KU LEUVEN

ARENBERG DOCTORAL SCHOOL
Faculty of Engineering Science

Times series forecasting, applications on low voltage grid data

Lola Botman

Supervisor:
Prof. dr. ir. B. De Moor

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

January 2024

Times series forecasting, applications on low voltage grid data

Lola BOTMAN

Examination committee:
Prof. dr. ir. P. Wollants, chair
Prof. dr. ir. B. De Moor, supervisor
Prof. dr. ir. R. Belmans
Prof. dr. ir. J. Suykens
Dr. ir. J. Lago
(Amazon)
Prof. dr. ir. F. Vallée
(University of Mons)
Dr. ir. K. Vanthournout
(VITO)
Dr. ir. C. Bastiaensen
(Elia)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

January 2024

© 2024 KU Leuven – Faculty of Engineering Science
Uitgegeven in eigen beheer, Lola Botman, Kasteelpark Arenberg 10, 3001 Leuven, B-3001 Leuven (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm, electronic or any other means without written permission from the publisher.

Abstract

In an era marked by the pressing need to address climate change, innovative solutions in energy management are becoming increasingly crucial. In this thesis, three main contributions to the field of energy forecasting are presented, each addressing distinct aspects of individual household and plug electricity consumption. The first contribution is a novel method for long-term forecasting of individual household electricity usage, demonstrating high accuracy and scalability. The second introduces a global probabilistic approach for short-term forecasting of individual households, particularly robust to changing patterns. The third establishes an innovative pipeline for optimizing energy efficiency in office buildings through plug scheduling. Each of these contributions underscores the pivotal role of advanced forecasting in energy management. But how exactly does energy time series forecasting intertwine with the broader goals of the energy transition?

Global efforts intensify to address the climate crisis. Adopting more renewable energies is essential to shape a sustainable future. The United Nations' Sustainable Development Goals are central to this evolution. Goal #7 is specifically related to the energy transition with the objective of "ensuring access to affordable, reliable, sustainable, and modern energy for all by 2030". A main source of progress towards Sustainable Development Goal #7 is decarbonization by increased electrification. Accurate energy forecasting then emerges as indispensable. As the transition from traditional vehicles to electric cars accelerates, the demand for home charging stations and public charging infrastructures intensifies. This not only increases overall electricity consumption but also poses the risk of grid congestion, especially during peak times when many return home to charge simultaneously. To efficiently manage and mitigate such challenges, precise forecasting of household electricity demand becomes imperative.

Energy forecasting literature has traditionally focused on long-term or coarse spatially aggregated predictions. These predictions typically represent national-

level or production-oriented estimates. The introduction of smart meters and smart plugs has made new data available. This allows for more detailed forecasting, at the individual household level or at the device level. Yet, inherent challenges accompany this shift. First, even though data is becoming more accessible, obtaining datasets remains challenging, largely because of privacy concerns. Additionally, it is crucial to have representative datasets that capture current trends, like the rise in heat pump usage or the increase in remote work due to the COVID situation. Secondly, maintaining high data quality is crucial. Any compromise can significantly degrade predictions. Thirdly, the stochastic patterns in individual household energy use need to be specifically addressed. The spikes in consumption are primarily influenced by human behavior, like when residents return home and begin using appliances. Lastly, ensuring scalability is essential, especially when forecasts are to be applied across thousands of households, for example.

Three principal research components form this thesis. The first tackles the so far unaddressed problem of long-term individual household forecasting. The proposed method for predicting both monthly and yearly electrical consumption of individual households one year ahead was developed applying pre-processing, data augmentation, clustering, and ensemble learning to produce precise forecasts. The method is able to predict an entire year ahead, even with limited historical data, and requires no specific household attributes. The method is scalable and accurate, as demonstrated by its top-ranking in the IEEE-CIS competition. This method can be applied to ensure accurate billing capabilities for suppliers.

The second method, producing short-term forecasts of individual household consumption, tackles the double peak penalty effect. This double penalty effect arises when point forecasting methods accurately predict a peak but with a slight timing error, incurring penalties both when the peak is unpredicted and when it is wrongly anticipated. However, predicting a peak with a slight delay is more valuable than missing it entirely. Employing probabilistic forecasts effectively mitigates this issue. A novel global probabilistic method is introduced, which requires only a week of historical data from the concerned household to make accurate one-day-ahead predictions, by leveraging other households historical data. Its adaptability and robustness has been assessed through evaluation across three different datasets and benchmarked with several state-of-the-art methods. The method, unlike all other benchmarks, adapts quickly when large changes in patterns occur, such as a new family moving in or a holiday period.

The third component establishes a comprehensive pipeline to reduce office building energy consumption. This involves automatically determining if a plug is idle or active, forecasting its usage, and then scheduling its operation. With electricity savings reaching up to 20%, this method showcases the

potential for demand flexibility and optimizing building energy efficiency. The automatic detection of idle versus active mode, using an ensemble of unsupervised clustering, fills a void in the literature. Existing approaches identify this idle/active threshold either by visual assessment or through hands-on experiments. The data, sourced from the University of California, San Diego, consists of over 150 individual plug consumption profiles, over more than a year. It is made openly accessible to the community, and is the first of its kind.

A significant overarching contribution of this research, is its emphasis on transparency and reproducibility. By providing access to the developed code, introducing a new dataset, and always utilizing open datasets, the study ensures that both the scientific community and industry professionals can derive value from it. This open approach facilitates rigorous comparisons in subsequent research and fosters a collaborative environment for further advancements in the field.

Beknopte samenvatting

In een tijd gekenmerkt door de hoge nood om klimaatverandering tegen te gaan, is innovatief energiebeheer steeds crucialer. In deze thesis worden drie belangrijke bijdragen op het gebied van energievoorspellingen gepresenteerd met betrekking tot het elektriciteitsverbruik van individuele huishoudens en het beheer van stopcontacten. De eerste bijdrage is een nieuwe methode voor langetermijnvoorspelling van het elektriciteitsverbruik van individuele huishoudens, die nauwkeurig en schaalbaar is. De tweede introduceert een globale probabilistische methode voor kortetermijnvoorspellingen van individuele huishoudens, die bijzonder robuust is voor veranderende patronen. De derde levert ten slotte een innovatieve pijplijn voor het optimaliseren van de energie-efficiëntie in kantoorgebouwen door middel van automatische schakeling van stekkers. Elk van deze bijdragen benadrukt de belangrijke rol van geavanceerde voorspellingen in energiebeheer. Maar hoe past tijdreeksvoorspelling van energieverbruik precies in het bredere plaatje van de energietransitie?

Wereldwijd worden de inspanningen om de klimaatcrisis aan te pakken versterkt. Het gebruik van meer hernieuwbare energie is essentieel voor een duurzame toekomst. De *Sustainable Development Goals* van de Verenigde Naties staan centraal in deze ontwikkeling. Doelstelling 7 heeft specifiek betrekking op de energietransitie met als doel “iedereen tegen 2023 toegang te geven tot betaalbare, betrouwbare, duurzame en moderne energie”. Een belangrijke stap om *Sustainable Development Goal 7* te halen, is decarbonisatie door meer elektrificatie. Naarmate de overgang van traditionele voertuigen naar elektrische auto’s versnelt, neemt de vraag naar oplaadpunten thuis en openbare oplaadinfrastructuren toe. Dit verhoogt niet alleen het totale elektriciteitsverbruik, maar brengt ook het risico van overbelasting van het elektriciteitsnet met zich mee, vooral tijdens piekuren wanneer veel mensen tegelijk thuiskomen om hun wagen op te laden. Om dergelijke uitdagingen efficiënt te beheren en te beperken, is het absoluut noodzakelijk om het elektriciteitsverbruik van huishoudens nauwkeurig te voorspellen.

De literatuur omtrent energievoorspellingen richt zich traditioneel altijd op voorspellingen op lange termijn of op brede, ruimtelijk geaggregeerde voorspellingen. Deze voorspellingen zijn meestal gebaseerd op schattingen op nationaal niveau of productiegerichte schattingen. Door de introductie van slimme meters en slimme stekkers zijn er echter nieuwe gegevens beschikbaar die meer gedetailleerde voorspellingen mogelijk maken, op het niveau van individuele huishoudens of apparaten. Deze ontwikkeling gaat echter gepaard met de bijhorende uitdagingen. Ten eerste blijft het verkrijgen van datasets een uitdaging, vooral door bezorgdheden over privacy, ook al worden gegevens toegankelijker. Daarnaast is het cruciaal om representatieve datasets te hebben die huidige trends vastleggen, zoals de toename in warmtepompverbruik of de toename in telewerk als gevolg van de COVID-situatie. Ten tweede is het van cruciaal belang om de kwaliteit van de gegevens hoog te houden. Elk compromis betreffende de kwaliteit kan de voorspellingen aanzienlijk verslechteren. Ten derde moeten de stochastische patronen in het energieverbruik van individuele huishoudens specifiek worden aangepakt. De pieken in het verbruik worden voornamelijk beïnvloed door menselijk gedrag, zoals het tijdstip waarop bewoners thuiskomen en apparaten gaan gebruiken. Ten slotte is schaalbaarheid essentieel, vooral wanneer voorspellingen moeten worden toegepast op bijvoorbeeld duizenden huishoudens.

Deze thesis bevat drie belangrijke onderzoekscomponenten. De eerste pakt het tot nu toe onopgeloste probleem van langetermijnvoorspellingen voor individuele huishoudens aan. De voorgestelde methode, om zowel het maandelijkse als jaarlijkse elektriciteitsverbruik van individuele huishoudens te voorspellen, werd ontwikkeld met behulp van *pre-processing*, *data augmentation*, *clustering* en *ensemble learning* om nauwkeurige voorspellingen te produceren. De methode is in staat om een heel jaar vooruit te voorspellen, zelfs met beperkte historische gegevens, en vereist geen specifieke eigenschappen van de huishoudens. De methode is schaalbaar en nauwkeurig, zoals blijkt uit de toppositie behaald in de IEEE-CIS competitie. Deze methode kan worden toegepast om leveranciers nauwkeurige factureringsmogelijkheden te bieden.

De tweede methode, die kortetermijnvoorspellingen van het verbruik van individuele huishoudens maakt, pakt het dubbele *penalty-effect* aan. Dit dubbele *penalty-effect* treedt op wanneer puntvoorspellingsmethoden nauwkeurig een piek voorspellen, maar met een kleine verschuiving in de tijd, waardoor er waardoor er tweemaal een fout wordt gemaakt, zowel wanneer de piek niet wordt voorspeld als wanneer er verkeerd op wordt geanticipeerd. Een piek met een kleine vertraging voorspellen is echter waardevoller dan hem helemaal missen. Het gebruik van probabilistische voorspellingen vermindert dit probleem effectief. Hier wordt dit probleem aangepakt door middel van een nieuwe globale probabilistische methode, die slechts een week aan historische gegevens van het betreffende

huishouden nodig heeft om nauwkeurige voorspellingen één dag vooruit te doen, door gebruik te maken van historische gegevens van andere huishoudens. Het aanpassingsvermogen en de robuustheid van de methode worden bevestigd aan de hand van een evaluatie op drie verschillende datasets en een benchmark met verschillende geavanceerde methoden. De methode past zich, in tegenstelling tot alle andere benchmarks, snel aan wanneer zich grote veranderingen in patronen voordoen, zoals een nieuw gezin of een vakantieperiode.

De derde onderzoekscomponent stelt een uitgebreide pijplijn voor om het energieverbruik van kantoorgebouwen te verminderen. Deze bepaalt automatisch wanneer een stekker inactief of actief is, vervolgens wat het gebruik is en de inschakeling ervan. Met elektriciteitsbesparingen die kunnen oplopen tot 60%, laat deze methode het potentieel zien voor vraagflexibiliteit en het optimaliseren van de energie-efficiëntie van gebouwen. De automatische detectie van de inactieve versus actieve modus van stekkers, met behulp van een ensemble van *unsupervised clustering*, ontbrak tot op heden. Bestaande benaderingen identificeren deze inactieve/actieve drempel ofwel door visuele beoordeling ofwel door handmatige experimenten. De gegevens, afkomstig van de Universiteit van Californië in San Diego, bestaan uit meer dan 150 individuele verbruiksprofielen van stekkers over een periode van meer dan een jaar. De gegevens zijn publiek toegankelijk gemaakt voor de gemeenschap en zijn een eerste in hun soort.

Een belangrijke overkoepelende bijdrage van dit onderzoek is de nadruk op transparantie en reproduceerbaarheid. Door toegang te geven tot de ontwikkelde code, een nieuwe dataset te introduceren en altijd open datasets te gebruiken, zorgt dit onderzoek ervoor dat zowel de wetenschappelijke gemeenschap als professionals uit de industrie er waarde uit kunnen halen. Deze open aanpak vergemakkelijkt rigoureuze vergelijkingen in volgend onderzoek en bevordert een samenwerkingsomgeving voor verdere vooruitgang in het veld.

Contents

Abstract iii

Beknopte samenvatting vii

Contents xi

List of Figures xv

List of Tables xvii

1 Introduction 1
- 1.1 United Nations Sustainability Development Goal #7: Ensure Access to Affordable, Reliable, Sustainable, and Modern Energy 2
- 1.2 The low-voltage grid as enabler of the United Nations Sustainability Development Goal #7 . . . 4
  - 1.2.1 The Belgian electricity grid . . . 4
  - 1.2.2 Digital twin of the low-voltage grid . . . 7
  - 1.2.3 The importance of data . . . 8
- 1.3 Time series forecasting . . . 10
  - 1.3.1 Time series forecasting literature and terminology . . . 10
  - 1.3.2 Focus on load forecasting literature . . . 14
  - 1.3.3 Challenges in low-voltage grid load forecasting . . . 15
- 1.4 Structure objectives and significance of the thesis . . . 19
- 1.5 Acknowledgments . . . 24

2 A scalable ensemble approach to forecast the electricity consumption of households 25
- 2.1 Introduction . . . 27
  - 2.1.1 Related work . . . 27
  - 2.1.2 Motivation and contributions . . . 29
  - 2.1.3 Organization of the paper . . . 30

2.2 Key attributes of the model 31
2.3 Methodology: Ratio approach 32
2.3.1 Data Augmentation & Normalization: relative consumption profiles 33
2.3.2 Clustering: grouping similar profiles 35
2.3.3 Prediction: the power of the median via ensemble learning 38
2.3.4 Ensemble learning 41
2.3.5 Post-processing: smoothing the predictions 42
2.4 Case Study 42
2.4.1 Problem statement IEEE-CIS Competition 42
2.4.2 Dataset 43
2.4.3 Performance Metrics 45
2.4.4 Benchmark Models 45
2.4.5 The proposed approach with the battery trick 46
2.4.6 Run times 46
2.4.7 Results & discussion 47
2.4.8 Details on the results 49
2.4.9 Hyperparameter sensitivity study of the proposed method 49
2.5 Conclusion 50

3 Probabilistic short-term forecasting of individual households consumption in low voltage grids 53
3.1 Introduction 55
3.1.1 Related work 55
3.1.2 Motivation and contributions 57
3.1.3 Organization of the paper 58
3.2 Background Knowledge 58
3.2.1 Empirical quantiles 58
3.2.2 Benchmark methods 60
3.2.3 Performance Metrics 66
3.3 Methodology: Global Approach 69
3.3.1 Data cleaning 69
3.3.2 Normalization 69
3.3.3 Selection of similar profiles 71
3.3.4 Prediction 72
3.4 Case Study 72
3.4.1 Problem statement 72
3.4.2 Datasets 73
3.4.3 Results & discussion 74
3.4.4 Hyperparameter sensitivity study of the proposed method 76
3.5 Conclusion 82

4 Building consumption reduction through plug load mode detection, forecasting and scheduling 83
4.1 Introduction . . . 85
4.2 Motivation and contributions . . . 85
4.3 Literature review . . . 86
4.3.1 Smart plug hardware . . . 86
4.3.2 Smart plug home vs. office set-up and integration . . . 86
4.3.3 Smart plug open data . . . 86
4.3.4 Non-invasive and invasive load monitoring . . . 87
4.3.5 Scheduling smart plugs . . . 88
4.3.6 Forecasting plug loads . . . 89
4.3.7 Plug operating modes detection terminology . . . 90
4.3.8 Plug operating modes detection methodologies . . . 91
4.4 Methodology . . . 95
4.4.1 Data preparation . . . 95
4.4.2 Operating mode detection . . . 96
4.4.3 Forecasting . . . 100
4.4.4 Scheduling . . . 103
4.5 Case Study . . . 105
4.5.1 Problem statement . . . 105
4.5.2 Dataset . . . 105
4.5.3 Results & discussion . . . 106
4.6 Future research directions . . . 119
4.7 Conclusion . . . 120

5 Conclusion and perspectives 121
5.1 Scientific conclusions . . . 121
5.2 Practical takeaways . . . 124
5.3 Research perspectives . . . 125
5.3.1 Perspectives in load forecasting . . . 126
5.3.2 Perspectives in the energy landscape . . . 128

Bibliography 133

List of Figures

1.1 The lay-out and stakeholders of the Belgian electrical grid. . . . 5
1.2 The Belgian energy mix in 2022. 6
1.3 Digital twin of the low-voltage grid. 9
1.4 Illustration of historical data, forecast origin, and forecast horizon. 13
1.5 Example of a point forecast versus a probabilistic forecast. . . . 13
1.6 Stochasticity of individual household consumption profiles. . . . 17
1.7 Concept drift. 18
1.8 Situating the research chapters along the spatial aggregation and horizon dimensions and their contributions to SDG7 and the load forecasting challenges. 23

2.1 Flowchart of the scalable ensemble approach. 34
2.2 Example of clusters of relative consumption profiles starting in January. 37
2.3 Example of clusters of relative consumption profiles starting in June. 37
2.4 Heat map of the monthly electrical consumption per smart household. 44

3.1 Empirical and true cumulative distribution function. 59
3.2 Flowchart of the approach Feedforward Neural Network & Empirical Quantiles. 63
3.3 Computing and grouping errors on the validation set to compute quantiles. 65
3.4 CRPS Visualizations. 68
3.5 Visualization of global approach and benchmark results. 78
3.6 Scatter plots of the global approach results compared with best second (or first) benchmark. 79
3.7 Visualization of global approach forecasts on profiles experiencing concept drift. 80

4.1 Proposed pipeline for plug load scheduling. 95
4.2 Heatmap of a plug. 99
4.3 Examples of mode detection results classified as high-fidelity. . 114
4.4 Examples of mode detection results classified as partial-fidelity. 116
4.5 Examples of mode detection results classified as low-fidelity. . . 118

List of Tables

2.1 Comparison of three benchmark methods (naive baseline, linear regression and ARIMA) and the top three methods in the IEEE-CIS contest. 48
2.2 Comparison of the top three methods' original submissions in the IEEE-CIS contest when removing faulty smart meters of the evaluation. 48
2.3 Comparison of the performance metrics of the method when varying the hyperparameters. 51

3.1 Table of global approach results and benchmarks. 77
3.2 Sensitivity analysis results. 81

4.1 Mode detection results. 107
4.2 Forecasting models evaluated on the test set, i.e., the last 20% of each plug load time series and averaged over all plugs. 110
4.3 The different forecast-driven schedules are compared with the naive baseline schedule and no schedule. 112

Chapter 1

Introduction

This dissertation dives into the urgent matter of climate change, aiming to leverage forecasting as a pivotal tool in the energy transition. It contributes new methodologies in energy forecasting, offering insights and practical solutions that can support the broader efforts in addressing climate change and facilitating the energy transition.

In the first part, the thesis introduces a novel approach to long-term household electricity usage forecasting, showcasing both accuracy and scalability. This method predicts monthly and yearly consumption up to a year ahead, without needing detailed household data. This long-term forecast, crucial yet relatively unexplored, plays a vital role in accurate billing and efficient grid operation planning. In the second part, the thesis delves into short-term household forecasting. The developed global probabilistic approach effectively addresses the double peak penalty effect, offering accurate next-day probabilistic forecasts. This method robustness is confirmed across diverse datasets, making it superior in forecasting changing consumption profiles like household composition shift or holiday periods. These forecasts are key in triggering mitigation strategies for network congestion. In the third part, the thesis presents a novel pipeline for energy savings in office buildings, focusing on plug mode detection, forecasting, and scheduling. This method can reduce electricity consumption by up to 20% by automating the detection of idle versus active plug modes in smart plug appliances.

We will further unravel the connection between these forecasting methodologies and their impact on the energy transition in this introductory chapter, highlighting both the scientific and industrial significance of the research.

1.1 United Nations Sustainability Development Goal #7: Ensure Access to Affordable, Reliable, Sustainable, and Modern Energy

The world is currently undergoing a profound energy transition, related to the pressing climate crisis, resonating through headlines, policy agendas, and conversations at all levels of society. In 2015, the United Nations introduced the "2030 Agenda for Sustainable Development", consisting of 17 Sustainable Development Goals to address pressing global issues. Among these goals, Sustainable Development Goal #7 (SDG7) stands out, explicitly stating the need to **ensure access to affordable, reliable, sustainable, and modern energy for all.**. While all Sustainable Development Goals are interconnected, SDG7 holds a special place in the context of the energy sector and the ongoing energy transition.

Access to affordable, reliable, sustainable, and modern energy services is crucial for the development of all sectors of society: agriculture, business, communications, education, healthcare, transportation, and technology. Lack of access to electricity has far-reaching consequences. Electricity is essential in healthcare, to power facilities and provide clean water for hygiene. It is crucial in education, to enable students to study and complete homework at night. In terms of economic opportunities, electricity is necessary to support competitive businesses and economic growth. Furthermore, electricity access for all would allow for more gender equality, to empower women and girls, and reduce their daily burden. Indeed, they suffer more from this lack of access, since they are still, up to day, made responsible for tasks such as fetching water and ensuring access to clean cooking solutions.

The latest Sustainable Development Goals report from 2023 [136] mentions some critical figures, that make the challenge of energy accessibility for all visible. Even as global access to electricity has improved, from 87% in 2015 to 91% in 2021, approximately 660 million people still lack access to electricity, primarily in low-developed countries and sub-Saharan Africa. Nearly 2 billion people rely on polluting fuels and technologies for cooking, and 675 million people remain unconnected to grids. The energy sector is the largest contributor to climate change, and accounts for about 60% of global greenhouse gas emissions [136]. Addressing energy-related emissions is vital for combating climate change.

SDG7 includes specific targets to guide global efforts in the energy sector [136]:

1. By 2030, ensure universal access to affordable, reliable, and modern energy services.
2. By 2030, substantially increase the share of renewable energy in the global energy mix.
3. By 2030, double the global rate of improvement in energy efficiency.
4. By 2030, enhance international cooperation to facilitate access to clean energy research and technology, including renewable energy, energy efficiency, and advanced and cleaner fossil-fuel technology, and promote investment in energy infrastructure and clean energy technology.
5. By 2030, expand infrastructure and upgrade technology for supplying modern and sustainable energy services for all in developing countries, including least developed countries, small island developing States, and land-locked developing countries, following their respective programs of support.

The European Union has aligned its policymaking and directives with the targets of SDG7. Among many initiatives, two directives are particularly important. First, the Energy Efficiency Directive suggests a 11.7% reduction of the energy consumption in the EU by 2030. Secondly, a revised Renewable Energy Directive sets target of 42.5% of renewable in energy consumption [3].

To meet SDG7, and related European targets by 2030, we must speed up electrification; increase investments in renewable energy sources such as solar, wind, and thermal; improve energy efficiency; and develop enabling policies and regulatory frameworks [136].

The aforementioned initiatives, such as the acceleration of electrification, give rise to two primary challenges in the context of energy supply and demand [57]. First, the increase in electrical appliances, exemplified by the transition from boilers to heat pumps in households and the expanding prevalence of electric vehicles, leads to an increase in overall electricity consumption, in peak consumption, and also in volatility, i.e., rapid fluctuations. Second, the growing share of renewable energy sources introduces uncertainty in energy production, particularly due to their susceptibility to weather conditions.

To address these associated concerns pertaining to **electrification** and the augmented integration of renewables, diverse mitigation strategies have been

designed [57]. These encompass storage control mechanisms, involving the strategic charging of storage units during periods of low demand to facilitate surplus energy discharge during peak demand intervals; demand-side response mechanisms, entailing the selective deactivation of specific appliances or street lighting in response to inadequate energy generation; coordinated electric vehicle (EV) charging schemes, aiming to distribute EV charging activities evenly to prevent network overloads during peak hours; and the establishment of local electricity markets to facilitate the exchange of surplus electricity among neighboring consumers. Each of these solutions requires accurate foresight of the electricity consumed and/or generated, which is where load forecasting comes into play.

1.2 The low-voltage grid as enabler of the United Nations Sustainability Development Goal #7

1.2.1 The Belgian electricity grid

The electrical grid is a sophisticated infrastructure designed to ensure the reliable and efficient distribution of electricity from diverse energy sources to end consumers. It comprises several crucial components and involves various stakeholders, each playing an essential role in the process, as illustrated in **Figure 1.1**.

Electricity is generated by a mix of technologies, including wind turbines, solar panels, gas-fired power plants, nuclear facilities, and more, illustrated on **Figure 1.2**. The energy mix evolves over the years as the renewable energies share is increasing. Belgium strives to align with its sustainability goals. The key **energy producers** in Belgium include Engie Electrabel, EDF Luminus and EON [34].

The energy producers inject the generated electricity into the high-voltage grid, also called the **transmission grid**, responsible for efficiently transporting electricity over long distances. It operates at voltage levels ranging from 30 kV to 380 kV, and is monitored and managed by the transmission system operator, Elia [40]. They have metering facilities, to record and check exchanges of energy on the grid every quarter-hour. Elia oversees this network of over $8,965$ km of connections, consisting of overhead lines ($5,596$ km) and underground cables ($3,369$ km). The high-voltage transmission network is connected to the distribution grid though transformer substations, electricity is then delivered to the majority of consumers. Additionally, numerous heavy industrial consumers are connected directly to the high-voltage transmission grid [69].

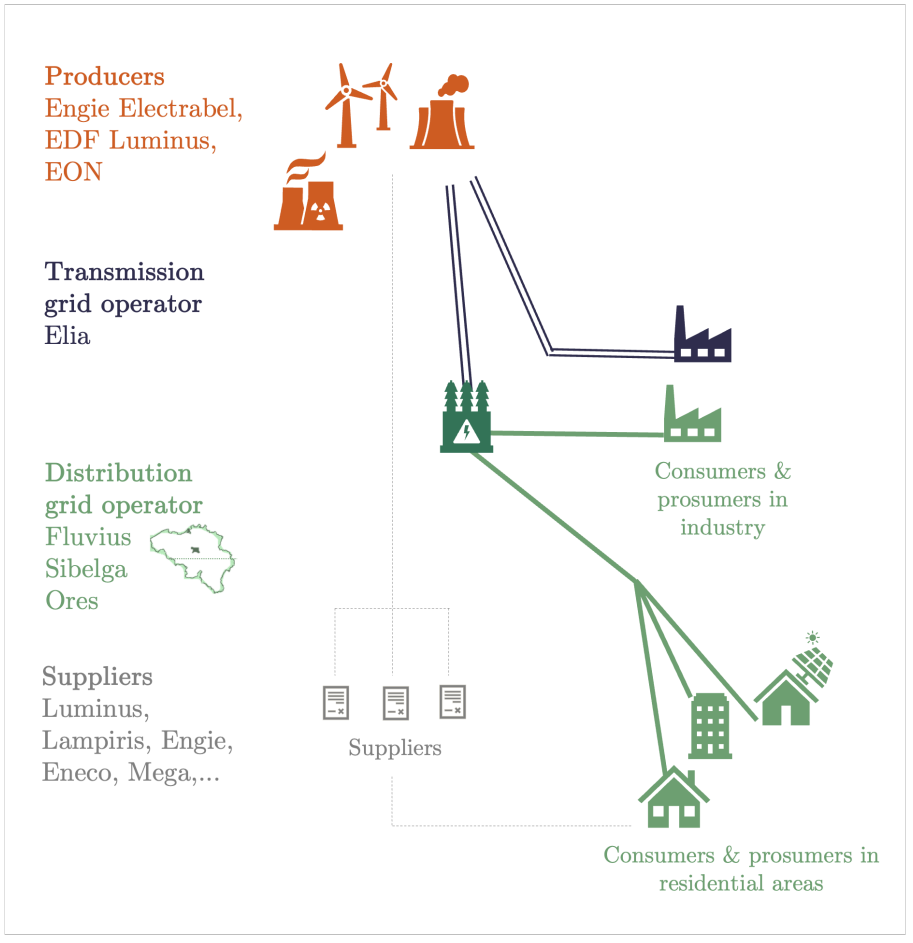


Figure 1.1. The lay-out and stakeholders of the Belgian electrical grid. The energy producers inject the generated electricity on the transmission grid, at high voltages. Transformer stations allow converting the electricity to lower voltages, suitable for the distribution grid, to be delivered to the individual consumers. Suppliers take care of selling the electricity to the consumers.

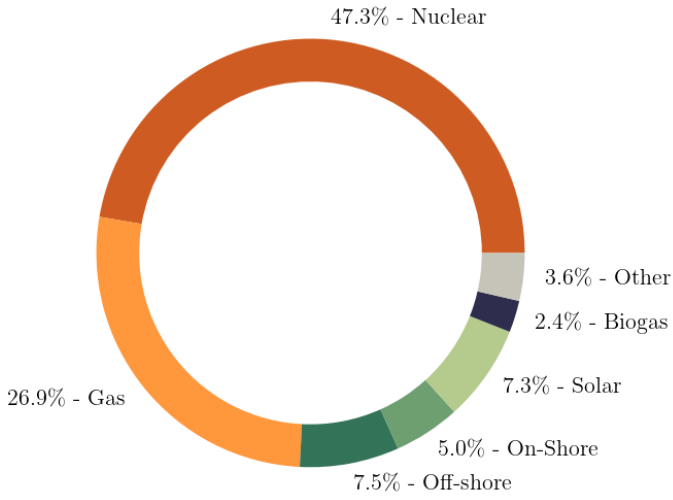


Figure 1.2. The figure illustrates the composition of Belgium's energy mix in the year 2022 [40], segmented into various sources. Nuclear energy comprises the largest portion at 47.3%, followed by natural gas at 26.9%. Renewable energy sources contribute significantly, with offshore wind accounting for 7.5%, onshore wind at 5.0%, and solar power at 7.3%. Biogas and other energy sources constitute a smaller fraction of the mix at 2.4% and 3.6%, respectively. This distribution highlights the country's reliance on nuclear and gas, while also emphasizing its investment in renewable energy.

The transformer stations, convert the voltage to lower levels, to then transmit the electricity to the **low-voltage network**, also called the distribution grid. In Flanders alone, the low-voltage grid encompasses an extensive network, comprising 3.5 million connections, 200,000 low-voltage feeders, 40,000 low-voltage transformers, and an impressive 84,000 kilometers of cable infrastructure, with a total asset value estimated at 11 billion euros. The behavior of the low-voltage grid is significantly influenced by two factors. The first is the grid topology, often characterized by incomplete or inaccurate data. Second, the offtake and injection profiles of consumers, exhibit high stochasticity and are constrained by limited measurements, even though now more and more digital meters are installed. The low-voltage grid plays a critical role in delivering electricity to individual households and smaller industrial consumers throughout Belgium. In addition to traditional end-consumers, who draw electricity from the grid, there are also more and more prosumers who generate electricity locally, often through technologies like solar panels. Prosumers may inject excess electricity back into the grid, contributing to the overall supply. Distribution System Operators (DSOs) provide new connections to the network and are responsible for reading electricity meters. The DSOs are different in each

region of Belgium: Fluvius in Flanders, Sibelga in Brussels, and Ores/RESA in Wallonia. Electricity is delivered to customers at the **suppliers** request, they are the entities selling electricity to the customers and making arrangements with the producers.

Regulatory oversight of the electricity market in Belgium involves various entities, including the CREG (Commission for Electricity and Gas Regulation) [34]. The CREG, an autonomous organization, ensures transparency, competition, and consumer protection in the electricity and gas markets. It plays a vital role in regulating the industry, approving transmission tariffs, and advising public authorities.

Belgium's electricity market is divided among different regions, and each region also has its regulatory body for energy. In Wallonia, it is the CWaPE (Commission wallonne pour l'énergie), in Flanders, it is the VREG (Vlaamse Regulator van de Elektriciteits- en Gasmarkt), and in Brussels, it is BRUGEL (Le régulateur bruxellois pour l'énergie). These regional regulators are responsible for overseeing the organization and operation of their respective electricity and natural gas markets, ensuring compliance with regional legislation, and providing mediation services to consumers in case of issues with suppliers or distribution network operators.

