits for commercial buildings: A review. In: *Energy* 89 (2015),

[43] Jessica Granderson, Samir Touzani, Claudine Custodio, Michael D. Sohn, David Jump, and Samuel Fernandes. Accuracy of automated measurement and verification (M&V) techniques for energy savings in commercial buildings. In: *Applied Energy* 173 (2016),

[44] Jessica Granderson, Samir Touzani, Samuel Fernandes, and Cody Taylor. Application of automated measurement and verification to utility energy efficiency program data. In: *Energy and Buildings* 142 (2017),

[45] ASHRAE. ASHRAE Guideline 14–2014, Measurement of Energy, Demand, and Water Savings. ASHRAE Atlanta, 2014.

[46] Charles W. Kurnik, Michael Li, Hossein Haeri, and Arlis Reynolds. The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures: January 2012 - September 2016. Tech. rep. NREL/SR–7A40-70472, 1463449. 2018.

[47] CalTRACK Initiative. CalTRACK.

[48] Margaret F. Fels. PRISM: An introduction. In: *Energy and Buildings* 9.1-2 (1986),

[49] Johanna L. Mathieu, Phillip N. Price, Sila Kiliccote, and Mary Ann Piette. Quantifying Changes in Building Electricity Use, With Application to Demand Response. In: *IEEE Transactions on Smart Grid* 2.3 (2011),

[50] Clayton Miller and Forrest Meggers. The Building Data Genome Project: An open, public data set from non-residential building electrical meters. In: *Energy Procedia* 122 (2017),

[51] Samir Touzani, Jessica Granderson, and Samuel Fernandes. Gradient boosting machine for modeling the energy consumption of commercial buildings. In: *Energy and Buildings* 158 (2018),

[52] Samir Touzani, Jessica Granderson, David Jump, and Derrick Rebello. Evaluation of methods to assess the uncertainty in estimated energy savings. In: *Energy and Buildings* 193 (2019),

[53] Suhaidi Mohd Aris, Nofri Yenita Dahlan, Mohd Nasrun Mohd Nawi, Tengku Ahmad Nizam, and Mohamad Zamhari Tahir. Quantifying energy savings for retrofit centralized HVAC systems at Selangor state secretary complex. In: *Jurnal Teknologi* 77.5 (2015).

[54] Haidong Wang, Yuantao Xue, and Yu Mu. Assessment of energy savings by mechanical system retrofit of existing buildings. In: *Procedia Engineering* 205 (2017),

[55] J. Kelly Kissock and Carl Eger. Measuring industrial energy savings. In: *Applied Energy* 85.5 (2008),

[56] Matthew Brown, Chris Barrington-Leigh, and Zosia Brown †. Kernel regression for real-time building energy analysis. In: *Journal of Building Performance Simulation* 5.4 (2012),

[57] M.J. Jiménez and M.R. Heras. Application of multi-output ARX models for estimation of the U and g values of building components in outdoor testing. In: *Solar Energy* 79.3 (2005),

[58] M.J. Jiménez, H. Madsen, and K.K. Andersen. Identification of the main thermal characteristics of building components using MATLAB. In: *Building and Environment* 43.2 (2008),

[59] M.J. Jiménez, B. Porcar, and M.R. Heras. Estimation of building component UA and gA from outdoor tests in warm and moderate weather conditions. In: *Solar Energy* 82.7 (2008),

[60] M.J. Jiménez, B. Porcar, and M.R. Heras. Application of different dynamic analysis approaches to the estimation of the building component U value. In: *Building and Environment* 44.2 (2009),

[61] Julián Arco Díaz, José Sánchez Ramos, M. Carmen Guerrero Delgado, David Hidalgo García, Francisco Gil Montoya, and Servando Álvarez Domínguez. 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. In: *Applied Energy* 224 (2018),

[62] Melek Yalcintas. Energy-savings predictions for building-equipment retrofits. In: *Energy and Buildings* 40.12 (2008),