1.2.2 Digital twin of the low-voltage grid

We focus our attention on the intricate web of the low-voltage grid, the millions of underground cables, bringing electricity to individual households. Historically, this network was constructed using a "fit and forget" approach, where oversized cables were installed to accommodate any electricity demand. However, the energy landscape is shifting, on the one hand, characterized by an increase in overall consumption, exemplified by the increase in electric vehicles, and on the other hand, by a change in consumption patterns, with local renewable energy sources, such as solar panels. These types of consumption and production also tend to happen simultaneously, generating consumption peaks. If many people in a neighborhood come back from work at the same time and start charging their electrical vehicles, this puts a high load on the grid around similar times. And similarly for solar panels, when the sun shines around midday, and not many people are home to consume that locally produced energy, the surplus is injected into the grid and this generates simultaneous high loads on the grid. The grid was never designed to handle these modern consumption habits.

While the "fit-and-forget" strategy was efficient back in the days, and implemented with the available information, it has become increasingly unsustainable in the face of these changing consumption patterns. It is no

longer a viable strategy, due to the scale and complexity of the low-voltage grid. In Flanders, replacing even a mere percentage of such an extensive grid would entail astronomical costs, prompting the need for a more intelligent, cost-effective solution. The aim is to prioritize where to strengthen the grid first, and by how much it must be strengthened.

To meet these challenges head-on, EnergyVille is pioneering the development of digital twin capabilities for the low-voltage grid[1]. **A digital twin** is a virtual replica of a physical system or process, incorporating real-time data and simulations to enhance understanding, management, and decision-making. The developed digital twin capabilities harness the power of artificial intelligence and power system engineering to offer a more detailed understanding of the grid dynamics. At the heart of this toolkit lies a simulation environment. It computes currents and voltages throughout the low-voltage grid, providing a granular view of how electricity flows through this complex network. This digital twin, as illustrated on **Figure 1.3**, needs several inputs to provide detailed information on the grid operation. This includes the individual household consumption profiles, recorded by smart meters, as well as the grid topology, meaning, which household is connected to which cable. In order for the digital twin to generate this detailed information on the grid operation in the future, it is necessary to forecast the individual household consumption profiles. This is the main focus of this research.

1.2.3 The importance of data

In the era of the energy transition, characterized by a growing emphasis on decarbonization and electrification, data has emerged as a pivotal and crucial asset [112]. The number of smart power meters worldwide exceeded one billion last year, a ten-fold increase since 2010 [62]. Power grids, too, are becoming increasingly data-rich, with approximately 320 million distribution sensors deployed globally [62]. These devices generate a continuous stream of data that holds the key to unlocking new opportunities for energy optimization. Historically, many studies in the field of energy relied on **simulated data**, where assumptions and simplifications were made to model energy consumption and generation. While these simulations provided valuable insights, they often fell short of capturing the complexity and nuances of real-world energy systems,

[1] EnergyVille is active in the fields of sustainable energy and intelligent energy systems and is a collaboration between different Belgian research partners. Two universities are involved, KU Leuven and UHasselt along with two research centers, namely the Flemish Institute for Technological Research VITO and international research and development organization (Interuniversity Microelectronics Centre). Note that only the KU Leuven and VITO are involved in the digital twin project.

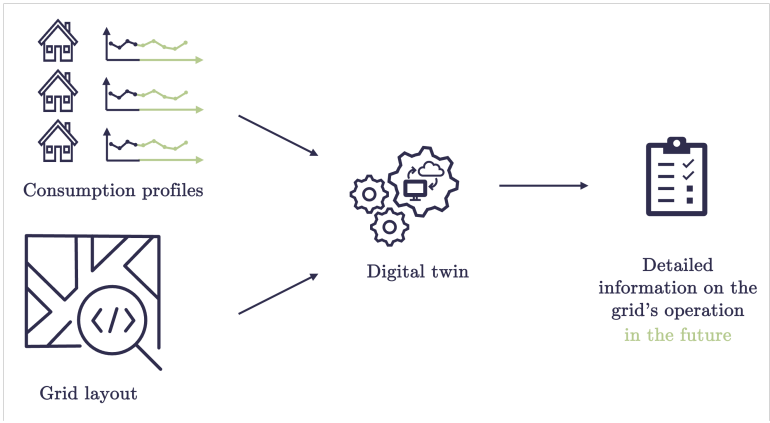


Figure 1.3. The digital twin, a virtual model, mirrors the physical low-voltage grid, improving its understanding and management. It requires household consumption profiles and grid topology as inputs to simulate the electrical currents and voltages within the low-voltage network. By analyzing these inputs, the digital twin provides precise information on the grid operations. Predicting the input consumption profiles allows to generate this precise information on the grid operations in the future.

especially at the low-voltage grid level. In contrast, the availability of **real-world data** offers opportunities for researchers to analyze and optimize energy systems with a higher degree of reliability.

While the potential benefits of data are vast, challenges persist. Accessing and harnessing the full potential of data in the energy sector remains a complex endeavor. Currently, only 2% to 4% of available smart meter data is used for the enhancement of grid operations. Transmission system operators acknowledge that control rooms often do not fully leverage data analytic applications, even in digitalized grids equipped with sensors and remote control capabilities. This underscores the untapped potential of data in optimizing energy systems.

To maximize the benefits of data in the energy sector, policymakers, and stakeholders must address several critical factors. Regulations should ensure easy consumer access to their data and enable seamless sharing with third parties, thereby promoting data availability. Additionally, the concept of "demand-response readiness" is gaining prominence, emphasizing the need for appliances to communicate using standardized protocols to simplify interoperability.

1.3 Time series forecasting

1.3.1 Time series forecasting literature and terminology

Forecasting models can be classified in three main groups, even if these groups and methods are connected in many ways [66]. The first to be developed were the statistical models, then the artificial intelligence and machine learning (ML) approaches rose in the 1990s [66] and more recently the deep-learning (DL) methods appeared.

Statistical methods encompass traditional time series methods like exponential smoothing, Autoregressive Integrated Moving Average, and state space models [36]. They generally offer efficiency, relying typically on a limited number of parameters. This makes them suitable for scenarios with limited historical data, or where model interpretability is needed (i.e., to earn trust with decision-makers [60]). Despite these benefits, statistical methods face challenges that make them unsuitable for complex forecasting problems: (i) they require strong assumptions about the data-generating process, (ii) to select the best approach, they require expert knowledge (iii) they struggle to model complex and nonlinear patterns.

Artificial intelligence and machine learning methods have emerged as powerful alternatives. These do not heavily rely on predefined assumptions about the data-generating process, making them more versatile in handling a wide range of data patterns [60]. ML models exhibit a "self-learning" property, employing nonlinear mechanisms and optimization algorithms to generate forecasts. ML methods also excel in their capacity to handle large, complex datasets and a multitude of parameters, the related drawback is the large amount of data necessary to train a model without risking overfitting. ML methods allow going over to *global* forecasting models, capable of providing accurate forecasts for multiple time series simultaneously. Traditional statistical approaches consider separate *local* models for each series. Furthermore, ML methods can efficiently incorporate exogenous variables as predictors, often enhancing forecasting accuracy. ML methods also demand a more extensive development effort compared to statistical models, as they involve fine-tuning hyperparameters and special care dedicated to preprocessing. Their computational cost, particularly during training, can be higher than that of traditional statistical models. This aspect is crucial for applications concerned with operational costs and timely forecasting. Some ML methods employed in time series forecasting include neural networks, regression trees, K-nearest neighbors regression, and support vector regression [60].

When it comes to **deep-learning**, no strict definition exists. DL can be

considered as neural networks with more processing layers, i.e., deeper neural networks, that aim to extract more advanced insights from the data [60]. DL methods include combinations of Recurrent Neural Networks and Convolutional Neural Networks (CNN), or CNNs and Long Short Term Memory networks. Lately, transformer models have shown high performance in language processing tasks, e.g., ChatGPT. Still, they have been used in time series forecasting tasks for the first time recently, hence literature is not abundant yet on the topic. In the latest review book on load forecasting [57], dating from January 2023, they have not been used in load forecasting (yet). DL methods share limitations with ML methods, i.e., they require more data and computational resources for training.

There is no consensus in the time series literature, nor in the load forecasting literature, about one model that would outperform all other models. It is highly dependent on the application and techniques should be chosen depending on different parameters such as the amount of data available, computing resources, the level of explainability required, the scalability, and the forecast uncertainty [57, 60].

Some specific forecasting concepts, definitions, and terminology used through the thesis are provided below, these come from the literature and more specifically from two recent books on forecasting [57, 60].

Historical data, forecast origin and **forecast horizon.** The historical data is the sequence of data points up to the current time, it is often used as a predictor. The forecast origin is the starting point for generating forecasts. The horizon specifies how many time steps beyond the forecast origin need to be predicted. If the horizon consists of only one value in the future, this can be called one-step-ahead predictions, as opposed to multi-step-ahead predictions where the horizon is larger than one. The concepts of historical data, origin, and horizon are depicted in **Figure 1.4** respectively in orange, gray, and green.

Spatial granularity. The spatial granularity, in load forecasting, is the level of load being forecast. A fine granularity could be a household load, and a coarser granularity is a city or countrywide load.

Point vs. probabilistic forecast. Point forecasts provide a single estimate for each time step in the forecast horizon, while probabilistic forecasts offer insights on the uncertainty by providing multiple values for each time step in the forecast horizon, such as quantiles or density functions. An example of a point forecast versus a probabilistic forecast is illustrated in **Figure 1.5**. The peak is underestimated in the point forecast, while the distribution prediction actually encompasses this peak. This figure is further detailed in Chapter 3.

Local vs. global models. A local model is trained and makes forecasts

on one single time series, while global models are trained and make forecasts on a set of (related) time series. Advantages of global approaches encompass enhanced computational efficiency, decreased requirements for development and maintenance efforts, as well as the ability to capture correlations among distinct series.

Ensemble models. Ensemble forecasts are built by combining the output of distinct models with mean, median, or more advanced weighting schemes. They have proven robust.

Recursive vs. direct strategies. Multi-step-ahead forecasting strategies include recursive strategies, i.e., one-step-ahead forecast is repeated using the previous forecast results, direct strategies, i.e., the entire horizon is predicted at once, and hybrid approaches that combine both.

Hierarchical forecasts. Hierarchical forecasts involve leveraging the hierarchy of the electrical grid and ensuring forecasts are coherent, e.g., the sum of the predicted loads of a street of households should match with the related substation load.

Peak forecast. In many applications, the peak load is of specific interest. This involves only predicting one value. It can be challenging since peaks are less frequent than base load and less representative training data are thus available. In some contexts, such as individual household load, the timing of the peak is also very relevant even though very stochastic.

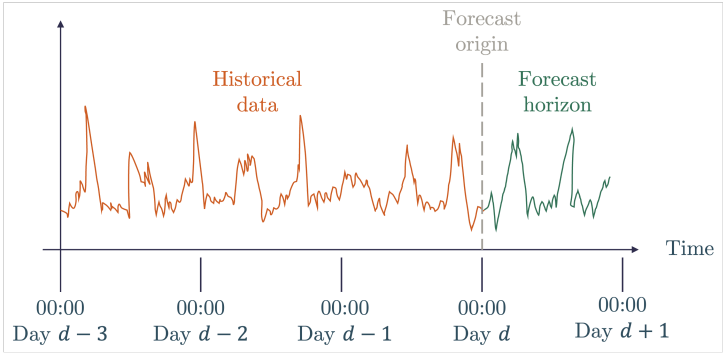


Figure 1.4. The historical data, depicted in orange, comprises the sequence of data points leading up to the forecast origin. In this example, it extends to three days before the forecast origin. The forecast origin, delineated by a gray dotted line, marks the start of forecast generation, here midnight on day d. The horizon is the extent of the forecast beyond the origin, which in this instance spans one day ahead. The number of time steps of the forecast is determined by the data resolution.

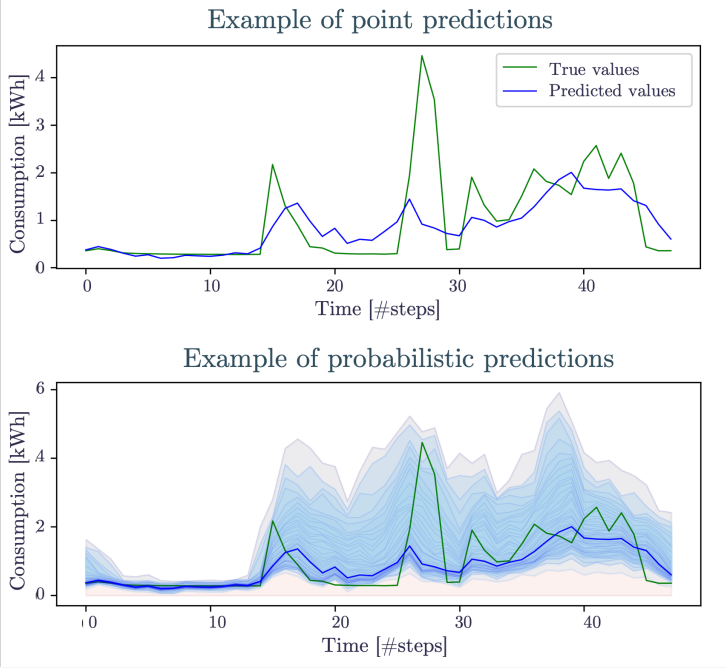


Figure 1.5. Example of a point forecast versus a probabilistic forecast. The peak is underestimated in the point forecast, while the distribution prediction actually encompasses this peak. This figure is further detailed in Chapter 3.

1.3.2 Focus on load forecasting literature

Load forecasting has been researched for over a century [66], ever since the first power plant was constructed in the United States. Beyond the model type, the load forecasting literature can be categorized along two dimensions: the **spatial granularity**, ranging from an individual appliance or household to neighborhood consumption, countrywide consumption, or even worldwide consumption; and secondly, the **forecasting horizon**, short-term (up to a day ahead), medium-term (up to a month ahead), or long-term (spanning several years). Different applications and end goals require different dimensions.

Most of the efforts in the literature have been focused on **long-term** forecasting of **coarse granularity**, e.g., cities, geographical zones [17], regions [107, 116] or countries [27, 45, 88]. This long-term coarse granularity research has primarily addressed operational concerns, production capacity estimations, and strategic planning [4]. These aggregated consumption profiles are relatively smooth, with discernible patterns and trends.

In recent times, the advances in smart metering rollout and subsequently, household-level data have spurred a growing interest in the **finer granularity of short to medium-term** forecasting [46, 77, 149, 150]. This navigates the complexities of forecasting within households, marked by the stochastic nature of the consumption profiles, highly dependent on human behavior. Its applications extend to peak and congestion forecasting, operational optimization, and management of more dynamic and unpredictable consumption patterns.

In parallel, a relatively underexplored frontier lies in **individual household long-term** forecasting, for applications in advanced payment systems and load scheduling. Existing research on long-term and fine granularity forecasting mostly focuses on large energy consumption units, e.g., supermarkets [26] or public buildings [92, 108], where the patterns are less stochastic than households.

Finally, the field of **plug-level** forecasting has emerged with the advent of smart plug systems. With data availability on the rise, this novel domain addresses challenges related to scheduling and optimizing energy consumption at the granularity of individual plugs, presenting opportunities for effective building energy management and resource allocation. A critical aspect within this field is the detection of plug operating modes, which entails discerning whether a plug power level indicates that the plug is active or not. A significant gap has been identified in the literature: the existing operating mode detection approaches all necessitate extensive manual work or tuning. These approaches can be broadly categorized into: Empirical techniques [10, 86, 141], Benchmarks and technical specification sheets [38, 76, 98], User feedback [49, 59], and Manual data labeling [87, 111, 146]. On the aspect of plug scheduling, the literature extensively

presents various approaches like time-based controls, occupancy-based controls, system-based controls, and more innovative hybrid controls [127]. However, a noticeable void is the absence of literature or methods proposing plug-level forecasting, which can then be used for scheduling. Notably, forecasting at the granularity of zones or rooms does exist [55], but individual plug load level forecasting remains unexplored.

1.3.3 Challenges in low-voltage grid load forecasting

Forecasting low-voltage grid loads at the individual household level and individual plug, i.e., individual appliance, presents a set of challenges that demand innovative solutions.

The first one has already been touched upon in Section 1.2.3: **data availability**. Obtaining relevant and timely data at this level of granularity is not easy. Typically, the recording and monitoring of demand is primarily done to the higher voltage levels of the network, and the roll-out of smart meters did not start until the 2010s [57]. Additionally, privacy considerations are substantial, as individual consumption data may potentially divulge sensitive information about occupants' daily routines and behaviors. Consequently, gaining access to such data often requires rigorous data agreements and privacy protections, which can be time-intensive to establish. Moreover, given the rapidly changing nature of energy consumption trends — marked by the increasing amount of electric vehicles, heat pumps, solar panels, and significant disruptions such as the COVID-19 pandemic — historical datasets, even those from just a few years ago, may no longer accurately reflect contemporary consumption patterns.

The second challenge is related to **data quality**. As the saying goes, "garbage in, garbage out": accurate forecasting depends on the quality of input data. This encompasses the often-invisible, yet immense, work of sourcing, formatting, merging, cleaning, and imputing data. These data intricacies can be a study in themselves, requiring rigorous attention before the actual forecasting research can start.

Thirdly, individual household consumption forecasting is highly subject to **stochasticity**, which impacts prediction accuracy. Unlike higher-level aggregations, i.e., substation-, city- or countrywide, where human behavior tends to average out, individual households exhibit unique consumption patterns influenced by intricate daily routines, as illustrated on **Figure 1.6**. We know that evening hours typically witness a peak in household consumption, but can we predict the exact moment each household returns home, flips on the lights, or starts cooking? Such unpredictability poses a significant challenge in forecasting. The **double peak penalty effect** [58] associated with specific

peak predictions frames this challenge. Traditional evaluation metrics or loss functions, such as Mean Absolute Error or Root Mean Square Error, struggle to accommodate to peak predictions in point forecasting models. When a forecast accurately predicts the magnitude of a peak but deviates slightly in timing, two penalties are incurred: one for forecasting a peak that does not materialize at that exact time and another for failing to predict the timing precisely when the actual peak occurs. As a result, conventional metrics may favor flat pattern forecasts, which ignore peaks altogether, even though a slightly miss-timed predicted peak would be more useful. Consequently, there is a pressing need for alternative assessment methods and loss functions in point forecasting that can more adeptly handle the burst-like patterns characterizing individual household consumption predictions. Another option is to focus on probabilistic forecasting methods, which can handle uncertainties in the data.

Fourth, global models also offer a solution to the **scalability** constraint of individual household and plug load forecasting. Many practical applications, including congestion forecasting and plug load scheduling, necessitate the simultaneous forecasting of hundreds of households or plugs. Therefore, even though the variable to be predicted is highly stochastic in nature, which usually requires more complex techniques, the developed models have to be scalable.

An additional challenge in time series is **concept drift**. This phenomenon occurs when models, usually trained on historical data, face significant changes in the environment they aim to predict. This is specific to individual household consumption. For instance, when a household changes owners, consumption habits may undergo substantial transformation, posing a unique challenge in model adaptability. Similarly, factors like vacations and unforeseen events like the COVID-19 pandemic can trigger sudden and unpredictable shifts in consumption patterns, making previously trained models inaccurate. Concept drift at the individual household level is illustrated in **Figure 1.7**. Global approaches have shown to adapt better to these changing patterns [60].

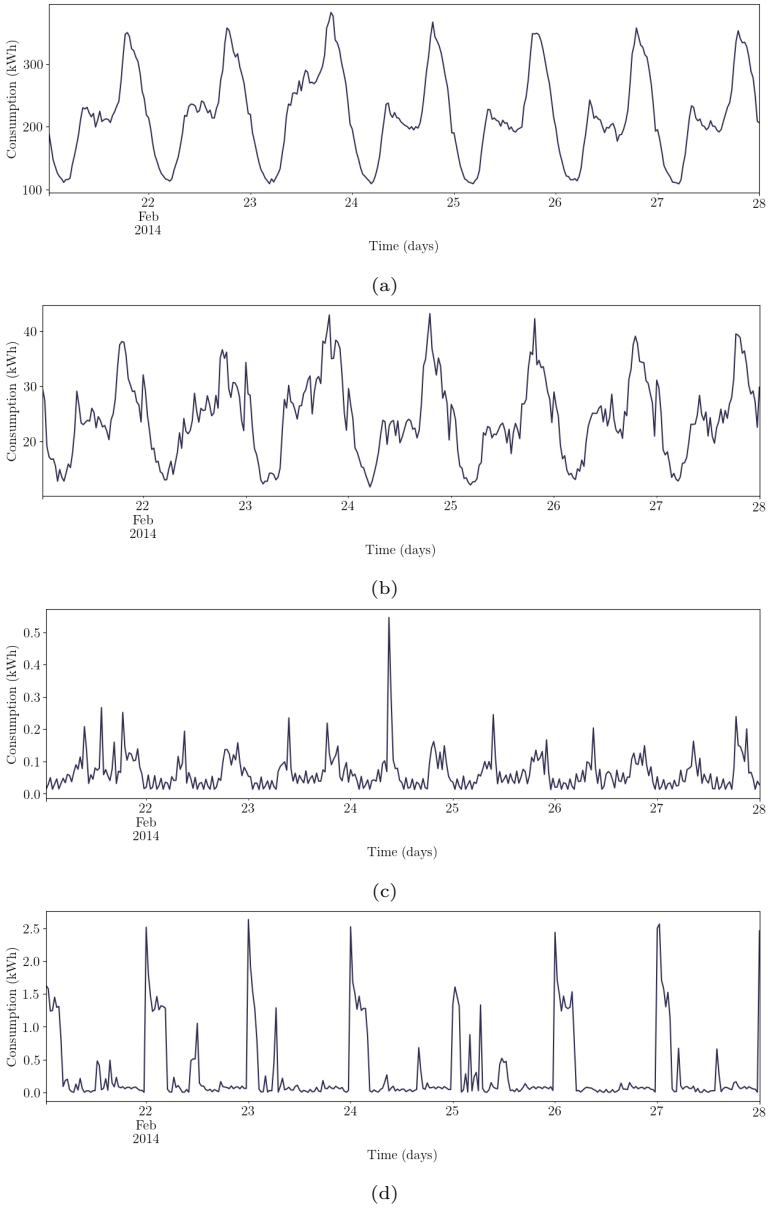


Figure 1.6. Example of different aggregation level and subsequent stochasticity of household consumption profiles (a) 1000 household profiles aggregated, (b) 100 household load profiles aggregated, (c) and (d) two different individual household profile from the same dataset [133], i.e., same country, same region, same days. The stochasticity of a single profile is higher than when profiles are aggregated. Predicting the consumption of a single household is complex, due to this stochasticity, highly dependent on human behavior.

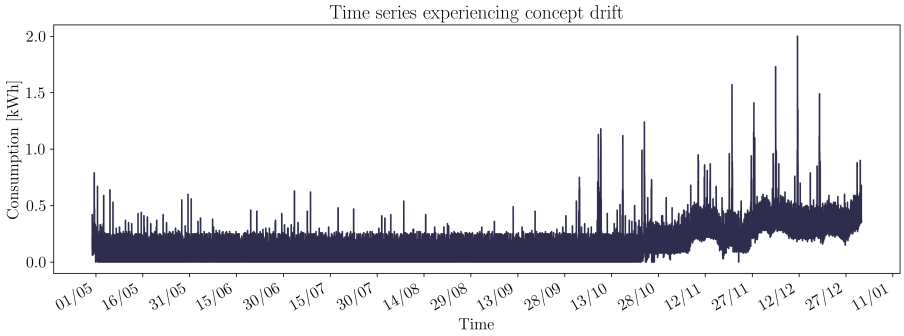


Figure 1.7. This time series illustrates the electricity usage of a particular household across a time period between May 2016 and December 2016. A noticeable change in the consumption pattern takes place towards the end of October. This is called concept drift. While the exact cause remains uncertain, it could be attributed to factors such as a change in occupancy, an update in the heating system, or possibly the purchase of a new electric vehicle. Such a concept drift presents a further obstacle in generating precise forecasts using historical data.

1.4 Structure objectives and significance of the thesis

The preceding sections have laid the foundation for understanding the multifaceted nature of ensuring access to affordable, reliable, sustainable, and modern energy, as emphasized by United Nations SDG7. We have explored the pivotal role played by the low-voltage grid to achieve this goal, as well as the innovative concept of a digital twin. Furthermore, we have underscored the importance and challenges of data in shaping the future of energy systems. We have skimmed on the history and techniques of time series forecasting with a deep dive into load applications, highlighting different end goals, gaps, and challenges. Building upon this knowledge, the subsequent chapters delve into specific aspects of load forecasting and energy optimization. Each chapter aspires to contribute to the broader understanding of how data-driven insights and forecasting techniques can enhance the efficiency and sustainability of energy systems.

The significance of this research extends beyond the realms of academia, resonating with industry and real-world applications, with the United Nations Sustainable Goal #7 detailed in Section 1.1 and the specific data challenges mentioned in Section 1.3.3. **Figure 1.8** positions the various research chapters along two dimensions: the forecasting horizon (short, medium, long-term) and the level of spatial aggregation (plug, household, city,...). It also highlights each chapter contributions to Sustainable Development Goal #7 and to data challenges, with each icon representing a distinct attribute as shown in the figure's legend. The background gradient provides insight into the volume of existing research within these fields, with darker shades indicating areas that have been extensively studied and lighter shades pointing to fields that are less explored.

Chapter 2: Long-term Household Consumption Forecasting

Chapter 2 of this thesis delves into long-term individual household consumption forecasting. By leveraging relative consumption patterns and ensemble learning, this chapter introduces a novel global algorithm that not only achieves state-of-the-art accuracy but also operates effectively even with incomplete data, requiring low computational power.

The development of this long-term forecasting method stemmed from a contest organized by the British electricity supplier company E.ON UK plc [132]. This applied research is not only useful, but also essential for electricity suppliers,

offering them a tool to enhance billing accuracy, production scheduling, and grid management. Importantly, this method addresses a notable gap in the field, as there is limited prior research on long-term household-level forecasting.

The contribution of this chapter to SDG7 is twofold. On one hand, it enhances the affordability of electricity access by avoiding overestimating household bills or underestimating them and then needing to compensate a large amount later on. On the other hand, it ensures more reliability since the forecasts allow grid management and production scheduling.

This global approach leverages historical household consumption data to make accurate predictions for individual households one year ahead, even with as little as one month of historical data. Additionally, the method does not require supplementary data, such as weather information or specific household attributes, when making predictions. These advantages effectively address the availability challenge discussed in Section 1.3.3. Furthermore, this approach enhances scalability by demonstrating up to twice the speed of other state-of-the-art methods.

> Further details on this topic are elaborated in the comprehensive paper: Botman, Lola; Soenen, Jonas; Theodorakos, Konstantinos; Yurtman, Aras; Bekker, Jessa; Vanthournout, Koen; Blockeel, Hendrik; De Moor, Bart; Lago, Jesus, "A Scalable Ensemble Approach to Forecast the Electricity Consumption of Households," in IEEE Transactions on Smart Grid, vol. 14, no. 1, pp. 757-768, Jan. 2023.

Chapter 3: Short-term Probabilistic Household Consumption Forecasting

Chapter 3 focuses on short-term probabilistic forecasting for individual household electricity consumption. This chapter proposes a global probabilistic forecasting approach, distinguishing itself from traditional local (point) forecasting models: the envisioned method generates accurate short-term quantile forecasts for one day ahead, offering insights into consumption uncertainties, and proving robust to change in consumption.

From an industrial point of view, the proposed global probabilistic forecasting method is set to be integrated into Fluvius digital twin. Fluvius, Flanders' distribution system operator, also underlined the practical applicability of this method for data imputation when smart meters fail to record a households consumption.

These short-term forecasts, allow anticipating potential congestion on the network, and trigger mitigation strategies. This enhances the reliability of access to energy to all and at all times, particularly during peak demand periods, therefore also aligning with the United Nations SDG7.

Similarly to the approach discussed in Chapter 2, this proposed global method effectively addresses several key challenges. Firstly, it demonstrates the ability to generate accurate one-day-ahead predictions using only seven days of historical data and no additional data source, thereby mitigating the availability challenge. Secondly, the utilization of quantiles as model outputs enables the method to handle stochasticity. Thirdly, its simplicity in training and global applicability contribute to scalability. Lastly, the potential for data imputation applications holds the promise of enhancing data quality.

Further details on this topic are elaborated in the comprehensive papers:

- Botman, Lola; Lago, Jesus; Agudelo, Oscar Mauricio; Vanthournout, Koen; De Moor, Bart, "A scalable method for probabilistic short-term forecasting of individual households consumption in low voltage grids," 2023 IEEE PES Grid Edge Technologies Conference & Exposition (Grid Edge), San Diego, CA, USA, 2023, pp. 1-5.
- Botman, Lola; Lago, Jesus; Becker, Thijs; Vanthournout, Koen; De Moor, Bart, "A global probabilistic approach for short-term forecasting of individual households' electricity consumption," **Submitted** in Applied Energy, November 2023.

Chapter 4: Smart Plug Mode Detection and Load Forecasting and Scheduling

In Chapter 4, we shift to non-residential buildings. Recent studies show that, in Europe as well as in the United States, the share of the energy consumption linked with buildings is 40% of the total energy consumption [42, 114]. This chapter explores the potential of smart plugs for building energy reduction. A comprehensive literature review of existing methods for plug mode detection is presented. Four approach types are identified, each based on heavy manual labeling, empirical decisions, or user input. Drawing upon a novel comprehensive dataset from the University of California, San Diego (UCSD), made openly accessible through this project, this research contributes to the literature by proposing the first automatic and unsupervised mode detection algorithm, i.e., accurately identifying whether a plug is active or non-active based on historical load values. A pipeline is proposed to schedule plugs on or off, based on one day

ahead power forecast along with their detected operating mode. The challenge of individual plug forecasting is tackled for the first time.

The proposed scheduling method has direct applications at UCSD, contributing to the buildings energy efficiency. In practice, over 600 smart plugs are already installed and recording load in over 15 buildings on the UCSD campus. Among all the installed smart plugs, around 200 are currently used for research and actual scheduling, although it should be extended to all plugs in the near future for the sake of energy savings.

By scheduling plug loads and saving energy, this chapter contributes to the sustainability aspect of the United Nations SDG7, fostering a more sustainable and efficient energy ecosystem for non-residential spaces. Additionally, this reduces the bill and contributes to more affordable electricity.

The uniqueness of the dataset utilized in this chapter marks a significant milestone, offering new insights into building management and demand flexibility. By openly sharing this unique dataset with the research community, we contribute to addressing the issue of data availability. Furthermore, similar to the applications discussed in previous challenges, the methods for mode detection, forecasting, and scheduling are designed to be scalable, and capable of handling the prediction needs of several thousand plugs required for monitoring an entire campus.

Further details on this topic are elaborated in the comprehensive paper: Botman, Lola; Lago, Jesus; Fu, Xiaohan; Chia, Keaton; Wolf, Jesse; Kleissl, Jan; De Moor, Bart, "Building consumption reduction with plug load forecasting and scheduling," **Submitted** in Applied Energy, November 2023.