[63] Wan n Nazirah Wan Md Adna, Nofri Yenita Dahlan, and Ismail Musirin. Development of Hybrid Artificial Neural Network for Quantifying Energy Saving using Measurement and Verification. In: *Indonesian Journal of Electrical Engineering and Computer Science* 8.1 (2017),

[64] Chenchen Chang, Jing Zhao, and Neng Zhu. Energy saving effect prediction and post evaluation of air-conditioning system in public buildings. In: *Energy and Buildings* 43.11 (2011),

[65] Bing Dong, Cheng Cao, and Siew Eang Lee. Applying support vector machines to predict building energy consumption in tropical region. In: *Energy and Buildings* 37.5 (2005),

[66] Abdiansah Abdiansah and Retantyo Wardoyo. Time Complexity Analysis of Support Vector Machines (SVM) in LibSVM. In: *International Journal of Computer Applications* 128.3 (2015),

[67] Muhammad Waseem Ahmad, Monjur Mourshed, and Yacine Rezgui. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. In: *Energy and Buildings* 147 (2017),

[68] Daniel B. Araya, Katarina Grolinger, Hany F. ElYamany, Miriam A.M. Capretz, and Girma Bitsuamlak. An ensemble learning framework for anomaly detection in building energy consumption. In: *Energy and Buildings* 144 (2017),

[69] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. Bayesian Data Analysis, Third Edition. 3rd. Chapman & Hall, 2014.

[70] Herman Carstens, Xiaohua Xia, and Sarma Yadavalli. Bayesian Energy Measurement and Verification Analysis. In: *Energies* 11.2 (2018),

[71] John A Shonder and Piljae Im. Bayesian Analysis of Savings from Retrofit Projects. In: *ASHRAE Transactions* (2012),

[72] David Lindelöf, Mohammad Alisafaee, Pierluca Borsò, Christian Grigis, and Jean Viaene. Bayesian verification of an energy conservation measure. In: *Energy and Buildings* 171 (2018),

[73] Yeonsook Heo and Victor M. Zavala. Gaussian process modeling for measurement and verification of building energy savings. In: *Energy and Buildings* 53 (2012),

[74] Michael C. Burkhart, Yeonsook Heo, and Victor M. Zavala. Measurement and verification of building systems under uncertain data: A Gaussian process modeling approach. In: *Energy and Buildings* 75 (2014),

[75] Jacques Maritz, Foster Lubbe, and Louis Lagrange. A Practical Guide to Gaussian Process Regression for Energy Measurement and Verification within the Bayesian Framework. In: *Energies* 11.4 (2018),

[76] Abhishek Srivastav, Ashutosh Tewari, and Bing Dong. Baseline building energy modeling and localized uncertainty quantification using Gaussian mixture models. In: *Energy and Buildings* 65 (2013),

[77] Samir Touzani, Baptiste Ravache, Eliot Crowe, and Jessica Granderson. Statistical change detection of building energy consumption: Applications to savings estimation. In: *Energy and Buildings* 185 (2019),

[78] Wei Tian, Yeonsook Heo, Pieter de Wilde, Zhanyong Li, Da Yan, Cheol Soo Park, Xiaohang Feng, and Godfried Augenbroe. A review of uncertainty analysis in building energy assessment. In: *Renewable and Sustainable Energy Reviews* 93 (2018),

[79] T. Agami Reddy and David Claridge. Uncertainty of "Measured" Energy Savings from Statistical Baseline Models. In: *HVAC&R Research* 6.1 (2000),

[80] Bill Koran, Erik Boyer, M. Sami Khawaja, Josh Rushton, and Jim Stewart. A Comparison of Approaches to Estimating the Time-Aggregated Uncertainty of Savings Estimated from Meter Data. In: (2017),

[81] Zadok Olinga, Xiaohua Xia, and Xianming Ye. A cost-effective approach to handle measurement and verification uncertainties of energy savings. In: *Energy* 141 (2017),

[82] Paolo Zangheri, Roberto Armani, Marco Pietrobon, and Lorenzo Pagliano. Identification of cost-optimal and NZEB refurbishment levels for representative climates and building typologies across Europe. In: *Energy Efficiency* 11.2 (2018),