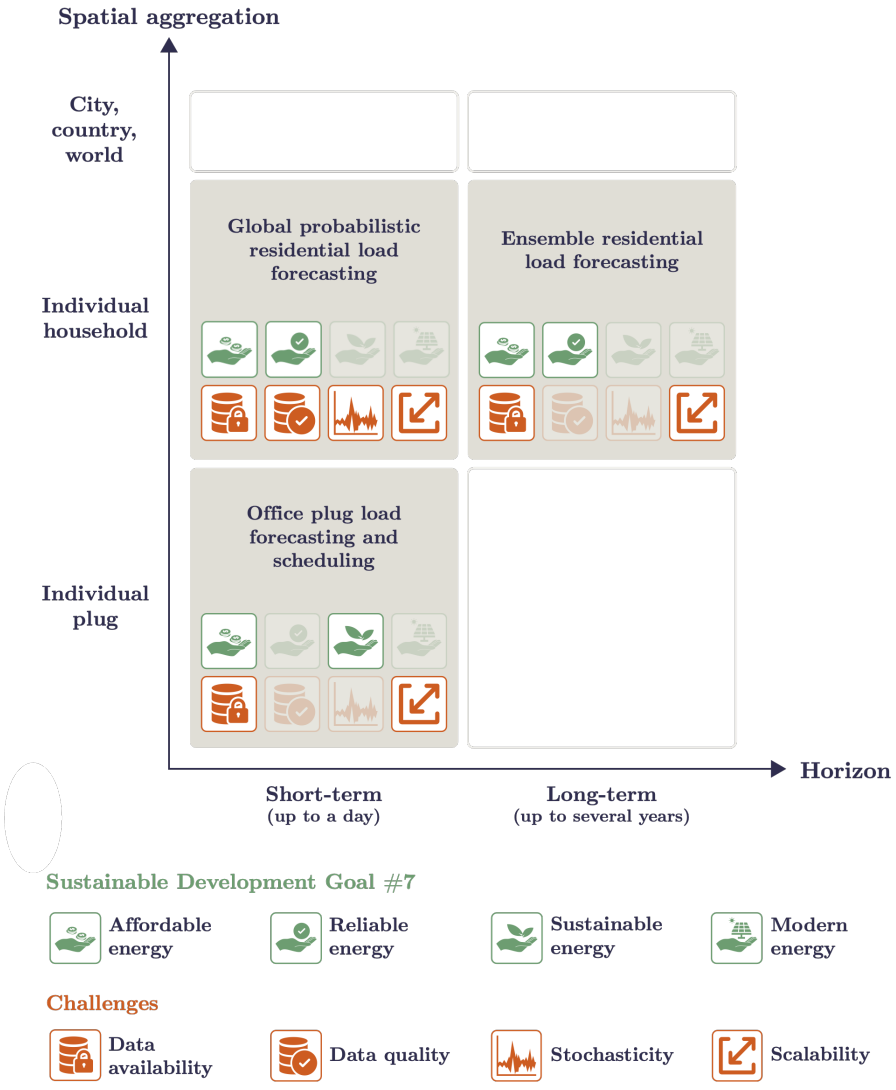


Figure 1.8. This figure organizes the research chapters of this thesis along two principal axes—forecasting horizon (ranging from short to long-term) and spatial aggregation level (from individual plugs to city-wide scales). The figure delineates the contributions of each chapter to Sustainable Development Goal #7, addressing affordable, reliable, sustainable, and modern energy, as well as specific data challenges including availability, quality, stochasticity, and scalability. Icons in color denote the focus of respective chapters, while grayed-out icons indicate areas not covered.

1.5 Acknowledgments

This research received funding from the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" program, from the KU Leuven Research Fund (projects C16/15/059, C3/19/053, C24/18/022, C3/20/117, C3I-21-00316) and from the European Research Council under the European Union's Horizon 2020 research and innovation programme (ERC Adv. Grant grant agreement N°885682).

The resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government.

The work of Chapter 2 has been carried out as part of a collaboration between VITO/EnergyVille and the research groups STADIUS and DTAI from KU Leuven. The authors greatly thank the other contestants of the IEEE-CIS competition: Wenlong Wu, Steffen Limmer and Kasun Bandara for providing useful information and kindly answering our questions. The authors would like to thank Prof. Isaac Triguero for setting up the competition and E.ON UK plc. for providing the data.

The work of Chapter 3 has been carried out as part of a collaboration between VITO (EnergyVille) and the research group STADIUS.

The work of Chapter 4 has been carried out as part of a research stay at the University of California, San Diego, in collaboration with the research group STADIUS from KU Leuven. This research was made possible thanks to an FWO travel grant for a long stay abroad (V408523N).

During the preparation of this work, the authors used ChatGPT in order to improve some language formulations. After using this tool, the authors carefully reviewed and edited the content as needed and take full responsibility for the content of the publication.

Chapter 2

A scalable ensemble approach to forecast the electricity consumption of households

This chapter is published and reprinted with permission from:

Botman, Lola; Soenen, Jonas; Theodorakos, Konstantinos; Yurtman, Aras; Bekker, Jessa; Vanthournout, Koen; Blockeel, Hendrik; De Moor, Bart; Lago, Jesus, "A Scalable Ensemble Approach to Forecast the Electricity Consumption of Households," in IEEE Transactions on Smart Grid, vol. 14, no. 1, pp. 757-768, Jan. 2023.

Abstract

Long-term individual household forecasting is useful in various applications, e.g., to determine customers' advance payments. However, the literature on this type of forecasting is limited; existing methods either focus on short-term predictions for individual households, or long-term prediction at an aggregated level (e.g. neighborhood). To fill this gap, we present a method that predicts the monthly consumption of individual households over the next year, given only a few months of consumption data during the current year. Utility companies can exploit this method to predict the consumption of any customer for the next year, even with incomplete data. The method consists of three steps: clustering the data using k-means, prediction using an ensemble of forecasts based on the historical median distribution among similar households, and smoothing the predictions to remove weather-dependent patterns. The method is highly accurate as it finished third in the IEEE-CIS competition (and ranks first when leveraging insights from another team), focused on forecasting long-term household consumption with incomplete data. It is also very scalable thanks to its low computational complexity and weak data requirements: the method only requires a few months of historical data and no household-specific or weather information.

2.1 Introduction

Electricity forecasting has been studied for many years and remains an important task. The roll-out of domestic smart meters enables the collection of consumption data of individual households. Electricity forecasting is one way in which these data can be used to create value. Accurate electricity consumption prediction is useful for a plethora of applications including, but not limited to, determining customers' advance payments, supporting day-to-day grid operations and strategic planning of energy grid extensions [137]. Different applications require different types of forecasts: day-to-day grid operations might require an hourly forecast for the next day, while strategic planning might require a monthly forecast for the next ten years [4]. In this paper, we propose a novel algorithm to forecast the monthly electricity consumption of individual households for the next year, given the monthly consumption during the available months of the current year.

The algorithm is one of the winning approaches at the IEEE-CIS competition [132]. The goal was to forecast the monthly consumption of multiple British households in 2018 using smart meter data (electrical consumption), weather data from 2017, and additional household-specific information such as the number of occupants, the number of bedrooms, etc. In the original competition, our approach ranked third out of 71 participating teams, reaching state-of-the-art accuracy, and we show that our method outperforms the approaches ranked first and second when disregarding smart meters without battery or when applying the same post-processing trick that the first approach applied to those faulty smart meters.

A key asset of the algorithm is that it can make predictions for a full year ahead even with incomplete data, e.g., even for households with one month of historical data. In addition, the proposed approach has the benefit of requiring low computational power, which grants the ability to scale up across millions of households. Finally, together with the other competition submissions, this model is one of the first to solve the challenge of long-term electricity forecasting at the individual household level in the literature.

2.1.1 Related work

The method proposed in this paper forecasts the monthly electricity consumption of individual households one year ahead, requiring only historical data of the previous year, at least the data from one month.

Electricity consumption forecasting is a wide subject of research. It has two dimensions: horizon and spatial granularity. The horizon is the forecast length in the future. It can be from few hours to multiple years. The spatial granularity is a geographical specification, i.e., consumption of an appliance or of an entire a country. Another relevant parameter is the sampling rate, which can vary from one second to one year. It is strongly linked with the horizon, e.g., if the electrical consumption of a city had to be predicted ten years ahead, an hourly forecast would not be accurate and thus not be considered.

Most of the efforts in the literature have been focused on long-term forecasting of coarse granularity, e.g., cities, geographical zones [17], regions [107, 116] or countries [27, 45, 88]. In long-term forecasting, due to the randomness and high volatility of the individual household electricity consumption, existing methods rely on averaging and aggregating these consumption patterns to a coarser spatial granularity.

Methods for individual household, i.e., fine granularity, consumption have been proposed [46, 77, 149, 150]. However, they are limited to short horizons. Typically, they forecast a few hours up to several days ahead and only for a small number of households. The latter is either due to the complexity of the algorithm used, leading to high computational requirements [150], or the lack of available input data. For example, in the Convolutional Neural Network – Long Short Term Memory (CNN-LSTM) model in [77], the authors used a configuration (i.e., three smart meters in specific household locations) that is nearly impossible to replicate in thousands of households. Another example of data scarcity is given in [46], where the case study is based on an unoccupied house and all activities are of an academic nature, which reduces stochasticity and simplifies the forecasting problem. The model proposed in [149] suffers the same drawback. It relies on typical daily load profiles depending on three parameters: the number of occupants, the time at which the first person gets up in the morning and the last person goes to sleep and the part of the day during which the house is unoccupied. The influence of the individual appliances is also considered (electric oven, TV, water heating, etc.). As with the previous works, this method can hardly generalize to multiple households due to the type of data required (e.g., disclosing the time of waking up or going to bed might conflict with privacy laws in several countries).

Short-term individual load forecasting might require different evaluation metrics [58] than long-term forecasting, e.g., to take in account the double penalty effect generated by the consumption peaks. Classical error metrics, such as RMSE, penalize twice for a peak that is correctly predicted by the algorithm in terms of amplitude and duration, but displaced in time: once when the peak actually happens and is not predicted and once again when the peak is predicted, but it does not actually happen. Long-term forecasting, based on monthly aggregated

data for example, are less prone to peaks, thanks to the averaging and are thus not subject to this double penalty effect.

Research on long-term and fine granularity forecasting mostly focuses on large energy consumption units, e.g., supermarkets [26] or public buildings [92, 108], where the patterns are less stochastic than households. Others investigate long-term residential consumption forecasting, per customer type, depending on average income and other demographics [105]. In the latter case, individual household consumption is not predicted, and the proposed method requires 30 years of historical consumption data in the training phase.

The lack of research on long-term household forecasting can further be seen in the extensive review of Zhao et al. [152], where out of the 92 extensively reviewed papers, only eight tackle residential load forecasting and just two consider fine granularity residential monthly forecasting. The first proposed method [103] tackles individual residential forecasting by disaggregating the household load into sixteen appliance categories, using detailed information such as the historical consumption, energy prices, weather information and demographics obtained through a questionnaire. Although potentially accurate, this method is not scalable due to required questionnaire data. The second one [102] proposes an approach in three steps in order to predict monthly consumption of six family houses based on only one month of historical data of the heating demand and the domestic energy demand as well as the indoor-outdoor temperature difference. The first step is to generate the consumption of a reference building using a simulation tool based on blueprint data of a similar house, the second is to generate additional data by scaling the measured data, the third is to train a neural network on the extended data and finally make predictions based on the actual indoor-outdoor temperature and the reference building. Several limitations are to be noted. Firstly, detailed blueprint data of a similar house is necessary, secondly the total household consumption has to be disaggregated into the energy used for space heating and the energy used for the electric appliances and hot water and finally, the data augmentation using scaling only makes sense because of the location of the households under investigation. Indeed, the houses are in Umeå, Sweden, where the temperature varies between $-30\,^{\circ}\mathrm{C}$ and $30\,^{\circ}\mathrm{C}$.

2.1.2 Motivation and contributions

This paper aims to fill a gap in the literature regarding long-term individual household forecasting, proposing a highly accurate and scalable method based on relative consumption patterns and ensemble learning. The proposed method is the first of its kind, as it

- can forecast *long-term* consumption of *individual* households,
- can operate on households with incomplete data,
- can handle missing time samples,
- does not have complex and exogenous data requirements such as weather data or household-specific attributes.

The method is highly accurate (as demonstrated by the award in the IEEE-CIS competition [132]), has low computational complexity and can be scaled to any number of households. As motivated by the competition, this type of forecasting is beneficial for customers, for accurate bill estimations. It is also interesting for producers, to schedule the right amounts of electricity production and procurement; for suppliers, to detect potential discrepancies between self-declared and real consumption; and for network operator in the grid management and strategic planning. Additionally, long-term individual household forecasting is one of the main research recommendation topics in the context of the energy transition [88].

The proposed method has to overcome a series of challenges that existing methods, due to the nature of the consumption granularity, do not have to overcome. For instance, household consumption patterns are more stochastic than aggregated data (on the city or national level) and also more stochastic than the consumption patterns for larger buildings such as supermarkets or offices. The algorithms that work for a supermarket, offices or a large building might not work for a household. Moreover, households change distributors every year in order to obtain the best rates. A commercial supplier might not have a full year of historical data to predict the next year. It is thus crucial to have a method which is reliable, works for households and can predict longer horizons than the available historical data.

2.1.3 Organization of the paper

The paper is organized as follows. Section 2.2 presents the key attributes considered to build the model, based on a set of realistic assumptions. The different steps of the method are detailed in Section 2.3, as well as the model parameters. Section 2.4.8 analyzes the impact of each step on the final performance metric. The case study proposed by the IEEE Data Port platform is explained extensively in Section 2.4, including information on the dataset, the benchmark models, the performance metrics and the results. Finally, the outcome and perspectives are discussed in Section 2.5.

2.2 Key attributes of the model

Before explaining the details of the model (see Section 2.3), it is important to outline the key attributes of the model and the assumptions made to derive these attributes. The proposed model forecasts monthly electrical consumption of individual households for a full calendar year using incomplete household data from the previous calendar year, where at least one month of measurements must be available. The model does not require household-specific information nor weather data. The methodology presented consists of four key attributes, which are based on assumptions, detailed in this section.

As the granularity of the predictions is monthly, most of the hourly and daily stochastic behavior of individual households averages out. In a monthly resolution, a lot of the stochastic behavior of individual households, e.g., what time one goes to sleep, plays little role. Instead, the factors that drive the consumption at that level are more deterministic, e.g., how large the household is. Since these deterministic factors are likely to be shared across many households, simple methods that use the mean or median consumption of households with similar consumption patterns, should be able to make highly accurate predictions. The first and second key attributes of the model stem from this assumption. The first attribute is that **the proposed model disregards hourly and daily patterns, and directly works with monthly data**. Therefore, even if hourly data from smart meters is available, we aggregate the data to the monthly level and disregards the stochastic patterns of lower resolutions. As a second key attribute, **we make predictions based on the median consumption of *similar* households**.

Although the absolute consumption of households differs significantly, relative monthly consumption patterns have less variance and are more similar. Therefore, the perceptual consumption of each month, with respect to the yearly consumption, is shared among many households. This can be explained by the fact that, independently of the household type or size, similar human behavior is shared across the households. Based on this observation, a third key attribute of the model is to **work with relative consumption patterns instead of absolute ones**. In other words, we normalize electricity consumption data.

The model assumes that, since relative consumption patterns are considered, additional household-specific information such as household size and weather data play little to no role. In particular, households with different types of appliances and a different number of occupants might have the same relative consumption profile, while households with the exact same number of occupants and/or appliances might have different relative profiles. Similarly, weather data are only available for the current year. Using this weather data to predict the

next year can lead to over-fitting the model to the current year. As an example, let us consider a household where November in the current year is colder than December: that household will likely have a larger consumption in November than December due to the weather. Yet, that specific situation will likely not generalize to the next year as December is, in general, expected to be colder. Moreover, weather forecasts are usually not available one year in advance and can thus not be used for this horizon. For these reasons, the fourth model attribute is to **disregard exogenous features**.

2.3 Methodology: Ratio approach

In order to forecast the household monthly consumption, different steps are applied sequentially. These can be divided in three main parts. Firstly, the training pipeline, Fig. 2.1a, consists of the preprocessing, data augmentation, normalization and clustering. Secondly, the inference framework, Fig. 2.1b, consists of the prediction step. Finally, the ensemble and post-processing steps are illustrated in Fig. 2.1c. All the modules are detailed in Subsections 2.3.1 to 2.4.5. In these sections, we provide a clear example for interpretability, while a general notation is used in the figures.

- *Preprocessing* A real life dataset usually contains missing data, due to defaults in the smart metering device, for example. For each household/smart meter, full missing days are imputed using linear interpolation on a daily level. Subsequently, the data are aggregated to the monthly level using summation.
- *Data Augmentation & Normalization* As discussed in Section 2.2, it is better to consider relative as opposed to absolute consumption. This conversion happens in the normalization step. However, a consumption profile depends heavily on the month in which the customer signs up with a supplier, and this is reflected in the normalization. Therefore, our proposed method explicitly considers the sign-up month. This is explained in more detail in Section 2.3.1.
- *Clustering* The relative consumption profiles are clustered together in order to group similar profiles. Different clusterings are executed per sign-up month. Profiles that signed up later are retroactively added to clusterings for earlier sign-up months. This is explained in more detail in Section 2.3.2.
- *Prediction* Each cluster can be used to make a prediction for a certain household. This operation is explained in Section 2.3.3. Since

multiple clusterings exist per household, several predictions can be made. Section 2.3.4 explains which clusters to consider and how to combine their predictions in an ensemble approach.

- *Post-processing* As a final step, the predictions are smoothed using a standard moving average technique in order to reduce the effect of weather patterns in the current year, as explained in Section 2.2. By using a moving average, the effects of weather are mitigated, as the final prediction is the average between nearby months. This step has a parameter w describing the window length for the moving average.

The entire process makes use of very few hyperparameters. Only three parameters require tuning (the minimum cluster size $n_{\text{min,cluster}}$, the number of nearest neighbors n_{similar}, and the window length w for smoothing), which are explained in Section 2.3.2, 2.3.3, 2.3.5 respectively. As detailed in Section 2.4.9, their value, although important, is not very critical as the method is rather robust to the selection of these three hyperparameters.

2.3.1 Data Augmentation & Normalization: relative consumption profiles

The proposed method is based on using relative consumption profiles, where a relative profile is the absolute profile divided by the yearly consumption. For instance, the relative consumption of a given household in April is the consumption of April divided by the annual consumption.

The vector of monthly consumption of a household x_j in a year is defined as

$$\mathbf{L}^{x_j} = [l_1^{x_j}, \ldots, l_{12}^{x_j}] \in \mathbb{R}^{12} \tag{2.1}$$

where the subscripts of the vector's elements denote the month, e.g., $l_1^{x_j}$ is the monthly consumption value for January. For a household x_j that signed up in month s, the first $s-1$ elements of the monthly consumption vector $\mathbf{L}^{x_j}$ are missing:

$$\mathbf{L}^{x_j} = [\underbrace{\text{NaN}, \ldots, \text{NaN}}_{(s-1)}, l_s^{x_j}, \ldots, l_{12}^{x_j}] \in \mathbb{R}^{12}. \tag{2.2}$$

Therefore, we introduce the monthly consumption vector

$$\mathbf{L}_s^{x_j} = \left[l_s^{x_j}, \ldots, l_{12}^{x_j}\right] \in \mathbb{R}^{12-s+1} \tag{2.3}$$

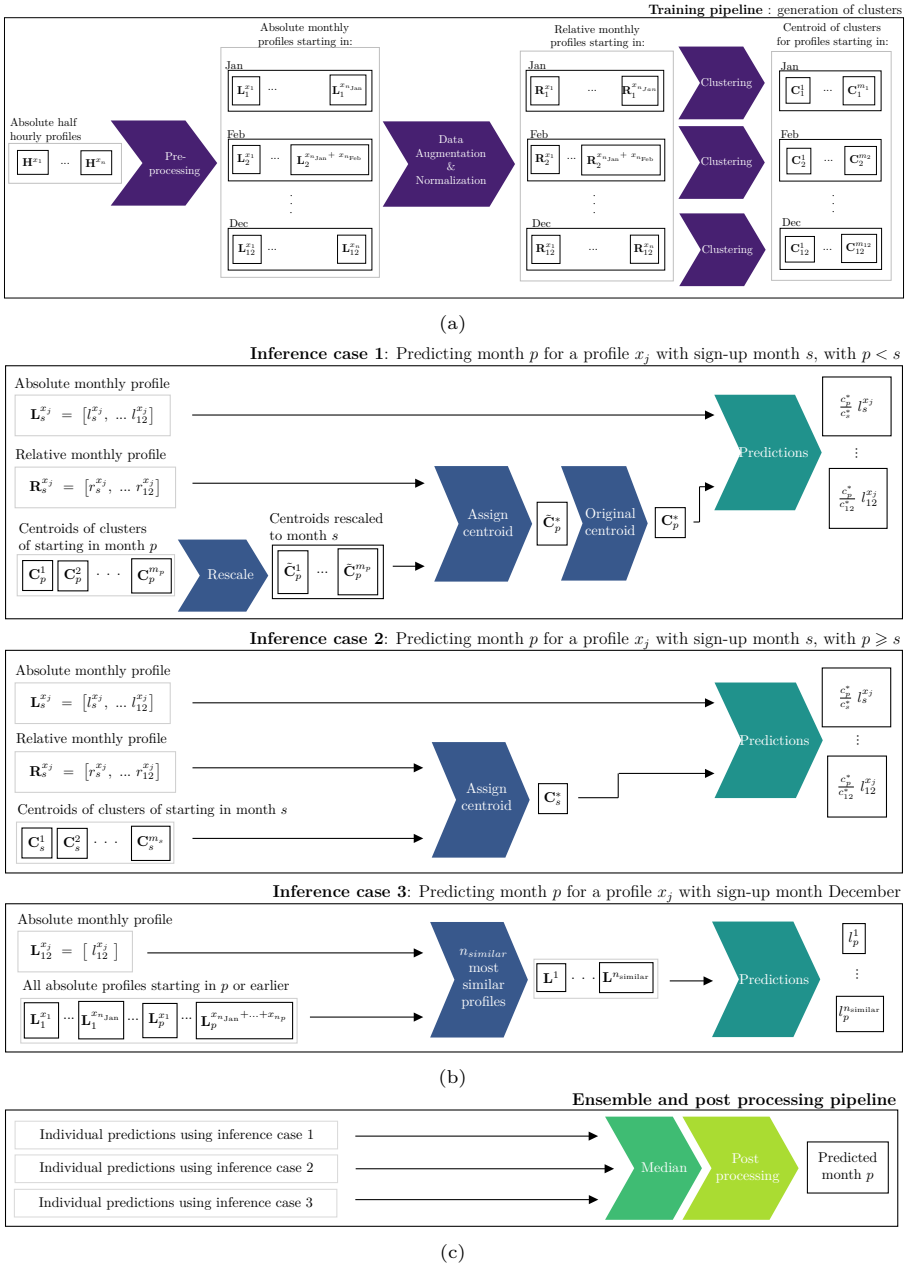


Figure 2.1. Flowchart of the method proposed in this paper. The input and output data are in the rectangles, and the steps executed are in the chevron arrows. (a) The training pipeline is applied to all the profiles and aims to build the clusters based on the relative profiles. The data augmentation step is highlighted using the indices of the profiles. There are n_{Jan} profiles which start in January, n_{Feb} in February, and so on. The total number of profiles is $n = n_{\text{Jan}} + n_{\text{Feb}} + \cdots + n_{\text{Dec}}$. (b) General framework for inferring the consumption of a month p for a profile with sign-up month s. (c) Ensemble and post-processing pipeline: the individual predictions from the inference cases, are combined using the median and smoothed in the post-processing step.

of household x_j starting from month s. This notation will also be used to discard the first few months of a profile, even though consumption data of the discarded months might be available.

To make the consumption profiles between households that signed up in the same month comparable, the consumption profiles $\mathbf{L}_s^{x_j}$ are normalized into relative consumption profiles $\mathbf{R}_s^{x_j}$ as

$$\begin{aligned} \mathbf{R}_s^{x_j} &= \left[r_s^{x_j}, \ldots, r_{12}^{x_j}\right] \in \mathbb{R}^{12-s+1}. \\ &= \frac{1}{\sum_{i=s}^{12} l_i^{x_j}} [l_s^{x_j}, \ldots, l_{12}^{x_j}]. \end{aligned} \tag{2.4}$$

After this transformation, the total relative consumption of each household sums up to one.

The relative consumption of households that signed up in different months are still not directly comparable. For example, a household x_j that signed up in November has a relative consumption in November that is around 50%, but if the same household has signed up in January the relative consumption during November would be closer to 8% ($\approx 1/12$).

To solve this issue, two steps are taken. In the clustering step, to find similar consumption patterns, a different clustering is made for each sign-up month. In the prediction step, only ratios between two months are considered, as these ratios are independent of the sign-up month (see section 2.3.3 for details).

Furthermore, a data augmentation technique is applied to maximally exploit the available data. When making a clustering for relative profiles that signed up in month s, we consider not only the data of households that signed up in month s, but also the data of the households that signed up earlier by discarding any consumption before month s. For example, a profile with sign-up month March, is also considered as if it signed up in April, i.e., by disregarding March. Similarly, it is also appended to the profiles signed up in May, i.e., by disregarding March and April; and so on. The relative profiles of the households that signed up earlier are re-normalized such that each relative profile sums up to one. In the prediction step, a similar procedure is applied, which is explained in section 2.3.3.

2.3.2 Clustering: grouping similar profiles

As households from different sign-up months have different relative values, clusters are built for each possible sign-up month. As an example, Fig. 2.2

illustrates the four clusters obtained when considering the households that signed up in January, using the k-means algorithm with Euclidean distance metric. Similarly, Fig. 2.3 illustrates the five clusters obtained when considering the households that signed up in June, including those that signed up before and from which the measurements of the months before June have been ignored. In both figures, for each cluster, the centroid (mean) of the cluster is plotted together with a 90% confidence interval. The k-means clustering is guaranteed to converge to an optimum, although this might be a local optimum [115].

In order to introduce the notation for the prediction step (Section 2.3.3), let $x_1, x_2, \ldots, x_N$ be the N households that signed up in month s or earlier. The set of these households is denoted by $\mathcal{X}_s$. For each of these households, independently of the data available, the clustering algorithm first defines a relative profile $\mathbf{R}_s^{x_j} \in \mathbb{R}^{12-s+1}$ starting in s, as in (2.4). Let $\mathcal{S}_s$ be the set of the N relative profiles starting in s (one per household that signed up in month s or earlier):

$$\mathcal{S}_s = \{\mathbf{R}_s^{x_1}, \ldots, \mathbf{R}_s^{x_N}\}. \tag{2.5}$$

The clustering step uses the K-mean clustering algorithm [91] with Euclidean distance to cluster the relative consumption profiles $\mathcal{S}_s$ into a set of m_s clusters with centroids

$$\mathcal{C}_s = \{\mathbf{C}_{\mathrm{s}}^1, \ldots, \mathbf{C}_s^{m_s}\}, \tag{2.6}$$

where the centroid of the k'th cluster $\mathbf{C}_{\mathrm{s}}^k$ is the mean of all profiles in that cluster.

$$\mathbf{C}_s^k = [c_{s,s}^k, \ldots, c_{s,12}^k]. \tag{2.7}$$

The double subscripts v, w in $c_{v,w}$ represent the sign-up month of the centroid and the given month within the centroid, respectively. For example, $c_{3,6}$ represents the month June within the centroid C_3, which has sign-up month March.

As k-means requires the number of clusters as an input, the elbow method [129] is employed to estimate the number of clusters. Moreover, to avoid unrepresentative clusters that do not generalize, we discard clusters with fewer than $n_{\min,\mathrm{cluster}}$ profiles, as they are likely to contain outliers and may not be representative enough.

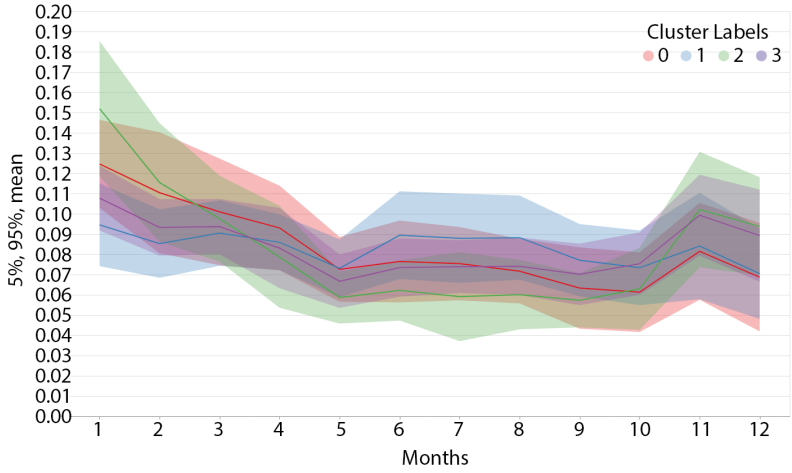


Figure 2.2. Example of clusters of relative consumption profiles starting in January for the IEEE-CIS competition data. The number of clusters obtained is four, determined using the elbow method. The centroid (mean) of the cluster is plotted together with a 90% confidence interval.

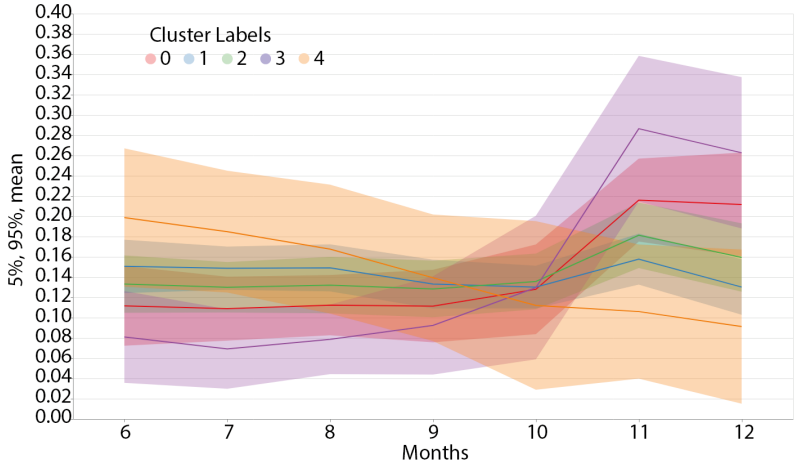


Figure 2.3. Example of clusters of relative consumption profiles starting in June for the IEEE-CIS competition data. The number of clusters obtained is five, determined using the elbow method. The centroid (mean) of the cluster is plotted together with a 90% confidence interval. It is clear, when comparing Fig. 2.2 and Fig. 2.3 that the relative values are different depending on the sign-up month.

2.3.3 Prediction: the power of the median via ensemble learning

The proposed model forecasts monthly electrical consumption of individual households for a full calendar year using (incomplete) household data from the previous calendar year. Here, "predicted month" refers to the monthly consumption value that we aim to predict in the next year, while "sign-up month" refers to the month in which the household signed up with the provider, meaning in the previous year.

Once we have clustered the relative profiles, we build the predictions. To do so, for each household to be predicted, we generate multiple predictions based on the clustered profiles and compute the final prediction as the median of these predictions. We combine predictions to exploit ensemble learning [154], i.e., combining predictions (or models) to reduce the bias or variance errors of the individual predictions (or models). In this context, we use the median instead of the mean to combine the predictions because it is less sensitive to outliers.

The proposed method considers a slightly different ensemble learning step depending on whether the prediction is made for an earlier or later month than the month in which the household signed up. In particular, the method distinguishes three cases, as illustrated in Fig. 2.1b: (1) predicting an earlier month than the sign-up month, i.e., a month for which there is no monthly data of the previous year available, (2) predicting a later (or equal) month to the sign-up month, i.e., a month for which there is monthly data of the previous year available, and (3) predicting the consumption for households that signed up in December.

To explain the prediction step, it is easier to consider an example. Let us consider again a household x_j that signed up in June. As explained in the previous sections, for household x_j, the relative profile is from June onward:

$$\mathbf{R}_6^{x_j} = [\mathbf{r}_6^{x_j}, \ldots, \mathbf{r}_{12}^{x_j}]. \tag{2.8}$$

Predicting an earlier month than the sign-up month

The prediction process for this case is represented in Fig. 2.1b, case 1. Let us assume that the month of March needs to be predicted for household x_j. For that, consider the centroids $\mathcal{C}_3 = \{\mathbf{C}_3^1, \ldots, \mathbf{C}_3^{m_3}\}$ obtained from the clustering with March as a sign-up month and denote the centroid elements as:

$$\mathbf{C}_3^k = [c_{3,3}^k, \ldots, c_{3,12}^k] \in \mathbb{R}^{10}. \tag{2.9}$$

Second, based on these m_3 centroids, a set of re-scaled centroids are generated $\{\tilde{\mathbf{C}}_3^1, \ldots, \tilde{\mathbf{C}}_3^{m_3}\}$ which match the scaling and size of the relative profile $\mathbf{R}_6^{x_j}$ of household x_j (with sign-up month June):

$$\begin{aligned}\tilde{\mathbf{C}}_3^k &= \left[\tilde{c}_{3,6}^k, \ldots, \tilde{c}_{3,12}^k\right] \in \mathbb{R}^7 \\ &= \frac{1}{\sum_{i=6}^{12} c_{3,i}^k} \left[c_{3,6}^k, \ldots, c_{3,12}^k\right].\end{aligned} \tag{2.10}$$

The part of the centroid before the sign-up month is disregarded and the centroid is re-scaled so that the new centroids sum up to one. Third, we determine which centroid $\tilde{C}_3^k$ lies closest to the relative profile $\mathbf{R}_6^{x_j}$, using Euclidean distance. We thus have the optimal centroid

$$\tilde{\mathbf{C}}_3^\star = \underset{\tilde{\mathbf{C}}_3^k \in \{\tilde{\mathbf{C}}_3^1, \ldots, \tilde{\mathbf{C}}_3^{m_3}\}}{\arg\min} \|\tilde{\mathbf{C}}_3^k - \mathbf{R}_6^{x_j}\|_2^2. \tag{2.11}$$

Then, the original full centroid $\mathbf{C}_3^\star \in \mathbb{R}^{10}$ associated with the optimal rescaled centroid $\tilde{\mathbf{C}}_3^\star \in \mathbb{R}^7$ is used to predict March's consumption for household x_j. Seven predictions are computed, based on the seven relative centroid values $c_{3,3}^\star, \cdots, c_{3,12}^\star$ of $\mathbf{C}_3^\star$ and the seven absolute measurements $l_6^{x_j}, l_7,^{x_j}, \cdots, l_{12}^{x_j}$ of household x_j. The predicted consumption of the month March $\hat{l}_3^{x_j}$ is calculated as the median of these individual predictions:

$$\hat{l}_3^{x_j} = \text{median}\left(\frac{c_{3,3}^\star}{c_{3,6}^\star} l_6^{x_j}, \ldots, \frac{c_{3,3}^\star}{c_{3,12}^\star} l_{12}^{x_j}\right). \tag{2.12}$$

In other words, to predict the consumption of March for household x_j, seven predictions are built: one for each available monthly consumption. In particular, for each available monthly consumption $l_q^{x_j}$, a prediction is built by multiplying the ratio between the representative relative consumption in March and the representative relative consumption in month q, i.e., $\frac{c_{3,3}^\star}{c_{3,q}^\star}$, by the monthly absolute consumption $l_q^{x_j}$. Then, the final prediction is built as the median of the individual predictions.

Predicting a later month than the sign-up month

This prediction process is represented in Fig. 2.1b, case 2. When predicting a later (or equal) month than the sign-up month, the same principle applies. However, instead of working with the clusters starting in the month that we want to predict p, the clusters associated with the sign-up month s are used.

As an example, let us consider again a household x_j that signed up in June. However, let us consider the case of predicting July rather than March. Now,

instead of using the m_7 centroids $\{\mathbf{C}_7^1, \ldots, \mathbf{C}_7^{m_7}\}$ associated with the predicted month (July), the m_6 centroids $\{\mathbf{C}_6^1, \ldots, \mathbf{C}_6^{m_6}\}$ associated with the sign-up month (June) are used.

The reasons for considering the centroids of the sign-up month instead of the predicted month are twofold. First, as the scale and length of each centroid $\mathbf{C}_6^k \in \mathbb{R}^7$ is the same as the relative profile $\mathbf{R}_6^{x_j} \in \mathbb{R}^7$, it is not necessary to re-scale the centroids. Second, by working with the clusters associated with the sign-up month, the number of predictions used to compute the median is maximized. For example, while for June as a sign-up month there are seven predictions, whereas for July there are only six. Similarly to the previous case, the closest centroid (in terms of Euclidean distance) to the relative profile $\mathbf{R}_6^{x_j}$ is determined and denoted by $\mathbf{C}_6^\star$, as in (2.11). Similar to (2.12) in the previous case, the prediction $\hat{l}_7^{x_j}$ is built by computing the medians of the ratios between the representative relative consumption of July and the representative relative monthly consumption of the available months multiplied by the absolute monthly consumption:

$$\hat{l}_7^{x_j} = \text{median}\left(\frac{c_{6,7}^\star}{c_{6,6}^\star} l_6^{x_j}, \frac{c_{6,7}^\star}{c_{6,7}^\star} l_7^{x_j}, \ldots, \frac{c_{6,7}^\star}{c_{6,12}^\star} l_{12}^{x_j}\right). \tag{2.13}$$

Predicting households that signed up in December

As the relative consumption for the households that signed up in December is always 1, the described method cannot be used to predict these households. Therefore, for these households, a different approach is considered, represented on Fig. 2.1b, case 3.

First, for each household x_j and predicted month p, the n_{similar} most similar households are computed in terms of absolute consumption in December that have data available for month p. That is, to predict March we consider all meters with data for March and December, in other words, the set of households that signed up in month p or earlier.

Then, the n_{similar} meters that are most similar to household x_j are found. Second, defining this set of similar households by $\mathcal{S}^{p,j} = \{x_1^{p,j}, \ldots, x_{n_{\text{similar}}}^{p,j}\}$, the prediction $\hat{l}_p^{x_j}$ is built for household x_j and month p as the median of the historical values of the similar households:

$$\hat{l}_p^{x_j} = \text{median}\left(l_p^{x_1^{p,j}}, \ldots, l_p^{x_{n_{\text{similar}}}^{p,j}}\right). \tag{2.14}$$

To find the n_{similar} most similar households, the k-nearest neighbors algorithm [33] is applied. In this context, although the parameter n_{similar} should be

optimized, empirical observation shows that it makes no difference for values $n_{\text{similar}} > 10$. We have tested values ranging from 10 to 100 during the IEEE-CIS competition and empirically observed that it does not have a significant impact. The value used in the final submission of the method is $n_{\text{similar}} = 50$.

2.3.4 Ensemble learning

As explained in the previous sections, the consumption of each (household, month) pair is predicted by building multiple predictions for each pair and then computing the median. The reason for doing so is ensemble learning [154], i.e., combining predictions (or models) to reduce the bias or variance errors of the individual predictions (or models).

In general, when combining multiple predictions, the larger the ensemble, the lower the variance or bias error and the better the prediction becomes. To that end, as an additional step during the prediction phase, a data augmentation technique is performed to increase the number of predictions for each (household, month) pair.

As before, let us explain this step with an example. Let us consider the household x_j that signs up in June and the process of predicting July. As explained in Section 2.3.3, this prediction is done by computing the median of seven individual predictions (cfr. (2.13)).

To improve this prediction, additional individual predictions are generated by assuming that the household did not sign up in June but signed up in July. To do so, the exact same procedure, described in Section 2.3.3, is repeated but assuming that the household has no data for June. This leads to a set of new six new predictions, similarly to (2.13).

The same procedure is repeated by simulating the cases that the household signed up in August, September, ..., up to November. The predicted month is now earlier than the sign-up month, as in Section 2.3.3 (cfr. (2.12)).

Furthermore, a prediction for July can also be computed by considering that the household signed up in December, by applying the procedure described in Section 2.3.3, similarly to (2.14).

So, in total, $7+6+5+4+3+2+1=28$ individual predictions are generated to compute the prediction

$$\hat{l}_7^{x_j} = \text{median}\left(\begin{array}{c} \frac{c^\star_{6,7}}{c^\star_{6,6}} \cdot l_6^{x_j}, \ldots, \frac{c^\star_{6,7}}{c^\star_{6,12}} \cdot l_{12}^{x_j}\Big|_{s=6} \\ \\ \frac{c^\star_{7,7}}{c^\star_{7,7}} \cdot l_7^{x_j}, \ldots, \frac{c^\star_{7,7}}{c^\star_{7,12}} \cdot l_{12}^{x_j}\Big|_{s=7} \\ \\ \frac{c^\star_{7,7}}{c^\star_{7,8}} \cdot l_8^{x_j}, \ldots, \frac{c^\star_{7,7}}{c^\star_{7,12}} \cdot l_{12}^{x_j}\Big|_{s=8} \\ \vdots \\ \frac{c^\star_{7,7}}{c^\star_{7,11}} \cdot l_{11}^{x_j}, \frac{c^\star_{7,7}}{c^\star_{7,12}} \cdot l_{12}^{x_j}\Big|_{s=11} \\ \\ \text{median}\left(l_7^{x_1^{7,j}}, \ldots, l_7^{x_{n_{\text{similar}}}^{7,j}}\right)\Big|_{s=12} \end{array}\right). \tag{2.15}$$

Each row in the parentheses considers a different sign-up month from June ($s=6$) up to December ($s=12$). This is also illustrated in Fig. 2.1c.

In general, for a household x_j that signed up in month s, this data augmentation step creates extra predictions so that we have a total of $\sum_{i=1}^{13-s} i$ individual predictions that are used to compute the median.

2.3.5 Post-processing: smoothing the predictions

As a final step, the predictions are smoothed using a standard moving average technique. The motivation behind doing so is to reduce the effect of weather patterns in the year of the given measurements (see Section 2.2 for details). By using a moving average, the weather effects are mitigated as the final prediction is the average between nearby months. This step has a parameter w describing the window length for the moving average. We tested three values as part of the IEEE-CIS competition and observed that, although the window size plays little role, five months is marginally better.

2.4 Case Study

2.4.1 Problem statement IEEE-CIS Competition

The problem tackled in this paper is the monthly electrical consumption forecasting of thousands of households one year ahead using the historical

consumption of the previous year. The proposed method was one of the award-winning methods in the IEEE-CIS competition [132]. For the sake of open-access and reproducibility, we use in this paper the same case study, as it allows any other researcher to verify the results and re-use the data. In the competition, the goal was to forecast the monthly consumption of 3248 households in 2018 using (i) half-hourly smart meter data from 2017 for the same households, with a different data availability for each household, ranging from one to twelve months representing different customer sign-up months, (ii) the weather data from 2017 for the households location at a daily resolution and (iii) additional household-specific information such as the number of occupants, the number of bedrooms, etc.

2.4.2 Dataset

The dataset is provided by E.ON UK plc., in the context of the IEEE-CIS competition on energy prediction from smart meter data [132]. It consists of half-hourly sampled time series describing the electrical consumption of 3248 households during the year 2017. It is a real world dataset containing all the challenges of a real application.

First, the profiles have a **different range of available historical data**, acknowledging that customers might have joined the measurement campaign at different times during the year, as is illustrated in Fig. 2.4. It is a two-dimensional representation of the electrical consumption recorded by the 3248 smart meters. It can be seen that the first 270 profiles (about a 12th of the profiles) have twelve months of historical data available (from January until December), the next 270 profiles have eleven months of historical data (from February until December) and so on until the last part of the profiles which have only one month of historical data available (December).

Secondly, **there are missing days within the available months**, which are the faint white lines in Fig. 2.4.

Thirdly, additional household-specific information were collected through surveys, such as the type of building, the number of rooms, the number of occupants, and so on. However, **these data are very sparse**, for example there are 48% of missing data in the type of building, 42% in the number of bedrooms, and always more than 97% for all the other additional information.

Finally, the competition also provides time series weather data, e.g., average, maximum and minimum daily temperature, associated with each household, in 2017. However, **the weather data of 2018 is not made available**, as well as the household consumption data of 2018. There is thus no validation

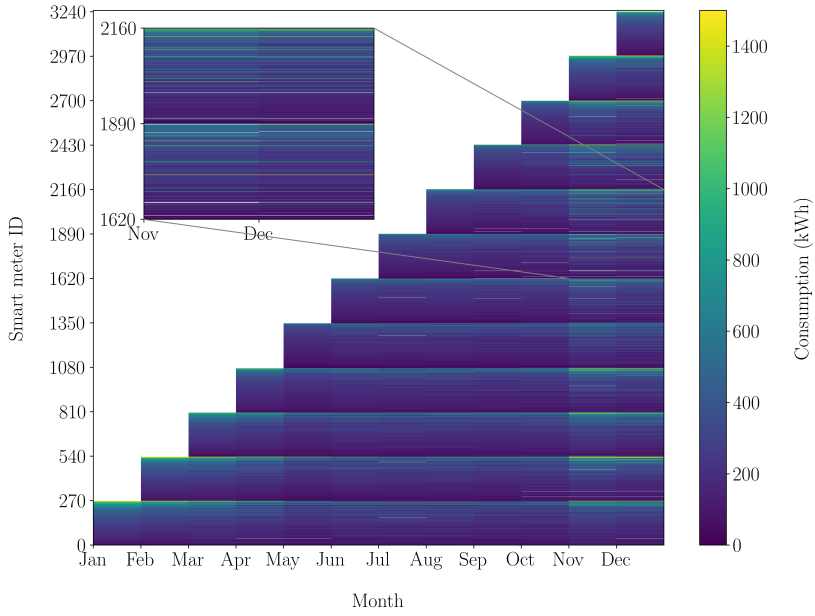


Figure 2.4. Heat map of the monthly electrical consumption per smart household (in kWh). The y-axis represents the individual profiles and the x-axis the month. The color gives an indication of the monthly values. In the colder months (November & December) the consumption is higher. The faint white lines are missing data. It is clear from the staircase structure that only the first 270 profiles have twelve months of data available. The next 270 profiles have eleven months of data available, and so on up to the last 270 profiles, which have only the month December.

dataset. It is subsequently not possible to assess the performance between the years. The performance could be checked on a leaderboard via the organizers. Further details about the data are provided on the competition website [132].

The objective of the competition was to predict the monthly consumption (in kWh) of the 3248 households for the year 2018 as well as the aggregated yearly consumption for the same year. As explained by the organizers, the yearly total consumption, computed as the sum of the monthly forecasts, is of interest for billing purposes while the monthly values are used by the energy trading teams to buy the right amount of electricity on the energy market. As the data are real and publicly available, we use it here for our case study. As explained in the next section, we also consider the metrics proposed by the organizers of the competition.

2.4.3 Performance Metrics

We use the metrics from the competition to measure the performance of our method [132]. It is the weighted average error of the yearly prediction error and the monthly prediction error. In particular, the yearly relative absolute error (year_{rAE}) is measured as

$$\text{year}_{\text{rAE}} = \frac{\frac{1}{n}\sum_{j=1}^{n} |\hat{l}^{x_j} - l^{x_j}|}{\frac{1}{n}\sum_{j=1}^{n} |l^{x_j} - \bar{l}|} \tag{2.16}$$

where n is the total number of households, $\hat{l}^{x_j} = \sum_{i=1}^{12} \hat{l}_i^{x_j}$ is the predicted total yearly consumption of household x_j, l^{x_j} is the true total yearly consumption of household x_j and $\bar{l} = \frac{1}{n}\sum_{j=1}^{n} |l^{x_j}|$. Then the relative monthly error ($\text{month}_{\text{rAE}}$) is considered

$$\text{month}_{\text{rAE}} = \frac{1}{n}\sum_{j=1}^{n} \frac{\frac{1}{12}\sum_{i=1}^{12} |\hat{l}_i^{x_j} - l_i^{x_j}|}{\frac{1}{12}\sum_{i=1}^{12} |l_i^{x_j} - \bar{l}^{x_j}|} \tag{2.17}$$

where n is the total number of households, $\hat{\mathbf{L}}^{x_j} = [\hat{l}_1^{x_j}, \ldots, \hat{l}_{12}^{x_j}]$ is the predicted monthly consumption of household x_j , $\mathbf{L}^{x_j} = [l_1^{x_j}, \ldots, l_{12}^{x_j}]$ is the true monthly consumption of household x_j and $\bar{l}^{x_j} = \frac{1}{12}\sum_{i=1}^{12} |l_i^{x_j}|$.

Finally, both metrics are considered equally important and aggregated as

$$\text{total}_{\text{rAE}} = \frac{1}{2}\text{month}_{\text{rAE}} + \frac{1}{2}\text{year}_{\text{rAE}}. \tag{2.18}$$

2.4.4 Benchmark Models

As part of the competition, we tested several different prediction models. In this paper, we compare our model, with a baseline model that we designed and two well-known models that were submitted to the competition by another contestant: a linear regression (LR) model and an Autoregressive Integrated Moving Average (ARIMA) model [83]. These have been submitted by Dr. Kasun Bandara [14, 15]. In addition, we outline the performance of the top three approaches (one of which is the one proposed in this paper).

The naive benchmark we proposed, computes the monthly average over the available months of 2017 for each individual household and uses this monthly average as a prediction for each of the months of the year 2018.

Wenlong Wu (first place) [132, 147] proposed a Machine Learning pipeline consisting of data preprocessing, feature engineering, algorithm modeling, post-processing, and ensemble fusion. Fuzzy C-Means is applied, extracting twelve

clusters of profiles. The model is based on a Light Gradient Boosting Machine (Light GBM) which makes use of bagging and boosting techniques. First, this model is applied on all the households individually, it learns the overall trend and makes day-level energy prediction. Secondly, one model is trained per cluster. Thirdly, a prediction is computed using the mean of November and December values. These three predictions are fused using a weighted average. Additionally, Wenlong Wu found 32 profiles with zero consumption values for November and December, which might be due to drained batteries during the acquisition of training data. Assuming that this problem is present in the collection of the ground-truth (i.e., test) data as well, he manually set their predictions to zero. We do not find this "battery trick" operation fair in terms of evaluation, because it exploits a specific problem in data collection. Finally, a scaling step is applied, down-scaling the summer months and up-scaling the winter months as he noticed this improved the performance. The weather data and additional household-specific information were not used. The six features used by Wenlong Wu are categorical household ID, mean meter reading of the household, and time-related features like day of week, day of month and month encoded cyclically using sine and cosine functions, such that January and December are defined as similar.

Steffen Limmer's approach (second place) [132] is based on k nearest neighbors, the following sequence of steps is considered: data preprocessing, outlier removal (by computing the base distributions and filtering out abnormal distributions using isolation forest), an ensemble strategy to predict the monthly consumption and finally scaling. In the prediction step, the distribution of a meter is predicted as the average over the distributions of k nearest base meters (without the outliers) and then the monthly consumption is computed based on the distribution prediction and the December consumption.

2.4.5 The proposed approach with the battery trick

In addition to our proposed scalable ensemble approach that ranked third in the competition, we also consider its modified version that uses the *battery trick*, which is applied in the method that ranked first and explained in Section 2.4.4. This modification would enable our proposed approach to outperform the first ranked method in the competition, as shown in Section 2.4.7.

2.4.6 Run times

The proposed method takes around 45 minutes to run on a machine with 9th Generation "Coffee Lake" 2.6 GHz 6-Core Intel Core i7 mobile processor (I7-

9750H), macOS operating system, Python version 3.8.3: 20 seconds to generate the clusters, 19 minutes to determine the closest cluster for all inference cases of all profiles, 25 minutes to build the predictions and 40 seconds for ensemble and smoothing. Besides the training and ensemble/post-processing pipelines, the method is parallelized on 4 threads. It could easily be parallelized with more threads to speed up the process even more.

The Light GBM model, by Wenlong Wu, takes 1h45 to run on the same machine[1]. It should also be noted that this method requires one light GBM model per smart meter individually, which reduces its applicability and scalability.

The run time of the KNN-Isolation forest ensemble method, by Steffen Limmer, could not be measured as the code was not made available.

2.4.7 Results & discussion

The results are presented in Table 2.1. The ratio based approach described in this paper is compared with several standard time series techniques presented in Section 2.4.4. These techniques have been considered as they are often used and recognized as accurate methods in forecasting problems. From Table 2.1, three main conclusions can be drawn. First, it is clear that for each method, the monthly performance is worse than the yearly performance. This makes sense as the stochasticity due to human behavior and the influence of external factors such as the weather, averages over the course of one year, the error is thus reduced. Secondly, the performances of the LR and ARIMA models are lower than the more complex and advanced top three methods. Finally, it can be noticed that the proposed method with the additional battery trick ranks first, with a higher scalability than the light GBM model.

Table 2.2 presents a comparison of the top three methods performance metrics, when the faulty smart meters are removed from the evaluation. The removed faulty smart meters are the same 32 smart meters treated by the "battery trick" detailed in Section 2.4.5. The method proposed in this paper outperforms the two other approaches.

A t-test [126] was performed on the two sets of monthly errors to assess the statistical difference between the performance of Light GBM model (which ranked first in the competition) and the proposed scalable ensemble approach with the battery trick (which outperforms Light GBM). The null hypothesis is that there is no statistical difference between the performance metrics of both methods. The p value obtained is 0.00958. The null hypothesis is thus rejected, and we can conclude that there is a statistical difference in performance metrics

[1]We have executed the code that was provided by the authors on the same machine.

Method	Monthly rAE	Yearly rAE	Total rAE
Our proposed approach (Original submission modified to include the battery trick)	0.9802	0.2861	0.6332
1st place: Light GBM (Original submission includes the battery trick)	1.0078	0.2864	0.6471
KNN-Isolation forest ensemble (Original submission modified to include the battery trick)	1.06905	0.28633	0.67769
2nd place: KNN-Isolation forest ensemble (Original submission includes outlier detection)	1.0728	0.2875	0.6801
3rd place: Our proposed approach (Original submission, does not include the battery trick)	1.0828	0.2892	0.6860
ARIMA	1.3852	0.3844	0.8848
Linear Regression	1.4461	0.3333	0.8897
Naive baseline (30th place)	1.4947	0.4345	0.9646

Table 2.1. A comparison of three benchmark methods (naive baseline, linear regression and ARIMA) and the top three methods in the IEEE-CIS contest [132]. The methods proposed in this paper, are highlighted. Total rAE stands for relative absolute error, which is the weighted average error of the yearly prediction error and the monthly prediction error. A smaller error indicates a better performance.

Method	Monthly rAE	Yearly rAE	Total rAE
Our proposed approach	0.9757	0.2885	0.6321
Light GBM	1.0165	0.2816	0.6491
KNN-Isolation forest ensemble	1.0654	0.2887	0.6771

Table 2.2. A comparison of the top three methods' original submissions in the IEEE-CIS contest when removing the faulty smart meter out of the evaluation [132]. The method proposed in this paper, is highlighted. Total rAE stands for relative absolute error, which is the weighted average error of the yearly prediction error and the monthly prediction error.

between both methods with $p < 0.005$. The same t-test was performed on the two sets of monthly errors after removing the faulty smart meters. The p value obtained is 0.01233, the null hypothesis is thus again rejected, and we can conclude that there is a statistical difference in performance metrics between both methods with $p < 0.005$.

2.4.8 Details on the results

To better understand the approach proposed in this paper, it can be divided into several subcomponents and the impact of each one of them on the final performance metric can be analyzed. We start from the naive baseline detailed in Section 2.4.4, with a $\text{total}_{\text{rAE}}$ of 0.96. Then a prediction was built based on the sign-up month. That lead to a more accurate prediction, reducing the $\text{total}_{\text{rAE}}$ to 0.86.

The second step considered was ensemble learning and data augmentation. The same average relative consumption across all meters was used but assuming that a meter could have signed up in different months. This improved the accuracy of the prediction down to a $\text{total}_{\text{rAE}}$ of 0.80.

Third, the clustering based on k-means was introduced for meters that signed up in November or earlier, which improved the prediction to 0.75 $\text{total}_{\text{rAE}}$.

Fourth, k-nearest neighbors was considered for meters that signed up in December, which improved it to 0.72 $\text{total}_{\text{rAE}}$.

Fifth, post-processing and smoothing were added, which improved the prediction to a $\text{total}_{\text{rAE}}$ of 0.69.

Finally, 32 profiles with zero values for November and December were detected and their prediction was set to zero. This improved the prediction to a $\text{total}_{\text{rAE}}$ of 0.63.

In short, each of the individual components of the method has a similar effect, as each yields a reduction between 0.03-0.05 in the $\text{total}_{\text{rAE}}$.

2.4.9 Hyperparameter sensitivity study of the proposed method

The method makes use of only three hyperparameters through the different steps: the minimum cluster size $n_{\text{min,cluster}}$, the number of nearest neighbors n_{similar}, and the window length w for smoothing. For the following parameter values, we have observed insignificant difference in the total absolute error: $w = 2, 3, \cdots, 7$; $n_{\text{similar}} = 10, 20, \cdots, 100$; $n_{\text{min,cluster}} = 5, 6, \cdots, 20$. The best performance was obtained with the minimum cluster size $n_{\text{min,cluster}} = 10$, the number of nearest neighbors $n_{\text{similar}} = 50$, and the window length $w = 5$.

The method is robust to the values of the hyperparameters, as independently of the chosen hyperparameters, the total relative absolute error remains in the interval $[0.6860, 0.7243]$, as is shown in Table 2.3. The mean and standard

deviation of the total relative absolute error are respectively equal to $0,6991$ and $0,01034$.

2.5 Conclusion

In this paper, a method has been proposed to predict the monthly and yearly electrical consumption of individual households one year ahead based solely on historical data of the previous year. The approach consists of different steps applied sequentially: pre-processing, data augmentation and normalization, clustering, prediction based on ratios and finally ensemble learning and post-processing. This novel method fills a gap in the literature by combining a forecast horizon (one year ahead) and forecast granularity (individual household) which has not been tackled yet. The model offers three main advantages. (i) Low data requirement: the method works even for households with only one month of historical data, and it does not require additional household-specific information nor weather data. (ii) Scalability: the method has low computational requirements and does not involve household-specific information collection through questionnaires. (iii) Accuracy: the use case is based on data from the IEEE-CIS competition, and the third position in the competition demonstrates the high performance of the model. The battery trick places the method in the first position, without altering the scalability. However, we think it is not a fair modification because it aims to estimate problems in the ground-truth data to improve the metrics, which wouldn't provide any benefit in practice.

In future work, the method could be improved by detecting outlier household profiles and defaults (i.e., measurement errors) in the historical data and treating these with a customized model. The model presented in this paper could also be applied on new datasets in order to assess the generalization of the method. As suggested in the literature [88], the approach could also be applied in the context of higher granularity forecasting by aggregating household predictions to the national level.

Minimum cluster size	Nb of nearest neighbors	Window size	Monthly rAE	Yearly rAE	Total rAE
10	50	2	1.1039	0.2860	0.6950
10	50	3	1.0993	0.2857	0.6925
10	50	4	1.1039	0.2860	0.6950
10	50	5	1.0728	0.2875	0.6801
10	50	6	1.1039	0.2860	0.6950
10	50	7	1.0987	0.2913	0.6950
10	10	5	1.0918	0.2879	0.6898
10	20	5	1.0908	0.2876	0.6892
10	30	5	1.0908	0.2873	0.6890
10	40	5	1.0909	0.2871	0.6890
10	60	5	1.0909	0.2869	0.6889
10	70	5	1.0912	0.2868	0.6890
10	80	5	1.0917	0.2868	0.6892
10	90	5	1.0919	0.2868	0.6893
10	100	5	1.0922	0.2867	0.6895
10	50	5	1.1208	0.2922	0.7065
5	50	5	1.1269	0.3146	0.7208
6	50	5	1.1459	0.3026	0.7243
7	50	5	1.1358	0.2994	0.7176
8	50	5	1.1328	0.2987	0.7157
9	50	5	1.1012	0.2898	0.6955
10	50	5	1.1193	0.2905	0.7049
11	50	5	1.1141	0.2901	0.7021
12	50	5	1.1254	0.2896	0.7075
13	50	5	1.1210	0.2903	0.7056
14	50	5	1.1257	0.2920	0.7088
15	50	5	1.1212	0.2911	0.7061
16	50	5	1.1175	0.2904	0.7040
17	50	5	1.0973	0.2875	0.6924
18	50	5	1.0972	0.2879	0.6926
19	50	5	1.1181	0.2890	0.7035
20	50	5	1.1074	0.2886	0.6980

Table 2.3. Comparison of the performance metrics of the method when varying the values of the three hyperparameters. The chosen hyperparameters are highlighted. A smaller error indicates a better performance. The method is robust to the values of the hyperparameters, as independently of the chosen hyperparameters, the total relative absolute error remains in the interval [0.6860, 0.7243].

Chapter 3

Probabilistic short-term forecasting of individual households consumption in low voltage grids

This chapter is published and reprinted with permission from:

Botman, Lola; Lago, Jesus; Agudelo, Oscar Mauricio; Vanthournout, Koen; De Moor, Bart, "A scalable method for probabilistic short-term forecasting of individual households consumption in low voltage grids," 2023 IEEE PES Grid Edge Technologies Conference & Exposition (Grid Edge), San Diego, CA, USA, 2023, pp. 1-5.

Botman, Lola; Lago, Jesus; Becker, Thijs; Vanthournout, Koen; De Moor, Bart, "A global probabilistic approach for short-term forecasting of individual households' electricity consumption," **Submitted** in Applied Energy, November 2023.

Abstract

Accurate short-term prediction of individual residential load is essential for various applications and low voltage grid stakeholders. Distribution system operators can utilize these forecasts in grid simulations, operation planning, and to anticipate unusual or varying consumption in certain neighborhoods, which can help to avoid congestions caused by peaks. Furthermore, these predictions can be used to optimize the low voltage grid for renewable energy sources and potential battery storage. Due to the highly stochastic nature of household consumption, point forecasting is not optimal. We apply state-of-the-art probabilistic methods from other applications, and we propose a novel probabilistic global approach. It is based on a selection of similar normalized household consumption time series, which are then used to compute empirical quantiles. The proposed method has been evaluated on two different open datasets from different countries, and it outperforms state-of-the-art models in both cases. It has also been applied to a private Belgian dataset. In all three cases, the proposed method consistently outperforms the state-of-the-art methods on highly stochastic and changing households, i.e. experiencing concept drift. This method is scalable, with low computational requirements, requires only seven days of historical data of the target household to make predictions and does not require household-specific or weather information.

3.1 Introduction

Accurately forecasting the electrical consumption of individual households, also called households load forecasting, is necessary for a wide range of applications. Different applications require different forecasting dimensions, i.e., horizon and spatial aggregation [97]. For example, predicting day-ahead consumption, including potential peaks, allows operators to anticipate and take actions to mitigate congestion.

In this paper, which is an extension of [19], we compare several state-of-the-art algorithms and propose a novel approach for short-term probabilistic forecasting of individual household electricity consumption. The selected benchmark methods are inspired by successful applications in other short-term probabilistic forecasting applications. The proposed method allows us to make probabilistic predictions for one day ahead, computing a predetermined number of quantiles per time step. It has the interesting property of being global, i.e., it uses one model for all households. This is in contrast with most methods proposed in the literature which are categorized as local, i.e., one model is trained per household. Grabner et al. [54] also discuss global versus local models in the case of point forecasting. We compute the forecasts as follows. First, we select similar households, based on seven days of historical data of the target household. Next, we calculate empirical quantiles. Additionally, our method only uses households load data and does not require other data sources such as weather or household-specific attributes.

3.1.1 Related work

Recently, research on residential consumption forecasting at the household level has gained more attention due to the installation of smart meters. It is however still relatively immature [57] compared to forecasting of loads at higher aggregations levels, such as city or national levels [17, 128]. Low voltage and household level forecasting are important for the efficient operation of the energy system, especially with the increasing use of renewable energy sources, electrical vehicles, and heat pumps, which all add complexity and uncertainty to household electrical consumption time series, also called consumption profiles [64]. Short-term household consumption forecasting is also necessary to make the energy network more intelligent [39] and to potentially influence customer behavior towards a reduction of energy usage and congestion avoidance mechanisms. The key difference between household consumption profiles and aggregated profiles at the substation or city level is the variability of the data. Household consumption

profiles display higher levels of stochasticity compared to aggregated profiles due to the individual inhabitants' behavior.

The majority of research in this field has focused on point forecasting methods, i.e., methods where a single value is predicted per time point in the future, while probabilistic forecasting methods, i.e., methods where a distribution or spread is predicted per time point in the future, are more appropriate and informative due to their ability to provide information about the uncertainty of predictions. The first review paper [64] on probabilistic forecasting included all forecasting dimensions, i.e., from short-term to long-term and also different spatial aggregations. However, there is limited research on individual household consumption forecasting, and the authors of [64] only refer to [58]. A recent comprehensive review on low voltage load forecasting [56] also highlights the lack of research at the household level and calls for a shift towards probabilistic methods. More recently, Haben et al. published an entire book on load forecasting at the low voltage level [57]. Household load forecasting is mentioned a few times, mostly to present the multiple challenges due to the dependency on human behavior, unpredictability and peaky nature of the profiles. In the 'Advanced and Additional Topics' chapter, a dedicated section further highlights the distinct difficulty and complexity of research in the domain of household load forecasting. There is a strong need for further study of probabilistic forecasting methods, case studies, and evaluation methods at the household level.

In the context of residential load forecasting, peak consumption is of significant interest, as a high simultaneity of peaks may result in congestion on the low-voltage grid. Conventional point forecasting methods and their accompanying loss function (error metric) tend to overlook peaks as they generate substantial errors even when correctly predicted in terms of height and duration, albeit with a small time lag. This phenomenon is referred to as the double peak penalty effect. To address this issue, alternative error metrics can be considered [58]. Another approach is to utilize probabilistic forecasting, which can overcome this limitation; as the predictions now account for the spread of the data, they can more accurately capture variations in the timing of peaks. Using probabilistic methods is the recommended alternative since they can be assessed objectively using proper scoring functions [57]. Scenario generation is another field of research that allows to generate realistic daylong electricity consumption profiles [122].

This paper aims to address certain limitations in the existing load forecasting literature. One limitation is the lack of comparisons with baselines and/or existing state-of-the-art models [57]. However, conducting such comparisons can be challenging, in cases where the proposed methods are applied to private datasets [70] (or datasets that are hard to access in practice) or when the code is not readily accessible [35]). These factors hinder reproducibility and

comparisons. Another limitation is the reliance of most papers on a single dataset, which makes it difficult to draw generalizable conclusions about a model's performance. Some studies have shown that incorporating additional attributes, such as household-specific information like the number of rooms or heating types, can improve forecasting accuracy [77, 149]. However, when there is a need to scale the forecasts to an entire region, it is not feasible to obtain detailed information about each individual household. To overcome these challenges, this paper proposes a new probabilistic approach for short-term forecasting of residential consumption. The proposed method relies only on historical data and calendar information; it does not require additional attributes. The method is evaluated on three distinct datasets, two of which are publicly available, and the accompanying code is provided [18].