[83] EnergyPlus Development Team. EnergyPlus engineering reference: The reference to EnergyPlus calculations. In: *EnergyPlus Version 9.1.0 Documentation* (2019),

[84] A.M. Rysanek and R. Choudhary. Optimum building energy retrofits under technical and economic uncertainty. In: *Energy and Buildings* 57 (2013),

[85] University of Wisconsin–Madison. Solar Energy Laboratory. TRNSYS, a transient simulation program. Madison, Wis. : The Laboratory, 1975., 1975.

[86] Paolo Valdiserri, Cesare Biserni, Giacomo Tosi, and Massimo Garai. Retrofit Strategies Applied to a Tertiary Building Assisted by Trnsys Energy Simulation Tool. In: *Energy Procedia* 78 (2015),

[87] J.A. Suástegui Macías, C. Pérez Tello, A. Acuña Ramírez, A.A. Lambert Arista, H.D. Magaña Almaguer, P.F. Rosales Escobedo, and A.H. Ruelas Puente. Assessment of electrical saving from energy efficiency programs in the residential sector in Mexicali, Mexico. In: *Sustainable Cities and Society* 38 (2018),

[88] Fabrizio Ascione, Nicola Bianco, Claudio De Stasio, Gerardo Maria Mauro, and Giuseppe Peter Vanoli. Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach. In: *Energy* 118 (2017),

[89] Ehsan Asadi, M. Gameiro da Silva, C. Henggeler Antunes, and Luís Dias. State of the Art on Retrofit Strategies Selection Using Multi-objective Optimization and Genetic Algorithms. In: *Nearly Zero Energy Building Refurbishment*. London: Springer London, 2013, pp. 279–297.

[90] V. Siddharth, P.V. Ramakrishna, T. Geetha, and Anand Sivasubramaniam. Automatic generation of energy conservation measures in buildings using genetic algorithms. In: *Energy and Buildings* 43.10 (2011),

[91] Youssef Bichiou and Moncef Krarti. Optimization of envelope and HVAC systems selection for residential buildings. In: *Energy and Buildings* 43.12 (2011),

[92] Fabrizio Ascione, Nicola Bianco, Rosa Francesca De Masi, Claudio De Stasio, Gerardo Maria Mauro, and Giuseppe Peter Vanoli. Multi-objective optimization of the renewable energy mix for a building. In: *Applied Thermal Engineering* 101 (2016),

[93] Fabrizio Ascione, Nicola Bianco, Claudio De Stasio, Gerardo Maria Mauro, and Giuseppe Peter Vanoli. Multi-stage and multi-objective optimization for energy retrofitting a developed hospital reference building: A new approach to assess cost-optimality. In: *Applied Energy* 174 (2016),

[94] Fabrizio Ascione, Nicola Bianco, Gerardo Mauro, Davide Napolitano, and Giuseppe Vanoli. A Multi-Criteria Approach to Achieve Constrained Cost-Optimal Energy Retrofits of Buildings by Mitigating Climate Change and Urban Overheating. In: *Climate* 6.2 (2018),

[95] Fanny Pernodet Chantrelle, Hicham Lahmidi, Werner Keilholz, Mohamed El Mankibi, and Pierre Michel. Development of a multicriteria tool for optimizing the renovation of buildings. In: *Applied Energy* 88.4 (2011),

[96] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. In: *IEEE Transactions on Evolutionary Computation* 6.2 (2002),

[97] Navid Delgarm, Behrang Sajadi, Saeed Delgarm, and Farshad Kowsary. A novel approach for the simulation-based optimization of the buildings energy consumption using NSGA-II: Case study in Iran. In: *Energy and Buildings* 127 (2016),

[98] Laurent Magnier and Fariborz Haghighat. Multiobjective optimization of building design using TRNSYS simulations, genetic algorithm, and Artificial Neural Network. In: *Building and Environment* 45.3 (2010),