An additional critical challenge in the field of time series forecasting, particularly pertinent to individual household electricity consumption, is the phenomenon of concept drift. Concept drift refers to the change in statistical properties of the variable over time, which can be due to evolving consumer behaviors, technological advancements, or changes in external factors like the covid pandemic. This drift often results in the degradation of forecast accuracy as the models, trained on historical data, fail to adapt to the new data distribution. This is especially problematic in short-term household consumption forecasting, where rapid shifts in user behavior or sudden introduction of new appliances such as electrical vehicle or heat pumps, can significantly alter consumption patterns. The challenge lies in developing forecasting models that are robust to such changes and can adapt over time [54, 60].

3.1.2 Motivation and contributions

This paper proposes an accurate and global probabilistic method to forecast residential electricity consumption based on empirical quantiles. Because of the high stochasticity of individual household consumption profiles, probabilistic methods are needed to gain insight into the uncertainty of the predictions. The paper contributes to the literature by proposing a method that:

- generates accurate short-term probabilistic forecasts of individual households electrical consumption, consistently outperforming state-of-the-art methods on profiles experiencing concept drift, using three different datasets;
- generates forecasts for households with only seven days of historical data of the target household;

- is compared with state-of-the-art methods from other energy applications, which have not been applied yet on short-term probabilistic forecasting of household data.

For reproducibility, two of the datasets used, are publicly available, and the code is made available to the scientific community [18]. Additionally, our method presents interesting properties. It does not rely on exogenous data, such as weather or household-specific attributes; it generalizes well, i.e., it outperforms state-of-the-art methods on several datasets; it is scalable, thanks to the low data requirement and low model complexity [21]; and the model is easily understood, unlike black-box models.

3.1.3 Organization of the paper

The paper is organized as follows. Section 3.2 presents the background knowledge necessary to understand the method, it includes details on empirical quantiles (Subsection 3.2.1), the different benchmark models (Subsection 3.2.2) and the performance metric (Subsection 3.2.3). The different steps of the method are detailed in Section 3.3, as well as the model parameters. The case study is explained extensively in Section 3.4, including information on the different datasets (Subsection 3.4.2) and the results (Subsection 3.4.3). Finally, the outcome and perspectives are discussed in Section 3.5.

3.2 Background Knowledge

3.2.1 Empirical quantiles

Quantiles are a fundamental concept in statistics and data analysis. They represent specific points in a distribution that divide the data into equal portions. They provide valuable information on the spread and variability of data. Rob J. Hyndman and Fan [68] examine the calculation of sample quantiles, highlighting variations in existing algorithms, mathematical assumptions and their impact on estimation. Sample quantiles, also called theoretical quantiles, are defined based on a theoretical (either given, known or assumed) distribution. Empirical quantiles, on the other hand, are calculated directly from observed data. An empirical cumulative distribution function (CDF) can be constructed by estimating quantiles. The approximation of the CDF is closer to the true CDF, if more observations are available. To estimate a given q-quantile x_q, the dataset is arranged in ascending order, and the q-quantile is the value in the

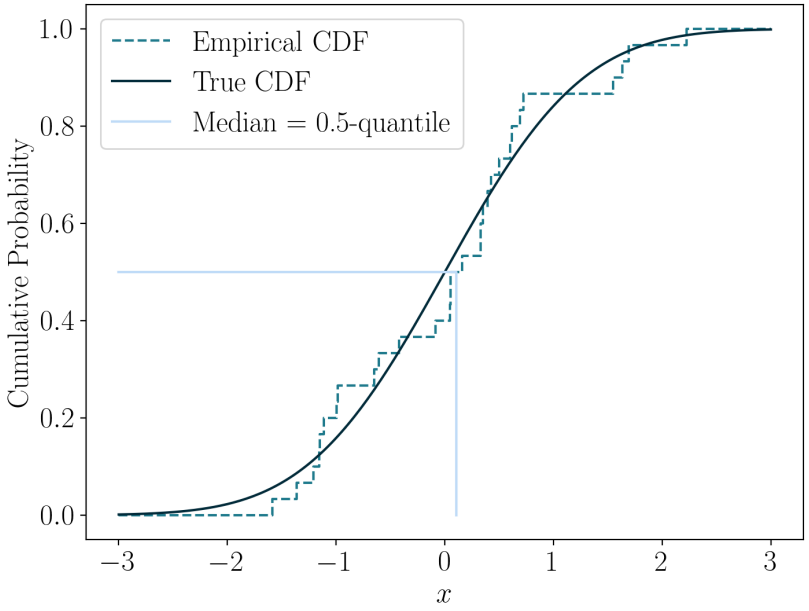


Figure 3.1. Empirical CDF (dotted blue) with true CDF (black) for the standard normal distribution. The empirical CDF is generated using 30 randomly drawn points from the true CDF. The median (0.5-quantile) is also illustrated with the light blue line. The figure originally appeared in [57], and has been reproduced in this paper with the addition of the median.

dataset that corresponds to the proportion of observations q that are less than or equal to that value. For example, to compute the median $x_{q=0.5}$, i.e., also called the 0.5-quantile, or the 50th percentile, half of the observations will fall below this value $x_{q=0.5}$ and half will fall above it. Figure 3.1, inspired from [57], illustrates an empirical CDF (in dotted blue line) with true CDF (in black) for the standard normal distribution. The empirical CDF is generated using 30 randomly drawn points from the true CDF. The median is also indicated with the light blue line.

Empirical quantiles play a crucial role in understanding the distributional characteristics of a dataset and in estimating percentiles without making assumptions about the underlying distribution. Ruppert and Matteson [110] explore the application of empirical quantiles in financial engineering, specifically in the context of probabilistic forecasting. By using historical data to estimate empirical quantiles, it becomes possible to generate probabilistic forecasts that capture the uncertainty associated with future observations. This approach is particularly useful in financial markets, where volatility and risk management

are of great importance. Likewise, a high stochasticity and variability can be observed in individual household data, which makes empirical quantiles relevant. Moreover, there is a parallel objective shared with distribution system operators, namely the efficient management of congestion risk while minimizing costs. Although [71] deals with long-term predictions, it highlights the added value of empirical quantiles in load forecasting. It's worth noting that terminology can vary between different sources and disciplines and that sample quantiles and empirical quantiles are for example used as synonyms in [57, 110], but not in [68].

3.2.2 Benchmark methods

The benchmark models utilized in this study encompass established and state-of-the-art approaches employed in the literature for this specific application or successful in other applications of probabilistic forecasting. The proposed baseline consists of simplistic approaches that rely on naive assumptions and do not require extensive modeling. Each of the models described in this section generates 100 quantiles at each time step, for a one-day horizon.

Linear Quantile Regression

A linear quantile regression model predicts quantiles, by optimizing the Pinball loss[1]. There is one model per quantile to be predicted in the future, in this study, we choose to estimate 100 quantiles. Quantile regression has the advantage of not making any assumption on the distribution of the dataset. The model utilizes two days of historical consumption data preceding the target day, in addition to the corresponding day one week ago and two weeks ago. Furthermore, calendar information are included as input variables, i.e., the hour of the day, the day of the month and the month of the year, each cyclically encoded, and a binary indicator for weekends versus weekdays. The python library Scikit-learn allows for a straightforward implementation. The reader is referred to [57, 78] for more details on quantile regression and to [143, 144] for the application of quantile regression to short-term probabilistic load forecasting.

Quantile Regression Averaging (QRA)

Quantile Regression Averaging is relatively new to the field of probabilistic load forecasting. It is inspired by successful applications in electricity price

[1] More details about the Pinball loss are provided in Section 3.2.3.

forecasting [65]. Liu et al. [89], applied it on aggregated load data. To the best of the authors' knowledge, it is applied to individual household load data for the first time in this paper. QRA starts by generating multiple point forecasts from different models, called "sister forecasts". In this work, the sister forecasts are built using four well-known point forecasting methods: a feedforward neural network [19], a lasso regression [130, 135], an XGBoost model [16, 28], and a naive persistence model, i.e., the predicted consumption for a given day is the consumption observed on the previous day. The point forecasting models produce a single value for each time step in the one-day ahead horizon. The inputs for the first three models are the same: historical consumption data, including the two preceding days, a full day from seven days ago, and a full day from fourteen days ago. These inputs capture recurring patterns in customer behavior. Additionally, the models incorporate cyclically encoded calendar and time features such as the hour of the day, day of the month, and month of the year, along with a binary indicator distinguishing weekdays from weekends. The sine and cosine encoding ensures the algorithm understands the cyclical nature of these features, e.g., 23h00 in the evening is close to 1h00 in the morning. QRA constructs quantiles by combining the ensemble of sister forecasts. This method employs quantile regression to generate 100 quantiles per time step, for a one-day horizon. The four point forecasts serve as inputs for the quantile regression process. This averaging step aggregates the individual predictions, resulting in an improved overall probabilistic forecast that captures a wider range of potential outcomes than a single point forecast. QRA has the advantage of leveraging the existing literature on point forecasting and does not rely on high-quality expert forecasts [89].

Bernstein Normalizing Flows (BNF)

This model uses Bernstein normalizing flows to transform a simple distribution to a more complex distribution. A neural network is trained to predict the parameters of the flows. These parameters yield the conditional probability distributions for all the time steps, one day ahead. Seven days of data prior to the predicted day are used as input and there are five additional exogenous inputs, i.e., the sine and cosine encoding of the day of the year, and the day of the week, as well as a binary indicator for the holidays. The sine and cosine encoding allows the algorithm to understand that the features are cyclical, e.g., 23h00 in the evening is close to 1h00 in the morning [106,138]. The authors claim that their method outperforms a simple Gaussian Model, a Gaussian Mixture Model and a Quantile Regression. The reader is referred to the full papers and implementation for details [11–13]. As this approach outputs distribution functions per time step, we computed quantiles from these distributions and then applied our own implementation of the discretized Continuous Ranked

Probability Score (CRPS) on the quantiles (detailed in Section 3.2.3). Note that the BNF method results presented in this paper are slightly different from the results in [12, 13] as we have re-run them locally, with our own (not normalized) CRPS implementation, in order to allow for exact comparison with the other methods.

Feedforward Neural Network & empirical quantiles

This approach, illustrated with the flowchart in Figure 3.2, consists of three main modules, applied sequentially, on each profile individually: preprocessing each profile individually; training the point forecasting model on the training set; applying the point forecasting model on the validation set; computing and grouping the errors on the validation set predictions per time step; and finally applying the point forecasting model on the test set and adding the empirical quantiles to point forecasts to generate probabilistic forecasts.

The data **preprocessing** is kept to its minimum, i.e., each household profile is normalized between $[0, 1]$ using MinMaxScaling [104]. Each profile is split in time sequentially: the first 60% of the time steps are used for **training**, the next 20% for validation and the last 20% for testing.

The next step is to **train the point forecasting model**. Many methods have been proposed for short-term household forecasting, such as classical times series approaches (Exponential Smoothing, Autoregressive Integrated Moving Average ARIMA) [31], Long Short Term Memory networks (LSTM) [79], Recurrent Neural Networks (RNN) [106], Convolutional neural networks (CNN) [9], Support Vector Regression (SVR) or more advanced methods such as Gradient Boosting (XGBoost) [139] or combinations of these [8]. Deep learning methods tend to perform better on short-term household load forecasting, However it is difficult to state which method performs best out of all published deterministic methods, as they use different datasets, different input features, different preprocessing techniques, etc. One method may perform better on a dataset and worse on another, and there is no evidence that the complex methods outperform the more simple ones [82]. Amarasinghe et al. [9] compared CNN, LSTM, SVR and FFNN for one step ahead residential consumption forecasting, on a same dataset, and stated that all four techniques resulted in similar performance metrics. Any of the existing point forecasting algorithm could be used in this step. We choose to use a Feed Forward Neural Network (FFNN), since it is widely known, it has similar performances to- and lower computational requirement than other machine learning methods, and is one of the easiest to implement. Artificial neural networks are also recommended by Kuster et al. [82] in the specific setting of short-term household load forecasting with

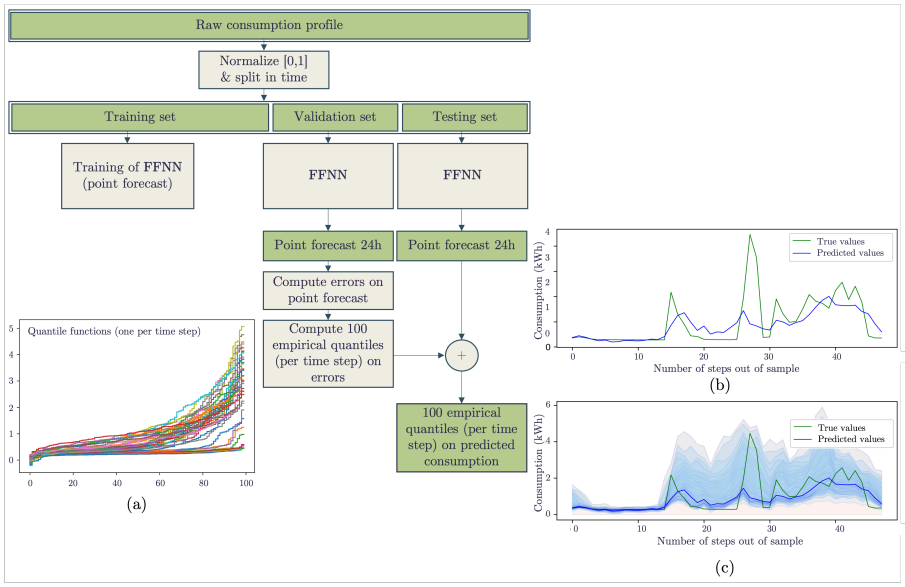


Figure 3.2. Flowchart of the proposed method, applied individually on each profile. Each consumption profile is first preprocessed, then split in training, validation and testing set. The Feedforward Neural Network (FFNN) is trained on the training set. The FFNN is then applied on the validation set. The empirical quantiles are computed on the errors of the validation set. Finally, the FFNN is applied on the test set and the empirical quantiles are added to the point forecast in order to obtain a probabilistic estimate of the future consumption. (a) 100 quantiles are computed per time step, there is one quantile functions per time step in the day. (b) Point forecasting results one day ahead. (c) Point forecasting with empirical quantiles.

sub-hourly resolution. The outputs of the neural network are the T consumption predictions one day ahead:

$$\hat{y}_d = [\hat{y}_{d_1}, \hat{y}_{d_2}, ..., \hat{y}_{d_T}] \in \mathbb{R}^T, \tag{3.1}$$

where d indicates the forecast day and T is the number of predicted time steps and depends on the dataset time resolution. The input of the model consists of a number of days of the historical consumption $y_i \in \mathbb{R}^T$: (i) the two days of consumption values prior to the forecast day y_{d-2} and y_{d-1}, (ii) one full day seven days prior to the forecast day y_{d-7} and (iii) one full day fourteen days prior to the forecast day y_{d-14}. Indeed, one can expect a typical customers' behavior (and his subsequent electrical consumption) on a Wednesday to be similar to the previous Monday and Tuesday, as well as the Wednesday of the two previous weeks. Additionally, calendar and time features of the day to be predicted are added: (iv) hod_d the hour of the day, (v) dom_d the day of the month, and (vi) moy_d the month of the year, are cyclically encoded using a sine

and cosine function, and a (vii) binary indicator wk_d is used for weekday versus weekend. The sine and cosine encoding allows the algorithm to understand that the features are cyclical, e.g., 23h00 in the evening is close to 1h00 in the morning [106, 138]. This means that in total, the input consists of four days of historical consumption values and seven calendar/time features. The input vector is denoted:

$$\begin{aligned}&[y_{d-2}, y_{d-1}, y_{d-7}, y_{d-14}, \sin(\text{hod}_d),\\ &\cos(\text{hod}_d), \sin(\text{dom}_d), \cos(\text{dom}_d), \qquad \in \mathbb{R}^{4T+7}.\\ &\sin(\text{moy}_d), \cos(\text{moy}_d), \text{wk}_d]\end{aligned} \tag{3.2}$$

The trained point forecasting model is applied on the **validation set**. It generates one day ahead forecasts $\hat{y} = [\hat{y}_1, \hat{y}_2, ..., \hat{y}_T]$ for the m days in the validation set. The errors e (in dark yellow to khaki green on Figure 3.3) between the predictions $\hat{y}$ (in blue) and the true consumption y (in green) are computed for the m days. There are m error vectors of size T:

$$e = y - \hat{y} = [e_1, e_2, ..., e_T] \in \mathbb{R}^T. \tag{3.3}$$

The **errors are then grouped** per time steps, i.e., the first group g_1 consists of all prediction errors made at the first time step (00h00), the second group g_2 consists of all errors made at the second time step (00h30), and so on. In total, there are T sets of m errors, one for each time step of the day:

$$\begin{aligned}g_1 &= [e_1, e'_1, ..., e_1^{(m)}] \in \mathbb{R}^m\\ g_2 &= [e_2, e'_2, ..., e_2^{(m)}] \in \mathbb{R}^m\\ &\vdots\\ g_T &= [e_T, e'_T, ..., e_T^{(m)}] \in \mathbb{R}^m.\end{aligned} \tag{3.4}$$

Based on each set of error g_i, 100 **empirical quantiles** are computed. The 100 quantiles represent the error distribution of the point forecasting at each time of the day. In total, there are 100 quantiles for each of the T time steps, there are T quantile functions. An example is shown in Fig. 3.2(a).

Finally, the point forecasting model is applied to the **test set**, the output is one day ahead deterministic predictions, as illustrated on Fig. 3.2(b). The previously computed quantiles are added to these point predictions, this generates uncertainty bounds around the predictions, as shown in Fig. 3.2(c).

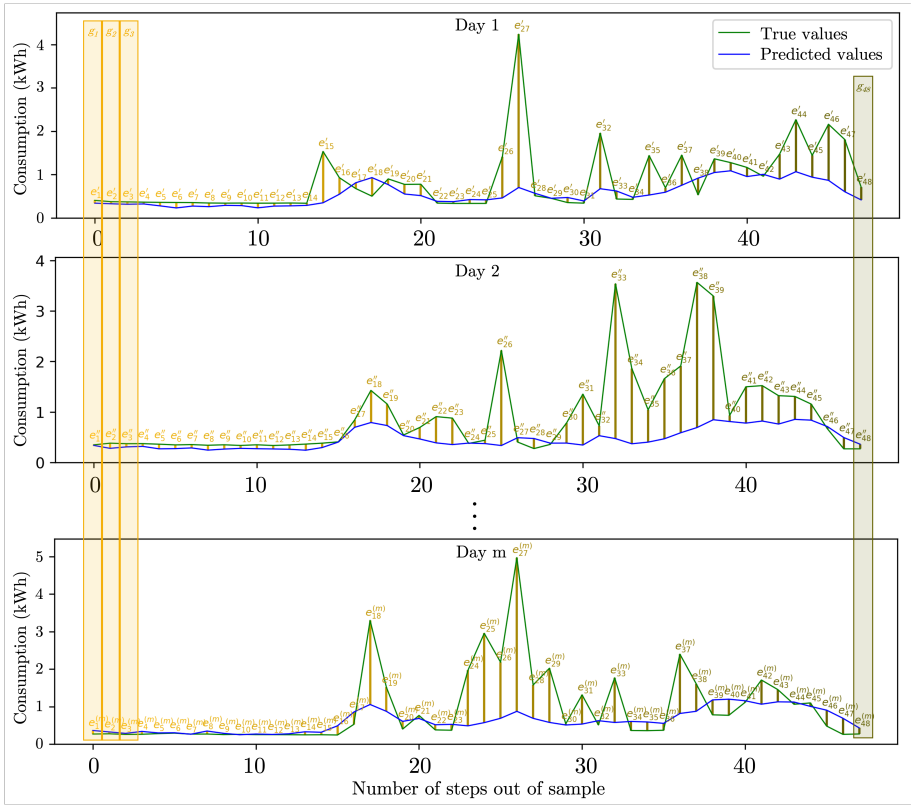


Figure 3.3. The errors between the predicted consumption (in blue) and the true consumption (in green) are computed, they are represented by the vertical lines between the predicted and the true consumption values at each time step. The errors are then grouped per time step to generate the empirical quantiles. The grouping are represented by the vertical rectangles.

Baselines

Two different baseline methods are proposed in this work. The first one is the same as in [19]. One can assume that, in the absence of mathematical formulations or complex modeling, the most reasonable hypothesis is that the current day's consumption will resemble that of the previous day, commonly known as a persistence model in point forecasting. In line with this assumption, we propose a probabilistic baseline that shares the same grounding. In Section 3.2.2, we introduced the FFNN + Empirical Quantiles model. Building upon this model, the baseline utilizes the persistence model for point forecasting instead of the FFNN, and subsequently applies the empirical quantiles' step.

The second baseline, named historical sampling, relies on the same assumption, but instead of going through a point forecasting step, the empirical quantiles are computed directly on the historical consumption, e.g., the predicted quantiles of the consumption of a Monday at 15:00 are computed using the consumption of all previous Mondays at 15.00.

3.2.3 Performance Metrics

In the context of point forecasting, where a prediction of a specific value at a given time step is made, the evaluation of the prediction's accuracy is commonly quantified using error measures such as the Mean Absolute Error (MAE) and the Mean Squared Error (MSE), where the absolute (or squared) error is computed at each time step and then averaged:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| = \frac{1}{n}\sum_{i=1}^{n} |e_i|\,, \tag{3.5}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2\,, \tag{3.6}$$

where y_i are the true consumption values, $\hat{y}_i$ are the predicted consumption values and n is the number of samples. In the case of probabilistic forecasting, where a distribution of potential future outcomes is predicted directly, the evaluation of the prediction's accuracy could be quantified by taking the mean or the median of the predicted distribution, and computing the Mean Absolute Error (MAE) and Mean Squared Error (MSE) with respect to the true values [12, 119]. However, these conventional metrics do not fully capture the probabilistic information included in the prediction. Therefore, it is necessary to use a metric that is better suited for probabilistic forecasts, such as the Continuous Ranked Probability Score (CRPS), which takes into account both the calibration (accuracy) and sharpness (spread) of the predicted distribution [51,52,101]. The CRPS is calculated individually for each predicted time step and then averaged over the entire horizon. The metric for one measurement can be computed as

$$\mathrm{CRPS}\left(\hat{F}, y\right) = \int_{-\infty}^{\infty} (\hat{F}(x) - \mathbb{1}\{y \leq x\})^2 \,\mathrm{d}x \tag{3.7}$$

where $\hat{F}$ is the predicted cumulative distribution function, y is the observed consumption value, and $\mathbb{1}$ is the indicator (also called Heaviside) function, i.e. takes value 1 when $y \leq x$ and 0 otherwise (in blue line on Figure 3.4). The CRPS is a commonly used metric for probabilistic forecasting and can

be intuitively interpreted as a sum of areas[2] between the predicted CDF and the actual value. It should be noted that the CRPS entails a more intricate calculation than solely computing the area between the predicted CDF and the observed value, as illustrated in details in Fig. 3.4 [24]. However, the CRPS evaluates the continuous CDF, while our proposed method generates quantiles. It is thus important to consider another version of the CRPS [101], which uses the quantiles, in order to evaluate the performance of the proposed method:

$$\mathrm{CRPS}\left(\widehat{F}, y\right) = \sum_{q \in [0,1]} \left\{ \mathrm{Pinball}\left(\widehat{Q}(q), y, q\right) \right\}, \tag{3.8}$$

$$\begin{aligned} &\text{where Pinball}\left(\widehat{Q}(q), y, q\right) \\ &\quad = \begin{cases} (1-q)\left(\widehat{Q}(q) - y\right), & \text{for } y < \widehat{Q}(q), \\ q\left(y - \widehat{Q}(q)\right), & \text{for } y \geq \widehat{Q}(q), \end{cases} \end{aligned} \tag{3.9}$$

and $\widehat{Q}(q)$ is estimated the q-quantile of the consumption and y is the actually observed consumption. Our models generate 100 quantiles, and the sum is thus computed for $q = [0, 0.01, 0.02, \ldots, 0.99, 1]$.

The Pinball loss is also called the quantile loss, and measures the accuracy of one quantile. It is an asymmetric loss function that takes the difference between the observed value and the predicted quantile and assigns a different weight based on the sign of this difference. Indeed, the q-quantile $\widehat{Q}(q)$ is accurately predicted when a proportion q of the observation are below $\widehat{Q}(q)$ [57]. The Pinball loss score, also called the quantile score, is obtained by averaging over multiple quantiles, and multiple time steps. A lower pinball score indicates a better forecast. The discretized CRPS can also be interpreted visually, as illustrated in Fig. 3.4(b), it is a sum of overlapping rectangular area, shaded in green. Furthermore, the CRPS is a generalization of the MAE, to which it reduces if $\hat{F}$ is a point forecast [50].

Note that we compute error metrics on the unnormalized values. Unnormalized error metrics offer valuable insights into the absolute performance of forecasting models, providing distribution system operators with the true error values relevant to their applications. Additionally, absolute metrics allow us to identify challenging profiles.

[2]Note that several references illustrate the CRPS as the exact area between the indicator function and the predicted CDF. While this intuition behind the CRPS is correct, the representation is not entirely accurate [22, 57].

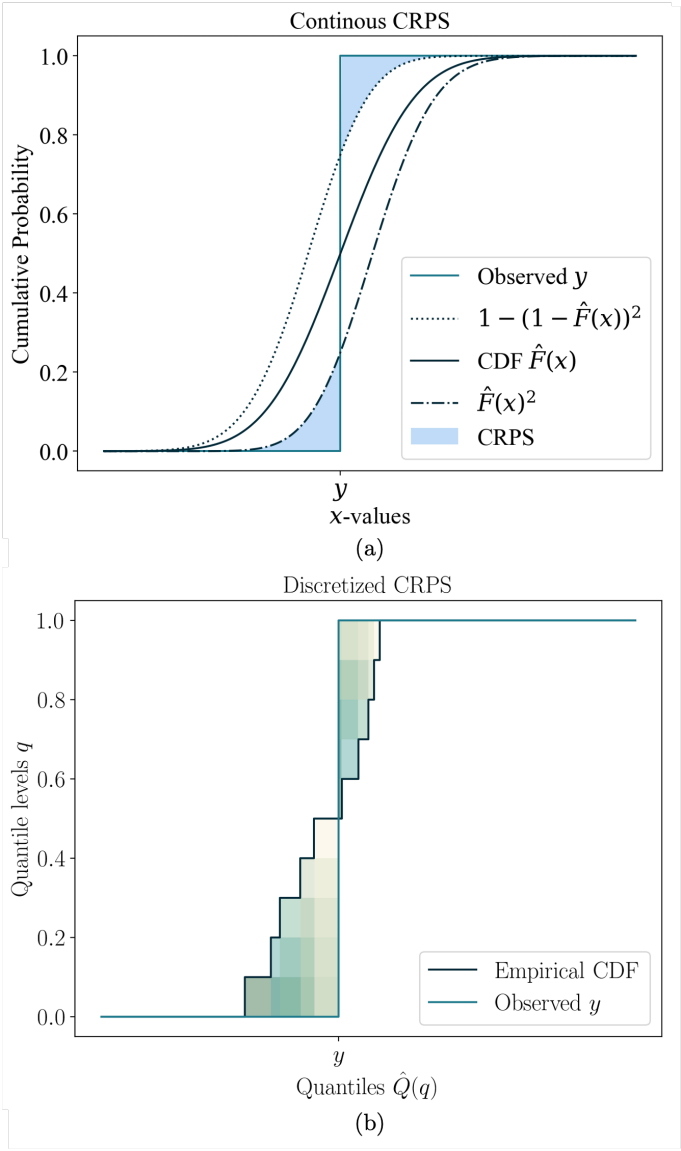


Figure 3.4. (a) Visual representation of the continuous CRPS computation between predicted cumulative distribution (black) and observed distribution (blue) [24]. The CRPS is the light-blue colored area. (b) Visual representation of the discretized CRPS computation between empirical cumulative distribution (black) and observed distribution (blue). The CRPS is calculated as the sum of all the overlapping rectangular areas shaded in green. It's important to note that for the sake of clarity in this illustration, only ten quantiles are shown, while the proposed method computes 100 quantiles.

3.3 Methodology: Global Approach

The method generates one day ahead probabilistic predictions for individual household consumption profiles, where 100 quantiles are provided for each time step. The data cleaning is presented in Section 3.3.1. The approach consists of three sequential steps applied on each profile: normalization, selection of similar households and prediction. The steps are detailed in subsections 3.3.2-3.3.4. The method can easily be adapted to a specific dataset thanks to the modular approach.

3.3.1 Data cleaning

The data cleaning and pre-processing is highly dependent on the dataset used. In each case, it is kept to the bare minimum to allow for reproducibility on other datasets. Only residential profiles are used, as it is of specific interest to this research. Unique missing values, i.e., one missing time step between two non-missing time steps, are linearly interpolated and incomplete profiles, i.e., profiles where several weeks or months are missing, are discarded. The amount of time steps in one day depends on the resolution of the dataset, e.g., half-hourly profiles have 48 time steps in one day, while quarter-hourly profiles have 96 time steps in one day. The number of time steps in one day is denoted by T, the daily consumption of a household j for day d is noted as:

$$\mathbf{y}_d^j = \left[y_{d_1}^j, y_{d_2}^j, \ldots, y_{d_T}^j \right] \in \mathbb{R}^T \tag{3.10}$$

where $y_{d_t}^j$ is the quarter-hourly or half-hourly electricity consumption of household j, at time step t, during day d.

3.3.2 Normalization

The goal of the normalization step is to enable the comparison of different profiles with each other, to find similar patterns and use these later on to make predictions. Different types of normalization are used, depending on the properties of the dataset. In this study, the normalization process involves scaling the data between 0 and 1 in the case of positive values only, or between -1 and 1 for datasets with negative values. Negative values correspond to an excess electricity production of solar panels injected on the grid by households. The

scaling is based on the maximum absolute value of the training set[3], ensuring that zero values remain unchanged throughout the normalization procedure, and that the test set is unused[4].

- Profile normalization[5]: the profile is normalized between 0 and 1, or between -1 and 1 for datasets with negative values, according to the maximum absolute value of the training set. The profile normalization allows to compare profiles with similar patterns, but different absolute values. For a quarter- or half-hourly value y_t^j, of household j, at time step t, the "profile normalized" value, indicated by the subscript p, is

$$r_{p,t}^j = \frac{y_t^j}{\left|y_p^j\right|_{\max}}. \tag{3.11}$$

  where $\left|y_p^j\right|_{\max}$ is the maximum absolute consumption in the training set of household j.

- Week normalization: Each set of seven days is normalized between 0 and 1, or between -1 and 1 for datasets with negative values. This allows for comparison between profiles and within a profile. We assume that weekly patterns might be present through the year, which might go unnoticed with the profile normalization. For a quarter- or half-hourly value y_t^j of a given week w of household j, at time step t, the "week normalized" value, indicated by the subscript w, is

$$r_{w,t}^j = \frac{y_t^j}{\left|y_w^j\right|_{\max}}. \tag{3.12}$$

  where $\left|y_w^j\right|_{\max}$ is the maximum absolute consumption in the week w of household j.

- Day normalization: Each day is normalized between 0 and 1, or between -1 and 1 for datasets with negative values. As in the week normalization,

[3]It should be noted that the normalized test data might be outside the [-1,1] interval if there is a value in the test set larger than the maximum value of the training set.

[4]We also implemented and applied a robust scaler, based on the median and the interquartile range [57], which recommended to avoid the effect of outliers. However, this did not yield better results and was not included in the paper.

[5]We also implemented profile normalization, after removing the unusual high peaks: first, the maximum $m\%$ of the training set absolute values are replaced by the maximum of the remaining $100-m\%$ of the training set and then the "no peak" profile is normalized between 0 and 1, or between -1 and 1 for datasets with negative values. The idea to remove the very high peak values is to allow for comparison, without taking into account households' potentially unusual and unrepresentative behavior. However, we did not notice any improvement and thus did not include it in the results.

the idea is to look for similar days between profiles and within a profile with a finer granularity than the week normalization. For a quarter- or half-hourly value y_t^j, of a given day d, at time step t, of household j, the "day normalized" value, indicated by the subscript b, is

$$r_{b,t}^j = \frac{y_t^j}{\left|y_b^j\right|_{\max}}. \tag{3.13}$$

where $\left|y_b^j\right|_{\max}$ is the maximum absolute consumption of in the day b of household j.