[99] Ehsan Asadi, Manuel Gameiro da Silva, Carlos Henggeler Antunes, Luís Dias, and Leon Glicksman. Multi-objective optimization for building retrofit: A model using genetic algorithm and artificial neural network and an application. In: *Energy and Buildings* 81 (2014),

[100] E. Iturriaga, U. Aldasoro, J. Terés-Zubiaga, and A. Campos-Celador. Optimal renovation of buildings towards the nearly Zero Energy Building standard. In: *Energy* 160 (2018),

[101] Daniel E. Marasco and Constantine E. Kontokosta. Applications of machine learning methods to identifying and predicting building retrofit opportunities. In: *Energy and Buildings* 128 (2016),

[102] Marco Beccali, Giuseppina Ciulla, Valerio Lo Brano, Alessandra Galatioto, and Marina Bonomolo. Artificial neural network decision support tool for assessment of the energy performance and the refurbishment actions for the non-residential building stock in Southern Italy. In: *Energy* 137 (2017),

[103] Graziano Salvalai, Laura Elisabetta Malighetti, Leopoldo Luchini, and Sara Girola. Analysis of different energy conservation strategies on existing school buildings in a Pre-Alpine Region. In: *Energy and Buildings* 145 (2017),

[104] Travis Walter and Michael D. Sohn. A regression-based approach to estimating retrofit savings using the Building Performance Database. In: *Applied Energy* 179 (2016),

[105] US DOE and Lawrence Berkeley National Laboratory. Building Performance Database (BPD) | Department of Energy.

[106] Yi-Kai Juan, Peng Gao, and Jie Wang. A hybrid decision support system for sustainable office building renovation and energy performance improvement. In: *Energy and Buildings* 42.3 (2010),

[107] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, eds. An introduction to statistical learning: with applications in R. Springer texts in statistics 103. OCLC: ocn828488009. New York: Springer, 2013.

[108] Clayton Miller. More Buildings Make More Generalizable Models—Benchmarking Prediction Methods on Open Electrical Meter Data. In: *Machine Learning and Knowledge Extraction* 1.3 (2019),

[109] Clayton Miller and Forrest Meggers. Mining electrical meter data to predict principal building use, performance class, and operations strategy for hundreds of non-residential buildings. In: *Energy and Buildings* 156 (2017),

[110] International Energy Agency. 2019 Global Status Report for Buildings and Construction. Tech. rep. 2019, p. 41.

[111] Na Wang, Patrick E. Phelan, Jorge Gonzalez, Chioke Harris, Gregor P. Henze, Robert Hutchinson, Jared Langevin, Mary Ann Lazarus, Brent Nelson, Chris Pyke, Kurt Roth, David Rouse, Karma Sawyer, and Stephen Selkowitz. Ten questions concerning future buildings beyond zero energy and carbon neutrality. In: *Building and Environment* 119 (2017),

[112] European Commission-DG Ener. Clean energy for all Europeans. Publications Office of the European Union, 2019.

[113] Minda Ma, Wei Cai, and Weiguang Cai. Carbon abatement in China's commercial building sector: A bottom-up measurement model based on Kaya-LMDI methods. In: *Energy* 165 (2018),

[114] Minda Ma, Xin Ma, Wei Cai, and Weiguang Cai. Low carbon roadmap of residential building sector in China: Historical mitigation and prospective peak. In: *Applied Energy* 273 (2020),

[115] Benedetto Grillone, Stoyan Danov, Andreas Sumper, Jordi Cipriano, and Gerard Mor. A review of deterministic and data-driven methods to quantify energy efficiency savings and to predict retrofitting scenarios in buildings. In: *Renewable and Sustainable Energy Reviews* 131 (2020),

[116] EnergyPlus Development Team. EnergyPlus engineering reference: The reference to EnergyPlus calculations. 2019.

[117] CalTRACK Initiative. CalTRACK. Publication Title: CalTRACK.

[118] Fintan McLoughlin, Aidan Duffy, and Michael Conlon. A clustering approach to domestic electricity load profile characterisation using smart metering data. In: *Applied Energy* 141 (2015),

[119] Ioannis P. Panapakidis, Theofilos A. Papadopoulos, Georgios C. Christoforidis, and Grigoris K. Papagiannis. Pattern recognition algorithms for electricity load curve analysis of buildings. In: *Energy and Buildings* 73 (2014),

[120] Alfonso Capozzoli, Marco Savino Piscitelli, Silvio Brandi, Daniele Grassi, and Gianfranco Chicco. Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings. In: *Energy* 157 (2018),

[121] Fateh Melzi, Allou Same, Mohamed Zayani, and Latifa Oukhellou. A Dedicated Mixture Model for Clustering Smart Meter Data: Identification and Analysis of Electricity Consumption Behaviors. In: *Energies* 10.10 (2017),

[122] Kehua Li, Zhenjun Ma, Duane Robinson, and Jun Ma. Identification of typical building daily electricity usage profiles using Gaussian mixture model-based clustering and hierarchical clustering. In: *Applied Energy* 231 (2018),

[123] Peder Bacher and Henrik Madsen. Load forecasting for supermarket refrigeration. In: *Applied Energy* 163 (),

[124] Caleb Robinson, Bistra Dilkina, Jeffrey Hubbs, Wenwen Zhang, Subhrajit Guhathakurta, Marilyn A. Brown, and Ram M. Pendyala. Machine learning approaches for estimating commercial building energy consumption. In: *Applied Energy* 208 (2017),

[125] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16* (2016),

[126] Peder Bacher, Henrik Madsen, Henrik Aalborg Nielsen, and Bengt Perers. Short-term heat load forecasting for single family houses. In: *Energy and Buildings* 65 (2013),

[127] Trevor Hastie and Robert Tibshirani. Generalized Additive Models. Chapman & Hall/CRC, 1990.

[128] Trimble Inc. SketchUp. 2019.

[129] The Big Ladder Software. Euclid. 2017.

[130] Powered by Dark Sky. Dark Sky API. 2019.

[131] Tobias Loga and Jens Calisti. TABULA WebTool. IWU – Institut Wohnen und Umwelt GmbH, 2017.

[132] Tobias Loga, Britta Stein, and Nikolaus Diefenbach. TABULA building typologies in 20 European countries—Making energy-related features of residential building stocks comparable. In: *Energy and Buildings* 132 (2016),

[133] Phil Ngo. OpenEEmeter Documentation.

[134] The Copernicus Programme. Copernicus Atmoshpere Monitoring Service (CAMS).

[135] Meteotest. Meteonorm.

[136] OpenWeather. OpenWeather.

[137] 2020 Global Status Report for Buildings and Construction. Tech. rep. Global Alliance for Buildings and Construction, 2020.

[138] Tracking Buildings 2020. Tech. rep. International Energy Agency (IEA), 2020.

[139] Tatiana Loureiro, Marta Gil, Rachel Desmaris, Annalisa Andaloro, Charikleia Karakosta, and Stefan Plesser. De-Risking Energy Efficiency Investments through Innovation. In: *Proceedings* 65.1 (2020),

[140] World Energy Investment 2020. Tech. rep. International Energy Agency (IEA), 2020.

[141] Zakia Afroz, H. Burak Gunay, William O'Brien, Guy Newsham, and Ian Wilton. An inquiry into the capabilities of baseline building energy modelling approaches to estimate energy savings. In: *Energy and Buildings* 244 (2021),

[142] Travis Walter, Phillip N. Price, and Michael D. Sohn. Uncertainty estimation improves energy measurement and verification procedures. In: *Applied Energy* 130 (2014),

[143] Charles Goldman, Michael Reid, Roger Levy, and Alison Silverstein. Coordination of Energy Efficiency and Demand Response. In: (2010),

[144] International Energy Efficiency Financing Protocol (IEEFP). Tech. rep. EVO (Efficiency Valuation Organization), 2021.