A normalized profile for a full day d, irrespective of the normalization used, is denoted by $\mathbf{r}_d^j$, in coherence with Equation (3.10):

$$\mathbf{r}_d^j = \left[r_{d_1}^j, r_{d_2}^j, \ldots, r_{d_T}^j\right] \in \mathbb{R}^T. \tag{3.14}$$

If relevant, the type of normalization is specified by the subscript: $\mathbf{r}_{p,d}^j$ for a "profile normalized" day, $\mathbf{r}_{w,d}^j$ for a "week normalized" day, and $\mathbf{r}_{b,d}^j$ for a "day normalized" day.

3.3.3 Selection of similar profiles

This section provides a detailed explanation of the procedure to select similar profiles, regardless of the normalization used. The general idea is to examine the consumption patterns of a given household during the seven days prior to the day to be predicted, and to compare these seven days with all seven-day periods of all households in the historical dataset. Utilizing the k-nearest neighbors algorithm, a number of households that exhibit the closest resemblance (along with their corresponding seven-day periods) are selected. Our method then uses the consumption data of these selected similar households, as detailed in Section 3.3.4 to make predictions, and thus only requires minimal historical data of the household to be predicted, i.e., the seven days prior to the target day.

First, the set $\mathcal{S}_d^j$ is formed, consisting of the seven normalized consumption days of household j, prior to the target day d, where normalized consumption values are denoted as $\mathbf{r}_d^j$:

$$\mathcal{S}_d^j = \left[\mathbf{r}_{d-1}^j, \mathbf{r}_{d-2}^j, \ldots, \mathbf{r}_{d-7}^j\right]^\intercal \in \mathbb{R}^{7\times T}. \tag{3.15}$$

Then, a collection $\mathcal{P}$ is built, using a rolling window of seven days, with a one-day increment (and thus a six-day overlap), over each profile up to and excluding day d. The collection $\mathcal{P}$ is noted as:

$$\mathcal{P} = \{\mathcal{S}_8^1, \ldots, \mathcal{S}_d^1, \ldots, \mathcal{S}_8^j, \ldots, \mathcal{S}_d^j, \ldots, \mathcal{S}_8^n, \ldots, \mathcal{S}_d^n\}. \tag{3.16}$$

The set of seven days prior to the day to be predicted $\mathcal{S}_d^j$ is then compared to the collection $\mathcal{P}$, and the k most similar sets are selected using the k-nearest neighbours algorithm [33], employing the Euclidean distance metric. Note that all the historical consumption of household j is included in $\mathcal{P}$ and used to make predictions. One could also choose to exclude household j from the collection $\mathcal{P}$, assuming that only seven days of historical consumption of the household to be predicted are available.

3.3.4 Prediction

One hundred empirical quantiles [68] are computed per time step to be predicted, based on the last day of the selected k-nearest neighbors. The method is based on successful examples of use in weather and price forecasting techniques [48,101]. For each predicted day d, of household j, the output of the model is a matrix of T vectors, i.e., one for each of the T time steps:

$$\hat{\mathcal{Q}}_d^j = \left[\hat{\boldsymbol{Q}}_{\boldsymbol{d_1}}^{\boldsymbol{j}}, \hat{\boldsymbol{Q}}_{\boldsymbol{d_2}}^{\boldsymbol{j}}, \ldots, \hat{\boldsymbol{Q}}_{\boldsymbol{d_T}}^{\boldsymbol{j}}\right] \in \mathbb{R}^{100 \times T}, \tag{3.17}$$

and each vector $\hat{\boldsymbol{Q}}_{\boldsymbol{d_t}}^{\boldsymbol{j}}$ contains the 100 predicted quantiles of household j, on day d, at time step t:

$$\hat{\boldsymbol{Q}}_{\boldsymbol{d_t}}^{\boldsymbol{j}} = \left[\hat{Q}_{d_t}^j(0.01), \hat{Q}_{d_t}^j(0.02), \ldots, \hat{Q}_{d_t}^j(1)\right]^\intercal \in \mathbb{R}^{100}. \tag{3.18}$$

3.4 Case Study

3.4.1 Problem statement

The proposed method produces household electricity consumption forecasts, in a probabilistic manner, i.e., the forecasts consist of 100 quantiles per time step for a daily horizon. Only seven days of historical consumption of the target household is necessary, and no other exogenous data is required such as household's attribute and weather information. The other profiles in the dataset are leveraged to make predictions.

3.4.2 Datasets

Irish Open Dataset

The first dataset used in this study is sourced from the Commission for Energy Regulation (CER) of Ireland and is publicly available on the Irish Social Science Data Archive (ISSDA) website [32]. The data was collected as part of the Smart Metering Project – Electricity Customer Behavior Trial, and comprises half-hourly ($T = 48$) electrical consumption measurements (in kWh) of over 5,000 Irish homes and businesses, spanning from July 2009 to December 2010. To ensure consistency in the benchmarking process, the preprocessing pipeline outlined in a previously published study [11, 12] is employed. As described in the cited study, non-residential profiles are removed, and only complete records are retained, as the focus of the model is on the stochastic behavior of residential customers. A subset of the remaining profiles is then selected randomly, specifically 10% of the dataset, equivalent to 363 customers.

British Open Dataset

The second dataset is made available by the UK Power Networks [133]. During a four-year innovation project, called Low Carbon London, they recorded the half-hourly ($T = 48$) electricity consumption of 5567 London households between November 2011 and February 2014. A representative sample of the Greater London population was recruited as participants. The data set includes two groups of customers: a subgroup of approximately 1100 individuals who were exposed to Dynamic Time of Use (dToU) energy prices during the 2013 calendar year, and a larger group of approximately 4500 individuals whose energy consumption was not subject to the dToU tariff. The dToU prices were communicated to the subgroup on a day-ahead basis through the Smart Meter In-Home Display or via text message. The remaining customers were on a flat rate tariff. We selected a random sample of 363 profiles out of the flat rate tariff customers, in order to have the same amount of profiles as in the previous dataset. Very simple preprocessing is applied: removing duplicate data and linear interpolation of missing values (only 0.14% of the data is missing).

Private Belgian Dataset

This private dataset is provided by Fluvius, the Flemish distribution system operator. They collected the electricity consumption of more than 3000 households in Flanders, for a year, between 2014 and 2016. For this application,

363 one-year profiles are randomly selected, as in the two previous datasets. Each profile consists of the quarter-hourly ($T = 96$) electrical consumption (in kWh). It should be noted that before being usable, this dataset required the most cleaning, since it is used for the first time. The reader is referred to [109] for more details on data cleaning and data quality of a sample of 100 household profiles, available publicly [44] from the same distribution system operator.

3.4.3 Results & discussion

The performances of the proposed approach and benchmarks on the CER, LCL, and Fluvius datasets are summarized in Tables 3.1a, 3.1b, and 3.1c, respectively. Additionally, a visual representation of the tables can be found in Figure 3.5. It is important to note that the magnitude of the CRPS values is influenced by the scale of the respective dataset, and that a smaller value of CRPS indicates a better performance of the algorithm. The percentage improvement compared to the baseline is also given.

For the CER dataset (Table 3.1a), our proposed global approach achieved a CRPS of 0.231002, representing a percentage improvement of 31.82% compared to the baseline. This outperformed all other methods, such as historical sampling and quantile regression, which showed slightly lower performances. Notably, the proposed approach demonstrated substantial superiority over Bernstein Normalizing Flows and the Feedforward Neural Network + Empirical Quantiles, with percentage improvements of 27.55% and 25.89% respectively.

In the case of the LCL dataset, Table 3.1b, our proposed global approach achieved a CRPS of 0.091197, resulting in a percentage improvement of 27.23% compared to the baseline. The method outperformed all other approaches such as quantile regression averaging, advanced quantile regression averaging, and historical sampling, which achieved CRPS values of 0.096869, 0.096869, and 0.098104, respectively. Additionally, the proposed approach demonstrated better performance compared to quantile regression and Feedforward Neural Network + Empirical Quantiles, with percentage improvements of 21.34% and 20.25%, respectively.

Lastly, on the Fluvius dataset, Table 3.1c, the proposed global approach achieved a CRPS of 0.030159, demonstrating a significant percentage improvement of 29.86% compared to the naive baseline. This method also outperformed the Feedforward Neural Network + Empirical Quantiles. Other approaches, including advanced quantile regression averaging, quantile regression averaging, and historical sampling showed superior performance.

The variations in performance observed on the Fluvius dataset may be attributed

to the dataset higher time resolution compared to the CER and LCL datasets, quarter hourly versus half-hourly. This increased frequency poses challenges for prediction tasks, as it becomes more difficult to accurately capture burst-like patterns and larger fluctuations in consumption because of the shorter timescale. The aggregation of data over longer time intervals in the CER and LCL datasets tends to smooth out these variations [70]. Another contributing factor is the forecast horizon. Studies have indicated that the forecast error tends to increase with the prediction horizon [57]. In the case of the Fluvius dataset, one-day ahead predictions require forecasting 96 data points, while the CER and LCL datasets only require 48 data points for the same one-day-ahead predictions. Furthermore, the Fluvius dataset stands out as the only one with negative values, indicating the injection of excess solar energy into the grid. Moreover, the authors acknowledge the comparatively lower data quality of the Fluvius dataset in comparison to the open CER and LCL datasets. These factors, namely the higher time resolution (and the subsequent more frequent bursts), the longer forecast horizon, the negative values and the lower data quality, likely contribute to the observed differences in performance between the Fluvius dataset and the CER and LCL datasets. Understanding these challenges can inform future research and methodological improvements for short-term forecasting tasks in datasets with varying time resolutions and forecast horizons.

Figure 3.6 displays three scatter plots illustrating the individual Continuous Ranked Probability Score (CRPS) values per household, comparing the global approach with the second (or first) best approach for each dataset. In the first sub-figure, the proposed global approach is compared with the historical sampling method using the CER dataset. The second sub-figure presents the comparison between the global approach and the classic quantile regression averaging technique applied to the LCL dataset. In the third sub-figure, the global approach is compared against the advanced quantile regression approach using the Fluvius dataset. The diagonal line within each scatter plot represents the CRPS values at which both methods yield equal performance. As mentioned in Section 3.2.3, the lower the CRPS, the better the prediction. The data points in the top triangle of the plot, in a dark color, represent the households where the global approach outperforms the alternative method. The data points in the bottom triangle, represented in a light color, indicate the households where the alternative method surpasses the global approach. Notably, across all three datasets, the majority of values closely align with the diagonal line, with no outliers observed in the light-colored data points. However, the dark-colored data points exhibits significant outliers. Consequently, we conclude that the two methods either perform equally well or exhibit comparable performance. However, in cases where there is a substantial difference in performance, the proposed global approach consistently outperforms the alternative method. The global approach manages to avoid outliers with very faulty forecast, and hence

is more robust.

The analysis and visualization of specific households in the scatter plots reveals valuable insights regarding the performance comparison between the global approach and the alternative method. Figure 3.7 provides a selected representation of households to support the argument. The observations can be extended to encompass a larger number of profiles, which cannot be included due to space constraints. Differences in the CRPS were computed on a per-household basis, indicating whether the global approach outperformed (positive difference) or the alternative method outperformed (negative difference). Several observations were made based on these differences. Firstly, across all datasets (CER, LCL, and Fluvius), households with the largest positive differences were characterized by significant changes in consumption between the training and test sets, i.e., profiles experiencing concept drift. A similar conclusion is made in [54], about a global point forecasting method, compared to local point forecasting benchmarks. Secondly, it was noted that the average negative difference was smaller than the average positive difference, indicating that when the alternative method outperformed the global approach, the difference was not as pronounced, whereas when the global approach outperformed the alternative method, the difference was more substantial. Lastly, there was no discernible pattern observed among households with the largest negative differences.

3.4.4 Hyperparameter sensitivity study of the proposed method

Table 3.2 presents a comparison of the proposed global approach's accuracy on Commission for Energy Regulation (CER) of Ireland dataset, London Low Carbon (LLC) dataset, and Fluvius dataset. The evaluation focuses on varying the two method's parameters: the normalization type and the number of nearest neighbors. The results reveal that the optimal parameter settings for achieving the best performance differ across datasets. This suggests that the inherent dynamics of each dataset influence the effectiveness of the approach. Furthermore, an analysis of the maximum improvement observed between the "worst" and "best" parameter combinations indicates an improvement of up to 8%. This highlights the non-negligible impact of selecting appropriate parameters on the model's performance.

Method	CRPS (CER Dataset)	Percentage improvement
Our Proposed Global Approach	0.231002	31.82%
Historical Sampling	0.232272	31.44%
Quantile Regression	0.233792	30.99%
Bernstein Normalizing Flows	0.245456	27.55%
Feed-forward Neural Network + Empirical Quantiles	0.251072	25.89%
Advanced Quantile Regression Averaging	0.254295	24.95%
Quantile Regression Averaging	0.254306	24.94%
Baseline	0.338826	

(a) Results on the Commission for Energy Regulation (CER) of Ireland open dataset.

Method	CRPS (LCL Dataset)	Percentage improvement
Our Proposed Global Approach	0.091197	27.23%
Quantile Regression Averaging	0.096869	22.70%
Advanced Quantile Regression Averaging	0.096869	22.70%
Historical Sampling	0.098104	21.72%
Quantile Regression	0.098567	21.34%
Feed-forward Neural Network + Empirical Quantiles	0.099935	20.25%
Baseline	0.125323	

(b) Results on the London Low Carbon (LLC) open dataset.

Method	CRPS (Fluvius Dataset)	Percentage improvement
Advanced Quantile Regression Averaging	0.028564	33.57%
Quantile Regression Averaging	0.028564	33.57%
Historical Sampling	0.028786	33.06%
Quantile Regression	0.029166	32.17%
Our Proposed Global Approach	0.030159	29.86%
Feed-forward Neural Network + Empirical Quantiles	0.030266	29.61%
Naive baseline	0.043003	

(c) Results on the Belgian Fluvius private dataset.

Table 3.1. A comparison of the benchmark methods, i.e., a naive baseline, historical sampling, a feedforward neural network enhanced with empirical quantiles [19], quantile regression averaging, quantile regression and Bernstein polynomial flows [13] on (a) the Commission for Energy Regulation (CER) of Ireland open dataset, (b) London Low Carbon (LCL) open dataset and (c) the Belgian Fluvius private dataset. The method proposed in this paper is highlighted. CRPS stands for Continuous Ranked Probability Score, smaller values indicate a higher accuracy of the model.

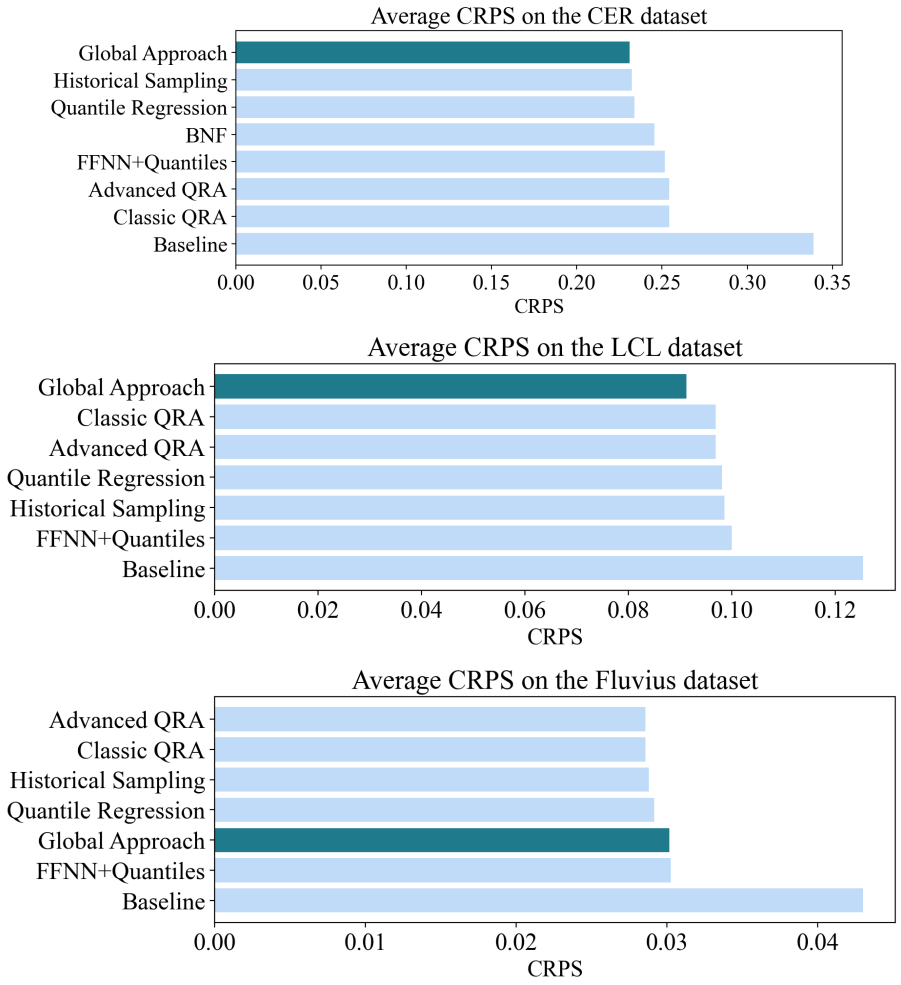


Figure 3.5. A visual comparison of the benchmark methods presented in Tables 3.1a, 3.1b and 3.1c, i.e., a naive baseline, historical sampling, a feedforward neural network enhanced with empirical quantiles [19], quantile regression averaging, quantile regression and Bernstein polynomial flows [13] on the Commission for Energy Regulation (CER) of Ireland open dataset, the London Low Carbon (LCL) open dataset and the Belgian Fluvius private dataset. CRPS stands for Continuous Ranked Probability Score, smaller values indicate a higher accuracy of the model. The proposed global approach is highlighted in a darker color.

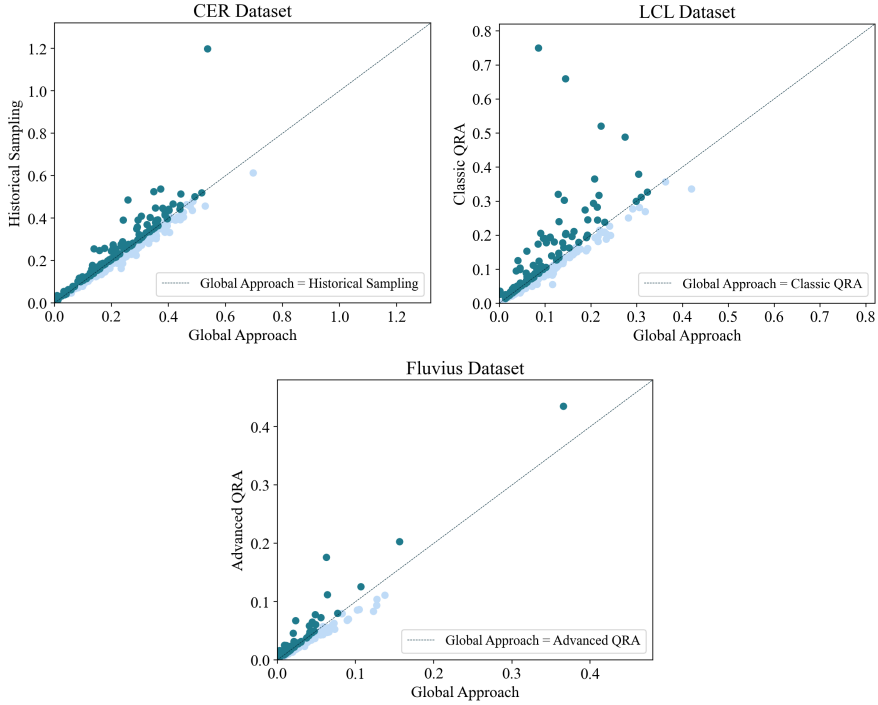


Figure 3.6. Each scatter plots compare the global approach with the second (or first) best approach for each dataset. Each dot represents one household, the x-value is the individual household's CRPS when the predictions for that household are made using the global approach, the y-value is the individual household's CRPS when the predictions for that household are made using the alternative second (or first) best approach. The first sub-figure compares the proposed global approach with historical sampling on the CER dataset. The second sub-figure shows the comparison between the global approach and classic quantile regression averaging on the LCL dataset. The third sub-figure compares the global approach with advanced quantile regression on the Fluvius dataset. The diagonal line in each scatter plot represents equal performance for both methods. The dark blue colored data points in the top triangle indicate households where the global approach outperforms the alternative method, while light blue colored data points in the bottom triangle represent profiles where the alternative method performs better. Notably, most values align closely with the diagonal line across all datasets, with no outliers in the light-colored data points. However, significant outliers exist in the dark-colored data points. These findings suggest that the two methods generally perform equally well or comparably, but the proposed global approach consistently outperforms the alternative method in cases of outlier households.

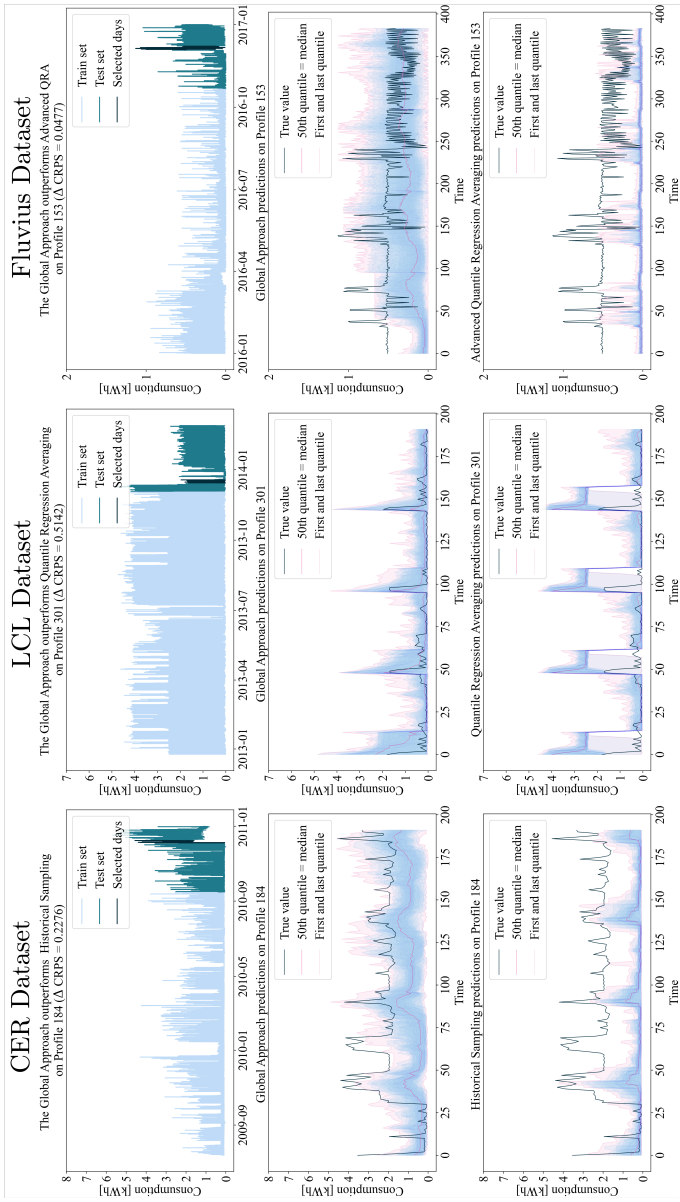


Figure 3.7. The figure showcases an analysis of one profile from each dataset CER, LCL, and Fluvius. These profiles are instances where the global approach significantly outperforms the alternative method. The top row of subplots provides an overview of the complete profiles, where the light blue colored segments represent the training set, the dark blue colored segments the test set, and the black colored segment the part of the test set for which the predictions are shown in the two bottom graphs. The first prediction stems from the global approach, while the second prediction reflects the alternative approach. The true values are in black. The lowest and highest quantiles are in pink, as well as the median. The intermediate quantiles are in blue. It becomes apparent that the global approach demonstrates superior adaptability to changes in consumption patterns, i.e., profiles experiencing concept drift.

Normalization	Number of nearest neighbours	CRPS (CER Dataset)	Percentage improvement
Per profile	100	0.231002	8.88%
Per week	100	0.238436	5.95%
Per day	100	0.239307	5.60%
No normalization	100	0.242696	4.27%
Per profile	50	0.250998	0.99%
Per profile	200	0.253509	

(a) Results on the Commission for Energy Regulation (CER) of Ireland open dataset.

Normalization	Number of nearest neighbours	CRPS (LCL Dataset)	Percentage improvement
Per week	100	0.09119	4.05%
Per week	50	0.09145	3.78%
Per week	200	0.09179	3.42%
No normalization	100	0.09376	1.35%
Per profile	100	0.09427	0.81%
Per day	100	0.09504	

(b) Results on the London Low Carbon (LLC) open dataset.

Normalization	Number of nearest neighbours	CRPS (Fluvius Dataset)	Percentage improvement
Per week	200	0.030159	5.02%
Per week	100	0.030271	4.67%
Per week	50	0.030498	3.95%
Per day	100	0.030840	2.87%
No normalization	100	0.031285	1.47%
Per profile	100	0.031752	

(c) Results on the Belgian Fluvius private dataset.

Table 3.2. A comparison of the proposed global approach accuracy on the three datasets when varying the normalization type and the number of nearest neighbours k. CRPS stands for Continuous Ranked Probability Score, smaller values indicate a higher accuracy of the model. The percentage improvement is computed in function of the lowest performing combination for each dataset.

3.5 Conclusion

This paper proposes a novel probabilistic global method for forecasting residential electricity consumption based on empirical quantiles. The method does not rely on exogenous data requirements such as weather data or household-specific attributes, simplifying the implementation process. It offers the advantage of generating forecasts needing only seven days of historical data of the target household. This data requirement is relatively low compared to other approaches, making the method applicable to situations where limited historical data is available. The model's scalability is thus enhanced by its low data requirement and low complexity. Moreover, the interpretability of the model predictions is a valuable feature, as they are not derived from a black-box model, enabling stakeholders to understand and trust the forecasted results.

The proposed method demonstrates superior performance compared to state-of-the-art models on profiles experiencing concept drift, as evidenced by the evaluation on multiple datasets. The global approach outperforms all other benchmarks on two of the datasets, the CER, and the LCL dataset. The Fluvius dataset presents unique challenges with its higher time resolution and longer forecast horizon, yet the proposed method still outperforms other methods on the typically challenging and changing profiles, i.e., experiencing concept drift. This highlights the effectiveness of the method in capturing the complex dynamics of household consumption patterns.

Furthermore, this research strives for reproducibility and comparability. The utilization of several open datasets and the sharing of code contribute to making this research accessible and transparent for the scientific community. By providing these resources, researchers can replicate the experiments and build upon the findings, fostering collaboration and further advancements in the field.

In conclusion, this paper presents a significant contribution to the literature by introducing an accurate and global probabilistic method for short-term individual household load forecasting. The method's performance superiority, minimal data requirements, comparison with state-of-the-art approaches, and commitment to reproducibility contribute to advancing the field of probabilistic forecasting and provide a solid foundation for further research and applications in the domain of household electricity consumption forecasting.

Chapter 4

Building consumption reduction through plug load mode detection, forecasting and scheduling

This chapter is published and reprinted with permission from:

Botman, Lola; Lago, Jesus; Fu, Xiaohan; Chia, Keaton; Wolf, Jesse; Kleissl, Jan; De Moor, Bart, "Building consumption reduction through plug load mode detection, forecasting and scheduling," **Submitted** in Applied Energy, November 2023.

Abstract

In an era of increasing energy demands and environmental concerns, optimizing energy consumption within buildings is crucial. Despite the vast improvements in HVAC and lighting systems, plug loads remain an under-studied area for enhancing building efficiency. This paper studies smart plug operating mode detection, specifically discerning between active and idle states, plug-level load forecasting and plug scheduling methodologies. This research leverages a unique dataset from the University of California, San Diego, consisting of readings from over 150 plugs in several buildings for more than a year, notably during the post-Covid era. This dataset is made publicly available. We propose a comprehensive literature review on plug operating mode detection, as well as a novel unsupervised learning approach to identify plug operating modes. A pipeline integrating the detected modes with forecasting and scheduling is developed, aiming at building energy consumption reduction. Our findings offer valuable insights and promising results into energy-efficient building management through smart plugs.

4.1 Introduction

Non-residential buildings account for 18% of the total electricity consumption of the United-States [38, 131]. This combined with the increase use of renewable energies, such as local solar panels, has led recent research to focus on building efficiency and demand flexibility. Strategies are developed to reduce building consumption. More specifically, plug loads in commercial buildings (not associated with lighting; heating, ventilation and air conditioning (HVAC) and water heating) accounted for 47% of the total commercial buildings energy consumption in the United States in 2020 [29,131]. HVAC systems and lighting have been improved and are more commonly monitored [84], however these plug loads must be targeted as well to keep increasing building efficiency. Note that in the literature, these plug loads are sometimes called plug and process loads (PPLs) [38], Miscellaneous Electric Loads (MELs) [72, 74, 85], small power equipment [98], or plug-in equipment loads [55].

This paper addresses smart plug operating mode detection, i.e., active vs. idle, load forecasting, and scheduling. The dataset originates from the University of California, San Diego and includes over 150 plugs distributed across several buildings over more than a year. This research builds upon the plug load management system described in [29]. The objective of this research is to contribute to the understanding of energy-efficient building management through smart plug analysis and to explore the potential for energy savings.

4.2 Motivation and contributions

The utilization of smart plugs to completely disconnect appliances, thus reducing standby energy consumption, presents a promising avenue for enhancing grid efficiency, as standby losses are significant [113]. Despite advancements in the efficiency of lighting and HVAC systems, consumption from unregulated plug loads has increased [74]. This underscores the importance of comprehensive study around miscellaneous and electronic loads as significant contributors to energy consumption. Yet, they are only minimally explored for potential energy-saving interventions, primarily due to data scarcity and the absence of economical technologies [49]. Existing studies emphasize the potential benefits of an open-source dataset comprising plug load measurements [74].

This paper sets several advancements in the study of smart plugs. The key contributions are as follows:

1. An novel open dataset is introduced, encompassing readings from over 150 plugs situated in various office buildings, spanning a duration exceeding a year, with specific emphasis on the recent post-Covid era.
2. The first literature review on methodologies for plug operating mode detection is presented, coupled with a compilation of terminologies that have been employed in prior research, to frame this challenge.

3. The paper proposes an unsupervised learning approach aimed at discerning the operating modes of a plug (or the attached device), addressing the challenges documented in existing literature.
4. The challenge of individual plug load forecasting is tackled for the first time.

Finally, a methodological pipeline is outlined, combining the detected modes with individual plug load forecasting techniques, designed to schedule plugs (on/off) a day in advance, with the overarching aim of decreasing building energy consumption.

4.3 Literature review

4.3.1 Smart plug hardware

The advent of smart plugs has led to significant advancements in both hardware and software research related to energy consumption and management. Smart plugs enable fine-grained monitoring and control of plug loads through Wi-Fi or cellular network [113]. Plug-level data is becoming more accessible, mainly thanks to these advances in hardware. Different plug and telecommunication solutions were proposed. For instance, a hardware plug that turns off if the consumption is below a user-defined threshold [61]. Another example is the smart plug equipped with a wireless Zigbee sensor, primarily designed to measure power consumption of electrical appliances in home energy management systems [6]. Another prototype for home units that can recognize device types based on statistics and labeled data, was described in [100].

4.3.2 Smart plug home vs. office set-up and integration

The smart plug research field mostly started by focusing on home energy management systems [1,7,67,86,117,118]. The installation of smart plugs in homes is easier, requires fewer devices and thus less investment, and is easier to maintain. The integration of these smart plugs into building systems however, presents its share of challenges. Chia et al. [29] discuss these challenges comprehensively, shedding light on data access, control, and integration within commercial buildings. The availability of such data remains limited, and the infrastructure for seamless integration is still evolving.

4.3.3 Smart plug open data

Comprehensive datasets are crucial for advancing research in the dynamic field of smart plug load. Making such datasets openly accessible not only fosters progress but also facilitates benchmarking. A recent overview of public datasets for home management systems and typical household appliances recordings is provided in [99]. While some papers have presented results without sharing their datasets, others have commendably

made them available in recent years. In 2013, Lanzisera et al. offered insights from a year-long study that meticulously inventoried over 4000 devices. With 455 plug load meters, they managed to monitor the MELs in an office building. However, this extensive data hasn't been made publicly accessible [85]. Likewise, in 2015, [73] analyses the characteristics that a competent open dataset should possess. While it has been described in detail, spanning devices like Desktops and Multi-Functioning Printing Devices, it remains unpublished. The authors have stated their intention to release it post a comprehensive analysis. To this date, this dataset has not been made available. To the best of the author's knowledge, the first smart plug open data set 'BLOND', made available in 2018, spans over 213 days with voltage and current recordings from 53 distinct appliances at 50 kSps (kilo Sample per second). Moreover, the dataset is taking 38.7 TB with nearly a million files [81]. In 2019, Doherty et al. published a dataset detailing 14 different plugs. The data collection spanned three months from October to December 2017, with power consumption recordings made every 5 minutes. Among the items tracked were common office appliances like Coffee Makers, Laptops, Microwaves, and Video Conference Cameras [38]. More recently, a three-week-long dataset from February to March 2020, recording data every minute was made available. This dataset incorporates information from 88 plugs, which include a diverse range of devices: 31 laptops, 9 desktops, 35 monitors, 13 fans, and 11 task lamps [127]. An open Matlab formatted dataset was also made available in 2020, extends its study beyond just smart plug data. It incorporates readings from 19 sensors situated in an office space. These sensors range from SmartSense Motion Sensors to Temperature-Humidity Sensors, Ultrasonic Sensors, and even Belkin Wemo Switches [63].