[145] International Performance Measurement and Verification Protocol. Tech. rep. EVO (Efficiency Valuation Organization), 2017.

[146] Herman Carstens. A Bayesian approach to energy monitoring optimization. PhD thesis. University of Pretoria, 2017.

[147] Richard McElreath. Statistical Rethinking. 2nd Edition. Chapman and Hall/CRC, 2020.

[148] Jim Albert. Bayesian Computation with R. Springer, 2009.

[149] Joop J. Hox, Mirjam Moerbeek, and Rens van de Schoot. Multilevel Analysis. Routledge, 2017.

[150] Wei Tian, Song Yang, Zhanyong Li, Shen Wei, Wei Pan, and Yunliang Liu. Identifying informative energy data in Bayesian calibration of building energy models. In: *Energy and Buildings* 119 (2016),

[151] Martin Heine Kristensen, Rasmus Elbæk Hedegaard, and Steffen Petersen. Hierarchical calibration of archetypes for urban building energy modeling. In: *Energy and Buildings* 175 (2018),

[152] Siyan Wang, Xun Sun, and Upmanu Lall. A hierarchical Bayesian regression model for predicting summer residential electricity demand across the U.S.A. In: *Energy* 140 (2017),

[153] Xiao-Chen Yuan, Xun Sun, Weigang Zhao, Zhifu Mi, Bing Wang, and Yi-Ming Wei. Forecasting China's regional energy demand by 2030: A Bayesian approach. In: *Resources, Conservation and Recycling* 127 (2017),

[154] Sergio Pezzulli, Patrizio Frederic, Shanti Majithia, Sal Sabbagh, Emily Black, Rowan Sutton, and David Stephenson. The seasonal forecast of electricity demand: a hierarchical Bayesian model with climatological weather generator. In: *Applied Stochastic Models in Business and Industry* 22.2 (2006),

[155] Amin Moradkhani, Mahmoud R. Haghifam, and Mohsen Mohammadzadeh. Bayesian estimation of overhead lines failure rate in electrical distribution systems. In: *International Journal of Electrical Power & Energy Systems* 56 (2014),

[156] Joel Villavicencio Gastelu, Joel David Melo Trujillo, and Antonio Padilha-Feltrin. Hierarchical Bayesian Model for Estimating Spatial-Temporal Photovoltaic Potential in Residential Areas. In: *IEEE Transactions on Sustainable Energy* 9.2 (2018),

[157] Mehdi Jafari, Laura E. Brown, and Lucia Gauchia. Hierarchical Bayesian Model for Probabilistic Analysis of Electric Vehicle Battery Degradation. In: *IEEE Transactions on Transportation Electrification* 5.4 (2019),

[158] A.T. Booth, R. Choudhary, and D.J. Spiegelhalter. A hierarchical Bayesian framework for calibrating micro-level models with macro-level data. In: *Journal of Building Performance Simulation* 6.4 (2013),

[159] Clayton Miller, Anjukan Kathirgamanathan, Bianca Picchetti, Pandarasamy Arjunan, June Young Park, Zoltan Nagy, Paul Raftery, Brodie W. Hobson, Zixiao Shi, and Forrest Meggers. The Building Data Genome Project 2, energy meter data from the ASHRAE Great Energy Predictor III competition. In: *Scientific Data* 7.1 (2020),

[160] Clayton Miller, Pandarasamy Arjunan, Anjukan Kathirgamanathan, Chun Fu, Jonathan Roth, June Young Park, Chris Balbach, Krishnan Gowri, Zoltan Nagy, Anthony D. Fontanini, and Jeff Haberl. The ASHRAE Great Energy Predictor III competition: Overview and results. In: *Science and Technology for the Built Environment* 26.10 (2020),

[161] Mauro Tucci and Marco Raugi. Analysis of spectral clustering algorithms for linear and nonlinear time series. In: *2011 11th International Conference on Intelligent Systems Design and Applications*. Cordoba, Spain: IEEE, 2011, pp. 925–930.