4.3.4 Non-invasive and invasive load monitoring

There are two main different approaches used to study building energy efficiency. **Non-Invasive Load Monitoring** (NILM) consists of disaggregation strategies, seeking to break down a building global energy consumption in the individual device consumption without the need for extensive hardware installations [99], requiring only one smart meter at the building level. A notable disaggregated model of an office building's plug loads was introduced by Doherty et al. [38]. This method uses power data from a few monitored devices combined with a detailed device inventory. The derived model provides a daily profile of the averaged metered load and creates models for non-measured loads based on logical assumptions and technical specifications. Crucially, these insights, when coupled with the disaggregated model, help identify devices that can be optimized for energy savings. Though NILM offers the advantage of reduced hardware requirements, making it more cost-effective and easier to maintain, it does have its challenges. Primarily, while it's been popular in the residential sector since the 1990s, there are concerns about its accuracy, especially in the commercial sector. Larger spaces tend to have a diverse range of devices, often with multiple instances of the same device, making them hard to distinguish. Skepticism persists regarding its precision in such settings [38]. However, it's worth noting that Lanzisera et al. demonstrated that by taking an inventory of just 40% of their building floor space,

they achieved less than a 10% error in their estimates, especially for computers and displays [85].

On the other hand, **Invasive Load Monitoring** (ILM), also termed component-based load monitoring or distributed sensing, operates by monitoring and recording the electrical consumption data –or voltage and current– of each individual device or appliance distinctly. It provides accurate disaggregation, particularly in commercial settings. The capability of smart plug technology in ILM is well-acknowledged. Yet, ILM's major challenge remains its cost, given the hardware and labor demands to monitor each device in a building individually [38, 99].

4.3.5 Scheduling smart plugs

Different smart plug control strategies can be used to optimize energy consumption in commercial spaces. Tekler et al. [127] discuss a range of these strategies: time-based (also called schedule-based), occupancy-based, and system-based controls. Timer-based controls are designed to deactivate plug loads according to a set schedule, typically during periods like non-office hours. If occupants have variable schedules, occupancy-based controls can be more dynamic, by determining their presence within a designated area and deactivating plug loads after prolonged absence. This aligns more closely with unpredictable occupancy patterns. System-based controls are components of a comprehensive building management mechanism. Beyond these established control strategies innovative hybrid automated strategies have been developed. Teckler at al. [127] propose the utilization of high-fidelity occupancy detection methods to trace the users' movements, combined with a plug load identification model. Through an interface, users can define their preferred settings per plug type. This ensures minimized energy consumption and alignment with the user's distinct habits. Hafer et al. [59] combined energy monitoring gamification on a mobile platform with plug load control schedules, resulting in a 21% reduction in plug loads across three buildings. Another innovative approach is seen in Choi et al. [30], where they introduced a location-aware energy-saving service. This service utilized smartphone Bluetooth connections to local beacons to determine the location and turn devices on or off accordingly, yielding energy savings of approximately 31.9% for PCs and monitors, and about 15.3% for lights. Ghatikar et al. [49] concentrated on plug load energy efficiency in IT office buildings. Smart plugs adjusted the schedule based on several factors: timing, usage, and company policies. Notably, their system had two distinct mechanisms. One was time-based, where devices operated during specified hours, and the other was usage-based, shutting off after reaching a set energy consumption limit. Moreover, they proposed a third mechanism, a master/slave-type control, where primary devices, e.g., PCs, would control the power state of their associated devices, such as monitors or speakers. The National Renewable Energy Laboratory report [131] offers a different perspective, recommending devices to switch to standby mode. This could either be a manual trigger or embedded in the device's hardware settings. For computers, triggers could be the power button, designated keyboard buttons, or even

an icon on the desktop. Multifunctional devices, like printers and scanners, should ideally shift to standby after 15 minutes of inactivity.

4.3.6 Forecasting plug loads

There are numerous forecasting techniques and papers about electricity consumption time series, designed to predict energy use at varying scales—from national to household levels. However, these broader approaches fall outside the scope of this paper and its accompanying literature review. We are specifically interested in individual plug load level. The literature appears sparse when it comes to forecasting at the finer granularity of individual plug or appliance levels, a niche that poses unique challenges. These challenges arise from the inherent stochasticity of appliance use, largely driven by unpredictable human behavior. Additionally, the data is frequently unbalanced, characterized by prolonged periods of low consumption values (indicating idle or off modes) with occasional higher consumption when the appliance is in use. Furthermore, any model designed for such forecasting must be inherently scalable, and not too complex, due to the potential need to predict usage across a multitude of plugs. The current literature landscape offers little insight into individual plug load forecasting, also underscored by Wang et al. [145]. However, some research has been identified that addresses aggregate plug load forecasting, whether at the room, zone, or building level, and these works will be the focus of this section.

Gunay et al. [55] constructed a model that splits plug load patterns across five distinct time frames: occupancy periods, intermediate breaks, weekday evenings, weekends, and vacations. By integrating the predictions of an occupancy model, the research demonstrates an adeptness in generating precise plug load forecasts at the room levels. Wang et al. [145] focus on HVAC predictive control, underscoring the necessity of plug load forecasts to estimate this source of internal heat gains within office environments. Their research findings highlight the significant role of occupant count in predicting building level plug load. They develop a model for an 8-hour horizon forecasting, utilizing features such as the last 24-hour consumption, the day of the week, hour of the day, and hourly occupancy data. They find that a Long Short Term Memory network outperform the ARIMA model by 4% and emphasized the impact of including occupancy data. Srinivasan et al. [125] discuss plug load forecasting, leaving certain specifics such as the forecasting horizon and the exact consumption entity (whether individual device or aggregate) undefined. Their work propose the Seasonal ARIMA model and show an example of a printer's energy consumption over a designated 4-month span. Abdullah et al. [1] also discuss forecasting, however they used an artificial neural network controller for appliance management, predicting the on or off states based on customer preferences and appliance priorities rather than focusing on individual plug load forecasting based on historical data. Malais et al. [94] work with residential data, comparing various data-driven forecasting models to predict aggregated lighting and plug load electricity demands in dwellings. While the gradient boosted regression trees model emerged as the best performing, the distinction between

the best models remained nuanced and dependent upon specific residential units and time horizons.

4.3.7 Plug operating modes detection terminology

The task of discerning the operating mode of an electrical device, specifically determining if it is active, standby, idle, or off, is an essential aspect of all forecasting and scheduling applications, mentioned in the previous section. Surprisingly, research addressing this topic specifically remains scarce in the literature. A notable challenge faced during the literature review is the lack of standardization in terminology. This absence of a consistent vocabulary not only complicates the review process but also makes comparisons between different research difficult. To palliate this lack of standardization this section provides an exhaustive list of the different terminology used during a span of over 20 years, in 19 papers which are addressing implicitly the problem of operating mode detection. In the context of smart plugs and related studies, the **procedure of operating mode detection**, as called in this paper, is expressed through: threshold detection, mode detection, status monitoring, load state estimation, recognition of appliances states or activity recognition.

Furthermore, **operating modes** themselves, as they are named in this paper, are referred to by multiple terms: mode of operation, operating patterns, operational characteristics, operation mode, operating state, operational energy use, power state, power status, power management modes, power levels, appliance status or appliance activity.

Lastly, when naming the **categories of operating modes**, there's a plethora of terms used both in academic research and industry practices. Some papers provide explicit definitions [59,76] while most studies assume a general understanding even though this understanding is different in each of them. As a generalization, and common ground between existing studies, four modes can be defined, expanding on [59] definitions:

(1) The mode off or energy zero represents a device that is completely turned off, consuming no power.

(2) The next mode is typically defined as a state the device automatically enters after prolonged inactivity [98]. It is known by terms like low-power, passive, suspend, sleep, asleep, standby, or even the descriptive "turned off but still consuming power". This is also sometimes called the vampire current.

(3) The following mode is denoted as idle or ready, it signifies a device that's turned on, not in use but ready for activation, consuming more power than the previous mode.

(4) The terms on, high-power, active, in use, and running are used interchangeably to denote this mode, representing the highest power consumption level of a device.

It should again be noted that there are varied conceptualizations across studies. Some consider only two states, off and a conglomerate of (2), (3) and (4). Others might recognize three distinct states off, a combination of (2) and (3) and on.

4.3.8 Plug operating modes detection methodologies

Approaches currently adopted to discern plug operating modes, applicable in both residential and commercial settings, span across a range of methods. It should be noted that operating modes detection is never the main research topic of a research paper, it is usually a byproduct of a broader study, or a step in a smart plug application such as plug scheduling. Four approaches for plug operating modes detection have been identified when reviewing the existing literature. Most of the studies focus on empirical techniques, like visualizing measurements and determining a standby/active threshold, or chose a threshold value based on measurements mean. Some researchers use literature benchmarks and refer to appliances specification sheets. Device users' feedback is sometimes used to qualitatively estimate which percentage of the time a device is on or off, necessitating questionnaire completion or interview of users. Finally, some papers rely on labeling a dataset manually and then applying supervised learning techniques. This section highlights each of these approaches along with the relevant papers.

Empirical techniques

Leonardi et al. [86] propose an innovative solution using machine learning to streamline metadata management for smart homes. While the study's core focus is on identifying swapped devices and newly connected devices on the network, it acknowledges the necessity of threshold identification for operating mode detection. They decide to empirically set the threshold at 15 W. Further analysis unveil that filtering out standby values improved device identification accuracy. Ambati et al. [10] propose "AutoPlug", an automated system adept at tracking and identifying devices plugged into smart outlets in real-time, without necessitating user intervention. This system presumes a smart building environment with interconnected smart outlets transmitting power consumption data in real-time. To segregate active from inactive periods, the researchers empirically establish a 5 W threshold. Tekler et al. [127] observe that many plug loads occasionally utilize a slight amount of standby power, approximately 2.3 W, when not in active use. Hence, the researchers designate a power threshold of 2.5 W to accurately detect active periods. Wang et al. [141] focus on home system management. They start by differentiating plugs as either continuous, e.g. a fridge, or intermittent, e.g. a toaster, power consumers. This categorization seems to have been done manually. They then determine the active vs. idle states based on the effective state ratio (ESR) that depends on the type of power consumption of the device (continuous or intermittent) and its measured mean consumption value. Chia et al. [29] adopt a visual method by examining power recordings of plug loads. Their empirical findings indicate that the majority of devices have distinct active use peaks

during business hours, with standby powers ranging from 2 W to 60 W. Siebert et al. [120] are unique in their approach by proposing both centralized and decentralized methodologies for scheduling algorithms. While the primary focus of the paper is not on ascertaining standby values, the determination of such values is crucial, given the study objective to mitigate standby consumption, also termed "vampire current". In the centralized approach, a home automation controller aggregates data from smart plugs, the smart meter, and utility information, like tariff tiers. The centralized algorithm initially determines the standby energy of devices by computing the daily average consumption for each smart plug. Subsequently, a threshold is set to deduce standby consumption. In essence, a plug is deemed in standby mode when its power lies between a lower and an upper threshold, defined respectively as a given percentage of the mean consumption and an upper limit. However, the case study section suggests that these parameters, are respectively set to 75% and 50 W empirically without detailed justification. The decentralized method proposes an in-built scheduling algorithm within the smart plug itself. Given the constrained processing capabilities of smart plug microprocessors, this methodology was designed for simplicity. To deduce standby consumption, the algorithm evaluates if the peak demand during a given assessment period is at least double the minimum. If not, the load is considered constant, implying no standby consumption. However, if this criterion is not met, the algorithm delineates the range between minimum and maximum values into ten segments. It then examines, starting from the minimum, whether the cumulative measures exceed a reference value (RV) which is a fraction of the total count. If this condition holds true, the standby threshold is recognized as the value of the last segment reviewed. Yet, similar to the centralized method, the paper's case study indicates an empirical selection of parameters, with RV set at 25% without an elaborate rationale.

Benchmarks and technical specification sheets

In a study by Kawamoto et al. [76], three distinct operating modes are explored: active, low-power, and off. The power consumption data for these modes is derived from the authors' unpublished measurements (empirically) or literature benchmarks. For servers specifically, measurements reflected power demands ranging between 50 W and 270 W, with an estimated average of 75 W. An interesting observation is made regarding copiers and laser printers. Their power levels vary considerably with the speed (measured in images per minute) of the device. Doherty et al. [38] center their study around the development of a disaggregated model to evaluate plug loads in office buildings. The novelty is the utilization of power data from a limited number of monitored devices combined with a device inventory. The methodology to ascertain standby load is twofold: direct measurements of specific devices in standby, for instance, observing the power consumption of an elliptical trainer for 30 minutes, or extracting the required data from specification sheets. Menezes et al. [98] focus on the creation of two models designed to estimate minor power consumption in office buildings, complemented by typical power demand profiles. The paper leans heavily on data obtained from published benchmarks, especially in the context of Energy

Star-rated devices, for which data can be sourced from their online database. For refurbished devices or specific appliances, short-term monitoring is deemed adequate to provide better input data. In their findings, they present a table outlining published energy requirements for both desktop and laptop computers in low-power and active modes. Additionally, another table captures the power levels of other equipment, i.e., computers, printers, copiers and screens, constructed after a thorough audit of their installed equipment.

User feedback

In the study by Ghatikar et al. [49], the primary focus is on the energy efficiency of miscellaneous electrical loads in IT office buildings, specifically in India and the U.S. using smart plugs. These smart plugs play a pivotal role in managing the operational schedules of MELs, drawing upon preset user inputs. Such inputs often hinge on factors like timing, usage patterns, and organizational guidelines. The researchers also suggest a control framework based on a master/slave model. In this setup, the power states of principal devices, such as personal computers, have an influence on the power states of their logically associated devices, like speakers, monitors, and printers. However, it's worth noting the system's dependency on persistent user engagement, along with the requisite infrastructure in terms of reliable communication technologies and hardware. In a study by Hafer et al. [59], assumptions about the duration spent in distinct operating modes were majorly extracted from user feedback gathered during preliminary phases of the research. The authors of this study mention that given the largely qualitative nature of these assumptions, there's scope for enhancement. Potential future studies could aim to refine these estimations by collating more rigorous data about the periods spent in each power state, with an emphasis on devices that see prevalent use.

Labeling data

Webber et al. [146] collected data on turn-off rates for an array of office equipment, including but not limited to computers, monitors, and printers. They visited office spaces and physically observed each device, and noted down its power status, essentially a form of manual labeling. For instance, a laptop's power status was determined by visual indicators, like a screen being on. Similarly, monitors usually manifested their low-power mode via a distinct light, and printers, particularly ink jets, were simply categorized as either "on" or "off." Notably, identifying the power status for copiers proved relatively straightforward. Ruzzelli et al. [111], while primarily focusing on domestic systems, introduce the RECAP system. RECAP is an acronym standing for the RECognition of electrical Appliances and Profiling in real-time. This system consists of three main parts: directing users to catalog electrical devices and establish a unique database of their distinct features, utilizing this database to train a computational model in order to differentiate and understand appliance behaviors, and implementing a load descriptor to facilitate comparisons between

similar appliance performances. The authors meticulously crafted a labeled database of distinct appliance signatures based on six parameters: real power, power factor, peak current, RMS current, peak voltage, RMS voltage. The challenge of operating mode detection (called device activity recognition in [111]), especially with appliances possessing multiple settings, was highlighted. In the work of Lin et al. [87], the primary objective is the identification of appliance states using circuit-level energy consumption. The study leans heavily on manually labeled data. Kalluri et al. [73] emphasized the significance of visual inspection for data labeling. A case in point was the BLUED dataset, wherein appliance activity transitions are manually labeled based on the visual assessment of energy signatures. Specifically, any power level fluctuation exceeding 30 W and persisting for a minimum of 5 seconds is documented. Masoso et al. [95] ventures into examining energy consumption during unoccupied hours, particularly in the distinct climates of Botswana and South Africa. Surprisingly, their research reveals that a substantial portion of energy (56%) is consumed during non-working hours, overshadowing the consumption during active hours (44%). These results are based on a series of detailed energy audits, with data acquired via on-site assessments and subsequent manual labeling.

Several other studies have acknowledged the significance of identifying plug operating modes, though their primary focus is not the detection of these modes. For instance, in the work of Ahmed et al. [5], there is mention of on/off states, but the emphasis is primarily on shifting peak consumption. Kamilaris et al. [74] underscores the importance of discerning operating modes in their conclusions, asserting that it is crucial for enhancing the analysis and understanding of total consumption. However, tangible solutions for this identification are not provided in their study.

Determining plug operating modes, both in domestic and office settings, has been approached through a diverse set of methods, as discussed in the preceding paragraphs. Four predominant methodologies are identified: empirical methods, utilizing benchmarks and specification sheet information, soliciting user feedback, and adopting supervised learning when a manually labeled dataset is available. However, each of these approaches presents its challenges. Empirical methods face challenges in their generalizability. Given the differences in power consumption levels across devices, from computers, to copiers, to coffee machines, crafting a one-size-fits-all model threshold value can be problematic. Moreover, addressing each device individually visually is labor-intensive. Relying on averages or percentiles for decisions can be an issue due to potential imbalances in device usage patterns, risking skewed interpretations. Turning to benchmarks and specification sheets, one is confronted with the challenging task of procuring information for a large number of devices. The information may sometimes be absent from specification sheets. Furthermore, devices, over their lifecycle, might deviate from their initial specifications. Additionally, the swift advancements in technology necessitate constant updates to benchmarks. The dynamic nature of a workplace, where devices are continuously added or replaced, further exacerbates the challenge. User feedback, while valuable, is inherently qualitative. Maintaining its relevancy demands consistent updates, which is both labor-intensive and challenging to scale. The approach of manual labeling followed by supervised learning, though promising, is not without

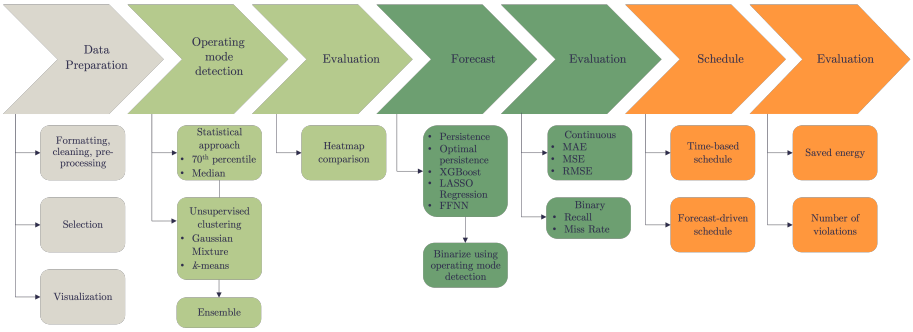


Figure 4.1. Pipeline for plug load scheduling composed of four main steps: (i) Data preparation, (ii) operating mode detection, (iii) plug load forecasting, and (iv) plug load on/off scheduling.

pitfalls. It demands meticulous experimentation and manual state annotations. The scalability issue resurfaces, coupled with potential inaccuracies due to human errors in labeling.

A recurring observation is that all of these methodologies for plug operating mode detection play a secondary role in broader research projects. The significance of plug operating mode detection is frequently understated, even though it can profoundly impact end applications. An additional layer of complexity arises from the lack of standard terminology, leading to potential misinterpretations and ambiguities in research discussions. This literature review underscores the knowledge gap and challenges and accentuates the need for focused research in this domain. The ideal solution would pivot towards unsupervised learning and automatically updating plug operating modes, removing the dependence on labor-intensive manual processes.

4.4 Methodology

We propose a pipeline for office plug load scheduling composed of four main steps, as illustrated in Figure 4.1: (i) Data preparation, (ii) operating mode detection, (iii) plug load forecasting, and (iv) plug load on/off scheduling.

4.4.1 Data preparation

A large data preparation process was necessary to obtain a usable dataset. This is a one-time procedure, and the specific details are presented in Section 4.5.2. In this section, some general guidelines to apply the pipeline in Figure 4.1 on any plug recording dataset are provided.

The dataset is expected to have the same frequency through the entire recording period, in our case it is a quarter hourly resolution. NaNs indicate missing recordings, that can occur due to system updates or failure, or employees unplugging appliances. These can either be filled using any data imputation technique, or left as such and removed during the training and testing. We opted for the latter.

Data exploration and analysis using visualization has been done iteratively, using time series plot, heatmaps, stacked diagrams with a weekly or yearly evolution. For example, heatmaps allowed to identify wrongly assigned device type, or fully zero profiles and to address those. The correct device type was reassigned in the metadata and fully zero profiles were dropped. This visual data exploration is highly recommended to obtain insights on the datasets, and take additional pre-processing steps but is out of scope for this paper.

Another recommended pre-processing step is to select only plugs that have reduced amounts of NaNs (missing values) and zero recordings, since the aim is not to schedule underutilized plugs. We decided to set the threshold to at most 30% of NaNs and 30% of zeros.

A selection based on the plug device type is applied. Our study focuses on scheduling devices such as printers, copiers, and TVs.

4.4.2 Operating mode detection

This section covers the proposed plug operating mode detection procedure. As concluded from the literature review, a recurring observation is that all the existing methodologies for plug operating mode detection play a secondary role in broader research projects. This section addresses the knowledge gap and challenges in plug operating mode detection.

In this study, we focus on two operating modes. The first mode corresponds to when the electrical appliance is active, as denoted by (4) in Section 4.3.7. This state is interchangeably referred to as high-power, on, or running in both academic and industrial contexts. The second mode encompasses states (1), (2), and (3) from the same section. With the aim of minimizing energy consumption in buildings, the final objective of the proposed pipeline is to ensure plug loads are tuned on only when in use, and turned off otherwise. This includes situations when appliances are off but drawing a minimal current, idle, or in standby.

The proposed solution is based on an ensemble of statistical and unsupervised learning methods. The method automatically detects plug operating modes, and removes the dependence on labor-intensive manual processes discussed in Section 4.3.8. Section 4.4.2 details the two statistical approaches, Section 4.4.2 explains the three unsupervised approaches, and finally Section 4.4.2 proposes an ensemble scheme to robustly combine these. Each method, is applied individually on each plug, and only uses the first 80% of the plug load time series. A visual evaluation is performed on the test set, i.e., the remaining 20%.

Statistical approaches

Two statistical approaches are proposed. They each generate a threshold, which is then used to binarize the power values. The power values higher than the computed thresholds are considered as on (1) and values lower than the thresholds are considered as off (0).

The first statistical approach considers the median of all recordings as the threshold. The median is favored over the mean, since it is less sensitive to outliers. A limitation of this approach is that this threshold assumes that an appliance is on average used more often than not.

The second statistical approach tries to address this limitation by selecting a threshold value equal to the 70th percentile of the recording values. The underlying assumption is made that an appliance is typically active during weekdays from 8 am to 6 pm. Mathematically, this implies the plug is active approximately 29.76% of the time. Therefore, the top 30% of values are considered as active, establishing a threshold to differentiate between active and idle states based on the 70th percentile. Percentile also have the advantage that no mathematical assumption about the underlying distribution of the data has to be made.

Unsupervised clustering approaches

Three unsupervised clustering approaches are proposed to classify the power values as the plug being in used (labeled as 1) or not used (labeled as 0), i.e., either in idle, standby or off operating mode. For each approach, the number of clusters is an input parameter and is set equal to two. The same pre-processing is applied in each method: the 0 W recordings and NaN values are dropped, as well as extreme positive outliers. Outliers are identified as values surpassing a cut-off, determined by the mean plus 30 standard deviations. This detection process is applied only when there are fewer than five such outliers. This approach seeks to mitigate potential data recording errors. Observing up to four values significantly deviating from the distribution suggests a potential recording anomaly. It is improbable for a device to exhibit such drastic consumption variations, especially for short 15-minute intervals. We observe that this outlier removal occurs for less than 5% of the plugs. Another purpose of this outlier detection is to address the sensitivity of both k-means and Gaussian mixture models to outliers.

The first unsupervised clustering approach uses the k-means algorithm. The missing values are dropped and k-means clustering, with default parameters, i.e., k-means++ initialization and squared Euclidean distances metrics. This method aims to partition data into distinct, non-overlapping subgroups, or "clusters". This algorithm assumes clusters to be spherical and of comparable size. The squared Euclidean distance renders it sensitive to outliers. Its scalability, especially with large datasets, makes k-means a popular choice across diverse research areas. For more details on k-means, the reader is directed to foundational literature [93, 104].

The second and third unsupervised clustering approaches are based on Gaussian Mixture Models (GMM). One of them uses the entire training set while the other only uses the most recent three months of data, and can be updated over time. Fundamentally, a GMM is a probabilistic model that assumes data points are derived from a mixture of several Gaussian distributions with unknown parameters. In this use case, the number of underlying Gaussians is set to two, since we are aiming to determine active vs. not active modes. The initialization used in both GMM approaches is the same as for the k-means approach. The model achieves convergence through the expectation-maximization (EM) algorithm. In terms of scalability, GMMs are well-suited for datasets of moderate size but may demonstrate slower performance with larger datasets. While k-means strives for clusters of similar spatial shape, algorithms like GMM allow for clusters with various shapes and sizes. The reader is referred to the existing literature on GMM and for theoretical and implementation details [96, 104]

In all methods, we consistently apply three post-processing steps. Initially, we ensure the cluster labeled as "0" corresponds to the values with the lowest average, given that our intent is for the "0" label to represent off, standby and idle mode; and "1" to indicate active mode. Subsequently, power readings of 0 W are labeled as "0", knowing the device is off. Lastly, and mainly relevant for the GMM clustering, values under the median of the "0" cluster but tagged as "1" are reclassified as "0". Conversely, values above the median of the "1" cluster but marked as "0" are re-labeled "1". This avoids the outlier effect, which GMM can sometimes produce.

A limitation of both the k-means and GMM algorithms is that they depend on the initialization step. Further experiments with varying initialization could be done in future work [121].

Ensemble

An ensemble technique [25] is introduced to combine insights from all the aforementioned methods. The ensemble works on a majority principle, with a safety margin, specifically designed for plug scheduling. If none of the five operating mode detection approaches assigned a power value as active, this power value is labeled as unused (0). Similarly, if only one method considered it active, it was still labeled as unused (0). When at least two out of the five methods agree upon active mode, a power value is considered active (1). The ensemble adopts a sort of weighted average, while prioritizing user convenience, keeping the end goal of plug scheduling in mind. It is preferable to label a power recording as active, ensuring the plug remains scheduled on, than to mistakenly label it as inactive and potentially schedule the plug off, while a user might utilize the connected appliance. Note that in the clustering literature, ensembles are sometimes also called consensus clustering.

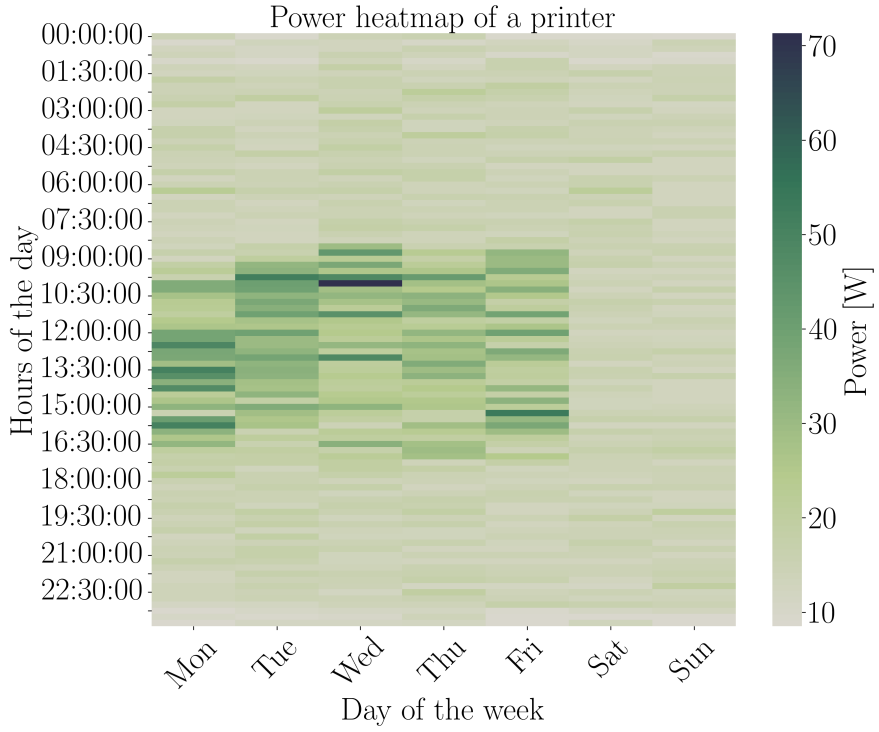


Figure 4.2. Representation of a three-months long time series of plug power values, visualized as a heatmap. The power level is indicated using color intensity, the days of the week are on the x-axis and hours of the day on the y-axis. Each cell represents the average power value for a combination of a day of the week and a time of the day (y-axis), e.g., the average power level on a Wednesday at 10am is around 70 W. A weekly plug usage pattern is clearly visible between Monday and Friday during office hours.

Operating mode detection evaluation

There is no ground truth (or "true threshold") to evaluate the proposed clustering ensemble approach. We suggest assessing the mode detection accuracy using heatmap comparisons. One can visualize the time series of a plug power levels as a heatmap, as depicted in Figure 4.2. A heatmap displays the plug power level using color intensity, with days on the x-axis and hours on the y-axis. This representation captures the weekly plug usage pattern. Two heatmaps of the same plug can be constructed: the original based on the original continuous power levels, while the latter presents the binarized levels, after applying mode detection. The binarized heatmap should mirror the original's weekly usage pattern. This pattern remains crucial as the pipeline end goal is to schedule plugs in line with their individual usage.

Quantifying the similarity between two heatmaps is not straightforward. Out of the many existing image similarity metrics in the literature, we selected three simple metrics: the MSE [80], the RMSE [37], and the Pearson correlation coefficient [53]. We also considered two more advanced and established metrics: the structural similarity index (SSIM) [90, 151] and the Kullback-Leibler (KL) Divergence [80]. However, none reflected the similarity as the human eye perceives it, nor is suited for this particular application. Patterns that align perfectly should receive high similarity scores, and minor deviations in timing and values be acceptable. A significant disparity, such as continuous high power on a Monday in the original data that is not present (marked as active) in the binarized heatmap, should not score highly, even if the distributions are mostly alike. The literature agrees that evaluating image similarity as perceived by human eye is challenging [80, 90]. Consequently, we decided to exclude these metrics and results from the paper, as they are unsuitable for numerical mode detection evaluation and did not provide added value.

Our proposed evaluation method involves visual qualitative heatmap comparisons. We classify the results into three distinct categories based on how well the algorithm captures the underlying usage pattern:

- High-fidelity detection: the binarized heatmap closely mirrors the original heatmap, capturing the usage pattern of the plug power levels. This suggests that the algorithm has successfully detected and represented the true underlying usage pattern.
- Partial-fidelity detection: while the binarized heatmap generally aligns with the original, there are discrepancies. For example, the power value is always marked as 'active' correctly when it is indeed active, but power levels are also sometimes incorrectly labeled as active during some of the inactive periods. This indicates that while the algorithm detects most active patterns, it does not provide a perfect match.
- Low-fidelity detection: the binarized heatmap significantly deviates from the original. This could manifest as the plug being marked as continuously active or continuously inactive despite contrary evidence in the original data, or the algorithm failing to capture the usage pattern at all.

This approach has limitations, such as the inability to quickly assess multiple plugs. A numerical metric would mitigate any potential human bias and would allow to accurately assess changes in the algorithm as improvements or not. Further research in that direction, to find, or design, an appropriate similarity measure with the right properties for this use case is out of scope for this paper.

4.4.3 Forecasting

The forecasting problem, highlighted in dark green in Figure 4.1, is characterized by two dimensions: spatial aggregation and horizon. Spatial aggregation is the level of load being predicted. In this context, we focus on forecasting individual plug loads.