[162] Margaret F. Fels. PRISM: An introduction. In: *Energy and Buildings* 9.1-2 (1986),

[163] Wilfred Keith Hastings. Monte Carlo sampling methods using Markov chains and their applications. In: *Biometrika* (1970).

[164] Simon Duane, A.D. Kennedy, Brian J. Pendleton, and Duncan Roweth. Hybrid Monte Carlo. In: *Physics Letters B* 195.2 (1987),

[165] Matthew D Hoffman and Andrew Gelman. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. In: (2014),

[166] Alp Kucukelbir, Dustin Tran, Rajesh Ranganath, Andrew Gelman, and David M. Blei. Automatic Differentiation Variational Inference. In: *arXiv:1603.00788 [cs, stat]* (2016). arXiv: 1603.00788.

[167] Michael I. Jordan, Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul. An Introduction to Variational Methods for Graphical Models. In: *Learning in Graphical Models*. Ed. by Michael I. Jordan. Dordrecht: Springer Netherlands, 1998, pp. 105–161.

[168] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational Inference: A Review for Statisticians. In: *Journal of the American Statistical Association* 112.518 (2017). arXiv: 1601.00670,

[169] Jake VanderPlas. Frequentism and Bayesianism: A Python-driven Primer. In: *arXiv:1411.5018 [astro-ph]* (2014). arXiv: 1411.5018.

[170] E.T. Jaynes. Confidence Intervals vs Bayesian Intervals. In: *Papers on Probability, Statistics and Statistical Physics* (1976).

[171] Alexandros Karatzoglou, Alex Smola, Kurt Hornik, and Achim Zeileis. kernlab - An S4 Package for Kernel Methods in R. In: *Journal of Statistical Software* 11.9 (2004).

[172] John Salvatier, Thomas V. Wiecki, and Christopher Fonnesbeck. Probabilistic programming in Python using PyMC3. In: *PeerJ Computer Science* (2016).

[173] Benedetto Grillone, Gerard Mor, Stoyan Danov, Jordi Cipriano, and Andreas Sumper. A data-driven methodology for enhanced measurement and verification of energy efficiency savings in commercial buildings. In: *Applied Energy* 301 (2021),

[174] EN-TRACK - Energy Efficiency Performance-Tracking Platform for Benchmarking Savings and Investments in Buildings. 2021.

[175] BIGG - Building Information aGGregation, harmonization and analytics platform. 2021.

[176] Lara Quijano-Sánchez, Iván Cantador, María E. Cortés-Cediel, and Olga Gil. Recommender systems for smart cities. In: *Information Systems* 92 (2020),

[177] Eleni Fotopoulou, Fernando Terroso, Antonio Skarmeta, Umutcan Şimşek, and Anna Fensel. Data Aggregation, Fusion and Recommendations for Strengthening Citizens Energy-aware Behavioural Profiles. In: (2017),

[178] Tiago Pinto, Ricardo Faia, Maria Navarro-Caceres, Gabriel Santos, Juan Manuel Corchado, and Zita Vale. Multi-Agent-Based CBR Recommender System for Intelligent Energy Management in Buildings. In: *IEEE Systems Journal* 13.1 (2019),

[179] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. In: *IEEE Transactions on Knowledge and Data Engineering* 17.6 (2005),

[180] Clayton Miller. Predicting success of energy savings interventions and industry type using smart meter and retrofit data from thousands of non-residential buildings. In: *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*. Delft Netherlands: ACM, 2017, pp. 1–4.

[181] Yujie Xu, Vivian Loftness, and Edson Severnini. Using Machine Learning to Predict Retrofit Effects for a Commercial Building Portfolio. In: *Energies* 14.14 (2021),

[182] Gerard Mor, Jordi Cipriano, Giacomo Martirano, Francesco Pignatelli, Chiara Lodi, Florencia Lazzari, Benedetto Grillone, and Daniel Chemisana. A data-driven method for unsupervised electricity consumption characterisation at the district level and beyond. In: *Energy Reports* 7 (2021),

[183] scikit learn. Multiclass and multioutput algorithms. 2020.