Regarding the horizon, the model produces forecasts for the upcoming day, with data recorded in quarter-hourly intervals, this yields 96 predictions. This section elaborates on several implemented forecasting models based on diverse approaches: a naive baseline, an enhanced naive approach, a linear regression, and finally a more advanced machine learning approach. It is good practice to clarify all forecast time parameters [148]. In this paper the forecast update rate is 24 h (daily at midnight), the forecast resolution is 15 min, the forecast horizon is 24 hours, and the forecast lead time is 0 h. An additional challenge encountered within the dataset is its intermittent nature. In the literature, intermittency is defined as prolonged sequences of consecutive zero values with brief periods of higher values. The domain of intermittent forecasting is predominantly explored within economic contexts, such as sales predictions for inventory systems [23, 60]. In this study, the data points do not always represent zero power levels; rather, they reflect prolonged lower power levels when the device is idle yet powered on, and higher power levels during shorter active usage periods. This aspect adds complexity to the forecasting task.

Persistence

A persistence model serves as a naive baseline, where no modeling is involved. In the absence of complex training, the most straightforward assumption is that today's consumption will be similar to yesterday's consumption.

Another possible assumption is that the load pattern of an electrical appliance on a Monday is more likely to look like the load pattern of the previous Monday than the previous day, i.e., a Sunday. Another persistence model using the same day of the previous week is proposed.

Optimal Persistence

An enhanced version of the persistence model, incorporating calendar information, holiday, and using averaging, is developed. Similar to the persistence model, this optimal persistence model is based on historical data without training. For Tuesday through Friday and Sunday, the forecast is generated by averaging the consumption values of the previous day and the corresponding weekday from the previous week. When forecasting Mondays or Saturdays, the prediction is generated by averaging the values from the same weekday of the prior two weeks. Additionally, when the forecast day is designated as a holiday in the UCSD calendar, the model selects the two nearest holidays or Sundays in the past and averages their consumption data.

Autoregressive model with exogenous input and Lasso regression

An Auto Regressive Model with eXogenous input (ARX) is developed, utilizing Lasso regression. This model has two days of historical data (one day ago and one week

ago) along with a holiday binary indicator as inputs. Historical data inputs are scaled between 0 and 1. The output of the model is one day ahead point prediction, i.e., 96 values are predicted since the dataset is in quarter-hourly resolution. The least absolute shrinkage and selection operator, commonly referred to as Lasso regression, is a technique in statistics and machine learning that combines variable selection with regularization. By constraining the sum of the absolute values of the regression coefficients to be below a set threshold, i.e., the L1-norm, Lasso regression effectively drives certain coefficients to zero [130, 135]. We set the parameter alpha, the penalty term that denotes the amount of shrinkage, equal to 0.001.

XGBoost regression

This forecasting model uses a more advanced machine learning approach, based on gradient boosting regression using the XGBoost library [28]. It has often proven successful in load forecasting research [16]. XGBoost models construct ensembles from decision tree models. Trees are incrementally added to the ensemble, each tailored to correct the errors of its predecessors. The fitting process involves minimizing a differentiable loss function using gradient descent. More details about XGBoost can be found in [28].

We implemented a local and a global approach. In the local approach, one XGBoost model is trained per plug load time series. In both approaches, seven days of historical data, scaled between 0 and 1, are used as inputs, and one day ahead is predicted, i.e., 96 values. In the global approach, early stopping and L1-regularization (alpha = 0.1) are implemented. In the local approach, more regularization is necessary to avoid overfitting. The parameters used are the following: the minimum sum of instance weight needed in a child is set to 8 (min_child_weight = 8), the maximum depth of the trees is set to three (max_depth = 3), the L1 regularization parameter alpha is set to 0.2 (reg_alpha = 0.2), the number of early stopping rounds is set to 20 (early_stopping_rounds = 20), the data is 80% subsampled (subsample = 0.8), and 80% of the features (col_sample = 0.8) are retained. The other parameters are kept to their default values.

Feedforward neural network regression

The proposed FeedForward Neural Network (FFNN) takes two days of historical data as inputs, i.e., one day and one week prior, along with a binary holiday indicator. Inputs are scaled between 0 and 1 to normalize the range of data values. The network generates a one-day-ahead point prediction for the plug load, i.e., 96 values since the dataset is in quarter-hourly resolution.

The architecture of the FFNN is adapted for two distinct approaches: a local and global approach. In the local approach, one FFNN is trained per plug. The structure includes an input layer with 128 neurons using the Rectified Linear Unit (ReLU) activation function, a hidden layer with 64 neurons also employing ReLU activation,

and an output layer comprising 96 neurons with linear activation to correspond with the 96 outputs for the day-ahead forecast.

In the global approach, one model is trained for all plugs. The architecture incorporates an input layer with 526 neurons followed by a dropout layer with a rate of 0.2 to prevent overfitting. This is succeeded by two hidden layers with 256 and 128 neurons, respectively, both utilizing ReLU activation. The architecture ends with an output layer with linear activation and 96 neurons, similar to the local approach.

In both scenarios, the training process is governed by two callbacks: early stopping to prevent overfitting and a reduction in the learning rate on a plateau to fine-tune the learning process for better performance. The Adam optimizer is utilized for its efficiency in both cases, with a maximum of 100 epochs for the local model and 200 epochs for the global model, ensuring adequate learning while avoiding excessive computational demands.

4.4.4 Scheduling

In the pipeline depicted in Figure 4.1, the scheduling step is emphasized in orange. The proposed approach integrates the outputs from both the ensemble clustering and forecasting models. Here, one-day-ahead forecasting is applied on the test set. Subsequently, the continuous forecast is binarized through the ensemble clustering model. This binary forecast then guides the individual plug scheduling, determining the activation and deactivation commands and their respective timings. We decided to call this schedule "forecast-driven". A baseline scheduling approaches is also proposed as a comparison. This baseline schedule, referred to as 'time-based' control in the literature, relies solely on predefined office hours. We also compare the baseline and the forecast-driven schedules with no schedule.

Time-based schedule

A fix time-based schedule is developed. Monitored plugs are turned off on weekends and every weekday between midnight and 5 am.

Forecast-driven schedule

In the proposed forecast-driven schedule, logic rules determine the on/off commands of the plugs based on the binary predictions. These binary predictions are the outputs of the subsequent application of the forecast model, in a one-day-ahead rolling window fashion, and the operating mode detection clustering. In the clustering output, a 0-label denotes that the plug is not in use, encompassing operating modes off, idle, and standby. Conversely, a 1-label signifies that the device connected to the plug is actively used. The rule for on/off scheduling is straightforward: if a device is consistently labeled as "0" (indicating it is unused) for a minimum of n consecutive time steps, it

is directed to switch off. In any other scenario, which includes instances where it is marked as "0" for fewer than n time steps or when it is labeled as "1", the device is turned on, or remains on.

Scheduling evaluation

The end schedule is evaluated using four different metrics. Each metrics allow us to make different conclusion on the approaches.

- Saved energy: The absolute saved energy is the difference in consumption between the true test set and the test set with the implemented schedule, i.e., time steps marked as turned off in the schedule have a zero consumption. The percentage of saved energy over the test set is computed by dividing the absolute value by the total energy. These relative and absolute saved energy can be computed per week, per plug, or even per day to obtain different insights. The schedule aims to save as much energy as possible.
- Number of violations: The number of violations is the number of times that the plug is scheduled as off while it is actually in use. The percentage of violations is also computed. These relative and absolute number of violations can be computed per week, per plug, or even per day to obtain different insights. The schedule should not prevent employees from working or using their appliances as they need. A minimum number of violations is desired for user convenience.
- Number of missed chances: The number of missed chances is the number of times that the plug is scheduled as on while it is not used. It is the number of time we could have saved energy but did not. The percentage of missed chances is also computed. These relative and absolute number of missed chances can be computed per week, per plug, or even per day to obtain different insights. When designing the schedule, a greater safety margin or expanded time-based schedules can result in increased missed chances for energy savings. It's crucial to strike a balance between energy conservation, the number of violations, and missed chances.
- Number of turn on/off commands: The number of times that a plug is turned on and off in a day or in a week can also be computed. This can help assess if the schedule might adversely impact appliances, as some appliances experience increased wear-and-tear when turned on or off.
- Energy efficiency: We also employ a metric termed energy efficiency. This metric is the ratio of useful energy to total energy consumed. Useful energy refers to the energy expended while the plug is active. If the plug operates solely when in use, the total energy consumed matches the useful energy, resulting in an efficiency of 100%.

4.5 Case Study

4.5.1 Problem statement

The pipeline proposed in this paper tackles the challenge of office plug scheduling. In a first step, the data is pre-processed. Secondly, the operating mode of each plug is detected, two modes are targeted: active or not active (encompassing idle, standby and off. Third, the plug load is predicted one day ahead. Finally, the forecast and the mode detection results are used in the scheduling step, along with logic rules to determine whether a plug should be turned on or off one day ahead.

4.5.2 Dataset

The raw dataset is collected through smart plug sockets, between the wall plugs and the electric appliance, with an integrated system as detailed in [29]. About 650 smart plugs are deployed in fifteen buildings of the campus of the University of California, San Diego (UCSD). The power level of each individual plug is recorded in mW at quarter hourly intervals or at five minutes interval in Jason files and accessible in a data lake.

The process of transforming the individual Jason file recordings into a usable dataset necessitated intensive data cleaning and formatting. The specific details lie outside the scope of this paper and will be elaborated upon in a subsequent publication centered on the dataset. This section provides a brief overview of critical pre-processing steps undertaken.

The dataset contained a blend of 5-minute and 15-minute recordings. To maintain consistency, these were resampled to a quarter-hourly resolution. Plug profiles that were only composed of zero recordings or lacking metadata were excluded. While there were no missing values in terms of empty recordings or NaNs in the raw data, instances were identified where the system was inoperative, such as during system-wide updates. It was observed that in such cases, the recording system substituted the gaps with zeros. We elaborated several scenarios to discern these erroneous zeros from actual zero power recordings and replace them with NaNs. One such scenario involved instances where all plugs in a building simultaneously displayed zero readings over a few time steps. In such cases, due to the improbability of all appliances being inactive for a consecutive 45 minutes for example, these readings were replaced with NaNs. Another scenario considered the start of a plug load time series with prolonged time of zero values, succeeded by consistent power level fluctuations for the rest of the recording period. This suggested non-usage during the COVID lockdown, followed by reactivation of the devices. The initial zeros did not truly represent genuine usage trends. Lastly, prolonged zero sequences towards the end of a plug load time series were substituted with NaNs, indicating the appliance's disconnection from the smart plug, preventing recordings. After these steps, 625 plug recordings spanning over at least a year remained.

In the subsequent phase, a subset of plugs was chosen for this particular study. Plug load profiles exhibiting extensive amounts of NaNs and zero values were discarded. Specifically, profiles with over 30% of NaNs or 30% of zeros in total were excluded. Additionally, it was ensured that the test set contained less than 3% NaNs, as the aim was not to schedule underutilized plugs.

The final selection was based on the plug device type. The current research focuses on scheduling devices such as printers, copiers, scanners, and TVs. Although a large number of computers were monitored, they were excluded from scheduling to avoid potential disruptions for employees, especially those who may be running models overnight or who might not have backed up recent work. This rigorous selection yielded a dataset of 169 high-quality smart plug time series spanning 498 days, from November 18th, 2022, to March 31st, 2023, ready for scheduling.

This final pre-processed dataset is openly accessible [20]. In the upcoming publication centered on the dataset, further analysis and visualizations that were a central part of the data exploration, will be detailed.

4.5.3 Results & discussion

Mode detection

The mode detection approach, based on clustering ensemble, is applied on the given dataset. This method converts the power level values into binary terms based on the detected mode. A label of 1 indicates an active plug, while a label of 0 signifies a non-active plug, encompassing states such as off, idle, or standby. As explained in Section 4.4.2, a visual evaluation of mode detection accuracy is undertaken through a heatmap comparison. The accuracy of the mode detection is classified into three categories: high-fidelity, partial-fidelity, and low-fidelity. The results are detailed in Table 4.1. Close to 70% of the plugs fall under the high-fidelity category, indicating that the algorithm correctly discerns the plugs usage patterns. Another 17% of the plugs demonstrate partial-fidelity detection. In these instances, while the algorithm identifies the core active usage patterns, it also mistakenly labels some non-active power values as active. Lastly, the algorithm exhibits poor performance for 12% of the plugs, categorizing them as low-fidelity.

Given the qualitative nature of the evaluation, we believe it is essential to provide a detailed illustration of these results. All figures discussed in this section are located in 4.5.3 to enhance textual continuity and readability. For each plug, the illustration consists of four plots. First, the complete time series is plotted, with the training set in light blue and the test set in dark blue. On the test set, the labels outputted by our mode detection algorithms are incorporated. Active power values are marked with green dots, while inactive values in orange dots. The first heatmap is build using the original test data. The second heatmap is generated with the binarized test data through our detection algorithm. The comparison of these two heatmaps is used to categorize our algorithm performance. The last heatmap is formed with the

	Number of plugs	Percentage of plugs
High-fidelity	119	70.41 %
Partial-fidelity	29	17.16 %
Low-fidelity	21	12.43 %
Total	169	100 %

Table 4.1. Mode detection results are presented based on a qualitative comparison between original and binarized heatmaps. High fidelity denotes accurate usage pattern detection. Partial fidelity suggests the detected pattern includes the original, but some values labeled as active may be inactive. Low fidelity means the algorithm does not capture the underlying usage pattern accurately.

binarized test data, based on power thresholds specified in product sheets. Appliance metadata, including aspects like type, brand, and model, facilitated the acquisition of the pertinent specification sheet and power threshold details. A dotted line on the initial time series plot also depicts this data sheet threshold. The discussion around these specification sheets thresholds concludes this section. Note that the units and scales of the heatmaps are different, as the original data is in Watts, while the data after mode detection is in binary format.

Figure 4.3 presents instances of high-fidelity detection results. In every example, our mode detection algorithm effectively captures the continuous power usage pattern in the binarized format. Devices showcased include a printer, a TV, and a copier.

Figure 4.4 displays several plugs classified as partial-fidelity. In each instance, the detected modes encompass the active usage pattern and more: the algorithm occasionally labels non-active power values as active. However, this does not pose an inconvenience to the user; the plug will not inadvertently turn off when it should remain on. For some plugs, there is no distinct original usage pattern, which could be a reason for the lower performance of the algorithm.

In our focused examination of the **low-fidelity** detection results, we aim to understand potential sources of failure and avenues for the improvement of the mode detection algorithm. We illustrate, discuss and propose solutions to each challenge. *The primary set of challenges relates to the inherent attributes of the plug load data.* There are appliances for which the absolute difference in Watts between the active and standby mode is very small, as illustrated in Figure 4.5a. Given the fact that GMM clustering allows overlapping clusters, if the small absolute difference between two modes makes them hard to detect, there exists a risk of unintentionally turning the device off during its active mode. It might be prudent to consider whether these devices should be scheduled at all.

Additionally, situations arise where three distinct modes are perceptibly present. However, the current design of the mode detection is oriented to discern two modes, grounded on this set input parameter in the k-means as well as the GMM clustering. This might potentially affect the efficacy of the proposed algorithm on these specific profiles. An example is illustrated in Figure 4.5b. To mitigate this, a solution would

be to apply mode detection with parameters adjusted for both two and three modes. Subsequent performance evaluation on a validation set would then determine the most optimal parameterization. Another observed phenomenon is the occurrence of concept drift [60]: over time, devices can exhibit changing behavior. An example is shown in Figure 4.5c. This might be due to wear and tear, software revisions, instances where users change the plugged appliances, or situation where users employ power strips. This challenge can be addressed by detecting concept drift and subsequently re-training the mode detection algorithm. Raising awareness among users about the scheduling system and the repercussions of unplanned alterations is also an option.

The second type of challenges is inherent to the intrinsic attributes of the implemented algorithm. It is well known that algorithms like k-means and GMM are sensitive to their initialization. To improve the robustness of the algorithm, one approach would involve running the mode detection task multiple times, and then averaging the outcomes. Alternatively, another strategy would entail focusing on adopting a more robust initialization method. Furthermore, in our algorithm, each power value is currently evaluated and labeled individually, not accounting for its sequential or temporal context. This can lead to overlooking specific transitional phases. Including additional features, like the time of day, could potentially increase the accuracy of the algorithm. Another potential improvement is linked to the resolution of measurements. Recordings with a higher frequency might improve detection. Indeed, devices like televisions or printers typically can transition between off and active modes within a 15 min interval; therefore 15 min averages may contain power values from multiple states.

Lastly, we compare our mode detection algorithm with thresholds from appliance specification sheets. For each example highlighted in this section, we determined the power range for both active and standby modes, as provided in the appliance specification sheets. Despite the extensive manual effort required, it is worth noting that such information is not always accessible. Out of the 13 appliances presented in Figures 4.3, 4.4, 4.5 only one specification sheet thresholds yields high-fidelity results: Figure 4.4(a). The printer in Figure 4.4(b) receives a partial-fidelity classification; it aptly labels the active mode but also mislabels certain power values that could be marked as inactive. All other mode detection results are categorized as low-fidelity. The remaining mode detection results using specification sheets fall into the low-fidelity category. It is worth observing that the copier in Figure 4.3(b) seems to capture the general usage pattern but mislabels several active power values as inactive, posing a risk of plug deactivation during actual use. One reason for apparently overly large specification sheet thresholds could be the 15 min averaging period. The specification sheet threshold assumes that the device is active during the entire 15 minute period. For example, for a copier or printer, this would mean continuous printing or copying for 15 minutes. Such usage is unlikely in practice. Most often, a 15 min interval will contain periods of activity together with inactivity or even standby modes. The inactive periods would then cause the 15 min average power to decrease below the specifications threshold.

Forecasting

The different forecasting model detailed in Section 4.4.3 were implemented in python and evaluated on an unseen test set, i.e., the last 20% of the dataset. Traditional distance metrics, the Mean Absolute Error, Mean Squared Error and Root Mean Squared Error were used. Smaller values indicate better performance. Table 4.2a presents the average metrics over all plugs. Additionally, the percentage improvement in terms of RMSE is computed, in function of the naive persistence model based on the consumption of the same day a week ago.

Another evaluation of these models is proposed, oriented towards the end goal of the forecasting step: the plug scheduling. The ground truth is composed of the labels of the mode detection algorithm applied directly on the true test set power values, as described in Section 4.4.2. The continuous point predictions from the forecasting are also binarized, using the operating mode detection algorithm. These are compared with the ground truth. The evaluation consists of two traditional and complementary classification metrics: the recall and the miss rate. The recall is also called true positive rate, sensitivity, or probability of detection. In our specific case, it is an indication on how many times the plug has been labeled as active out of all the times it was actually active. We aim to have a recall as high as possible. The miss rate, also termed false negative rate, is computed as $1 - \text{recall}$. It highlights the number of time that the plug was labeled as inactive while it was active. We want the miss rate to be as low as possible, as this risks to incur discomfort to the users of the appliances. Additionally, the percentage improvement in terms of recall is computed, in function of the naive persistence model based on the consumption of the same day a week ago. These results are presented in Table 4.2b.

In both tables, the models are ordered from the best performing to the least performing based on the respective metric. This ordering varies significantly depending on the chosen evaluation metric. This variance underscores the importance of selecting appropriate metrics that align with the end goals of the forecasting process. However, in both evaluation methods, the global feedforward neural network model ranks as the top performer. These findings align with those of Spiliotis et al. [124] focused on intermittent time series forecasting. However, it is important to note that their experiments were limited to one-step-ahead forecasts. This global FFNN model achieves a 26% improvement in RMSE compared to the baseline and a 42% improvement in recall. Interestingly, the addition of holiday data does not enhance RMSE performance for both local and global feedforward neural networks, nor for the LASSO autoregressive model. However, it does seem to reduce the miss rate by nearly 5% for the local FFNN.

It is important to note that the ground truth used in these evaluations is derived from the preceding operating mode detection algorithm and is not an absolute measure of the ground truth. Consequently, while this ranking offer useful insights into model performance, they should be interpreted with some caution.

	MAE	MSE	RMSE	Improvement (%)
Global FFNN	3.85	458.26	10.53	26.0
Individual FFNN	3.53	482.49	10.79	24.17
Global FFNN with holidays	3.65	467.21	11.07	22.21
Individual FFNN with holidays	4.0	489.28	11.15	21.64
Global XGBoost	2.86	529.23	11.51	19.11
Individual XGBoost	4.67	522.19	12.18	14.41
Optimal Persistence	2.79	642.54	12.28	13.7
ARX	4.49	606.44	12.33	13.35
ARX with holidays	4.51	597.54	12.38	13.0
Persistence 1 day	2.61	809.95	13.91	2.25
Persistence 1 week	3.03	817.65	14.23	/

(a) Forecasting results based on the evaluation of Mean Absolute Error, Mean Squared Error and Root Mean Squared Error on the continuous predictions.

	Recall	Miss Rate	Improvement (%)
Global FFNN	74.09	25.91	42.55
Individual XGBoost	71.27	28.73	40.27
ARX	69.55	30.45	38.8
ARX with holidays	68.59	31.41	37.94
Global XGBoost	59.72	40.28	28.72
Global FFNN with holidays	59.17	40.83	28.06
Individual FFNN with holidays	58.68	41.32	27.45
Individual FFNN	54.14	45.86	21.38
Persistence 1 day	48.73	51.27	12.65
Optimal Persistence	45.52	54.48	6.48
Persistence 1 week	42.57	57.43	/

(b) Forecasting results based on the evaluation of recall and miss rate on the binarized predictions. The continuous predictions are binarized using the operating mode detection algorithm.

Table 4.2. Forecasting models evaluated on the test set, i.e., the last 20% of each plug load time series and averaged over all plugs.

Scheduling

The scheduling results are presented in Table 4.3. The "true test set schedule" is the schedule obtained using the true values of the test set, as if the predictions were perfect. These are binarized using the operating mode detection algorithm, and the schedule is generated using the forecast-driven approach described in Section 4.4.4. This can be considered as the perfect schedule in terms of forecast, still keeping in mind that there is no fundamental ground truth for the operating mode detection step. The "no schedule" approach considers the plugs as always on and the baseline schedule is the simplistic time-based approach, as described in Section 4.4.4. These three specific schedules are interesting to compare the forecast-driven methods, in the sense they each give us insights into the realistic expectation of the pipeline. The true test set schedule indicates the maximum amount of energy that could be saved, i.e., 42.07% along with the best energy efficiency that could be obtained, i.e., 87.32%. This perfect schedule also shows that there are at least 6.28% of missed chances, these are occurrences when the plug is marked as inactive for such a short period that the schedule does not deem it long enough to actually turn the plug off. This perfect schedule has one of the highest number of turn on/off commands. When no schedule is applied, the percentage of missed chances, indicating the situations when the plugs are marked as inactive and not turned off, is 70%. This confirms our assumption in Section 4.4.2 that the plugs are, on average, approximately used 30% of the time. In this no schedule case, we also have an indication of the worst energy efficiency, i.e., 39.79%. Finally, the baseline schedule provides insights on the metrics that can be attained without complex computing, training nor data collection and the proposed approaches should at least be better than this.

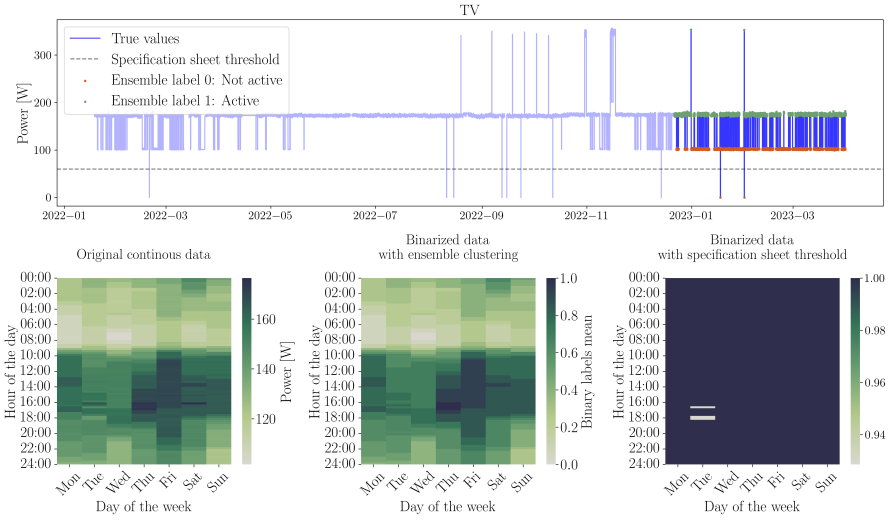
Selecting the right schedule depends on the specific application. For instance, when scheduling critical appliances, minimizing the number of violations is important to ensure reliability and continuity for the users. Conversely, in scenarios where non-critical appliances are scheduled, the objective may shift towards maximizing energy savings. Formulating an optimization problem that integrates these different considerations, would allow selecting a schedule that aligns precisely with the specific needs.

The results table clearly indicates that the global Feedforward Neural Network (FFNN) leads to the most effective forecast-driven scheduling. The proposed pipeline, while innovative and promising, also highlights that enhancements in forecasting could substantially increase energy savings and reduce violations. The potential improvements and future research directions for this pipeline are detailed in Section 4.6. A separate section is dedicated to this future direction discussion, given the breadth of research areas that this pipeline spans. The implemented models are chosen for their fast training capabilities and minimal data requirements, which are essential considerations. The objective is to have a model that is efficient and scalable. This ensures that the forecasting model remains practical for widespread application.

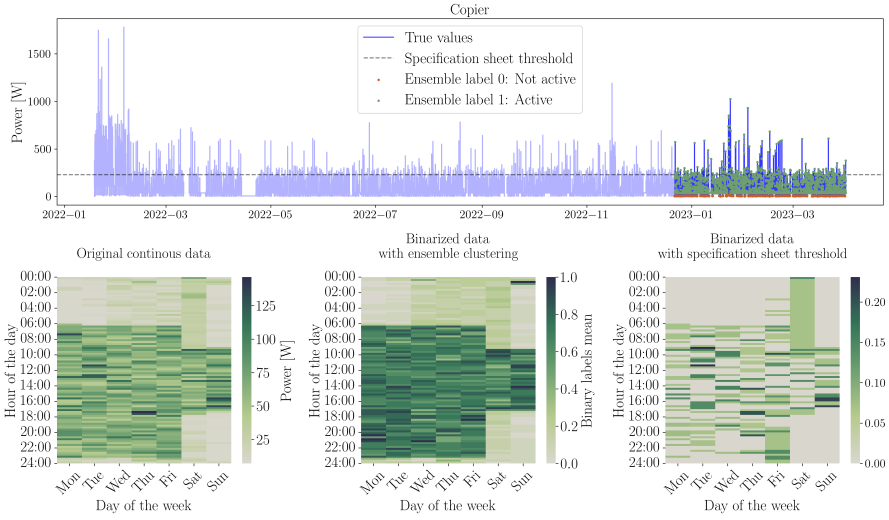
	Number of violations (%)	Missed chances (%)	Energy saved (%)	Number of turn on/off commands per plug per day	Energy Efficiency (%)
True Test Set Schedule	0.0	6.28	42.07	4.33	87.32
No Schedule	0.0	70.26	0.0	0.0	39.79
Global FFNN Schedule	2.21	33.11	21.92	1.62	45.89
Individual XGBoost Schedule	2.75	35.82	16.51	1.82	44.77
Global XGBoost Schedule	3.27	27.67	25.84	3.33	47.92
Individual FFNN Schedule	4.61	23.77	29.38	3.71	47.85
ARX Schedule	5.09	28.63	20.38	2.18	47.35
Optimal Persistence Schedule	6.05	10.03	40.15	3.4	55.77
Persistence 1 day Schedule	6.05	12.7	46.17	4.26	54.51
Persistence 1 week Schedule	7.95	14.3	47.7	4.32	50.31
Baseline Schedule	11.4	38.09	41.88	1.45	43.28

Table 4.3. The different forecast-driven schedules are compared with the naive baseline schedule and no schedule.

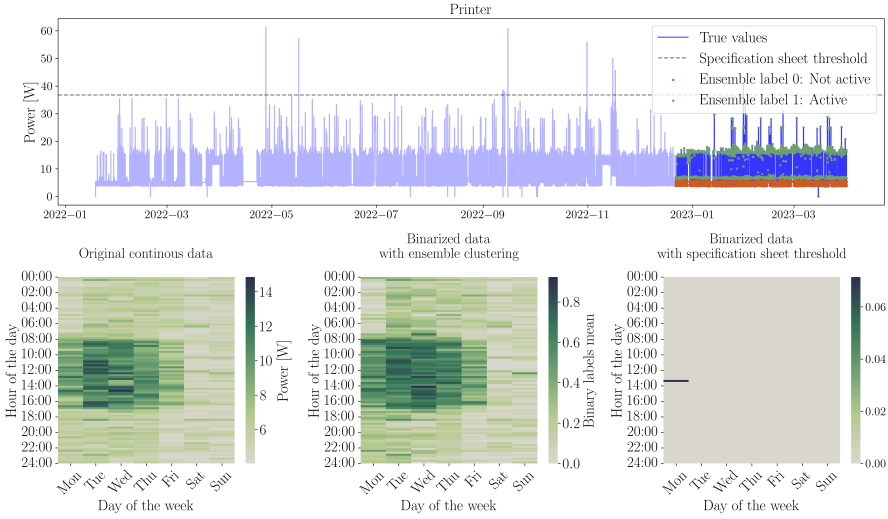
Illustration of the operating mode detection results



(a) High-fidelity mode detection of a TV.

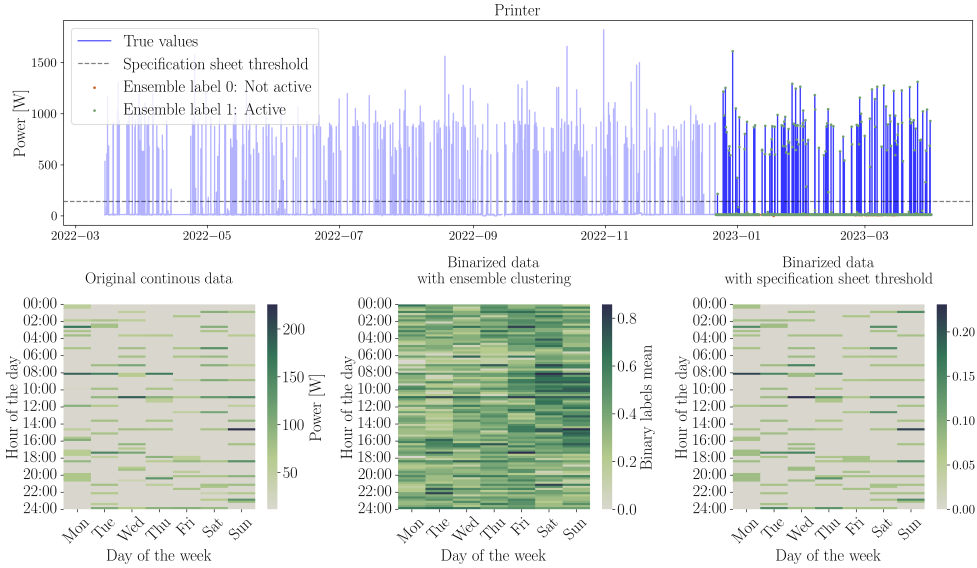


(b) High-fidelity mode detection of a Copier.

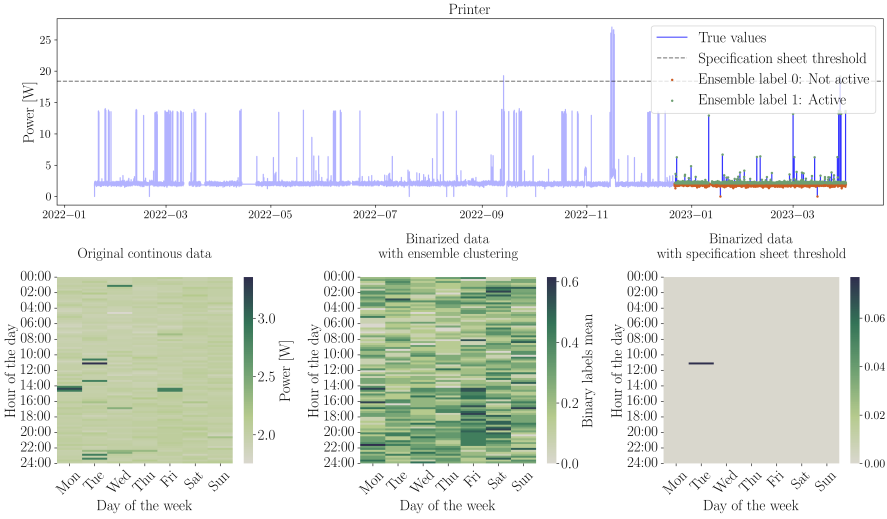


(c) High-fidelity mode detection of a printer.

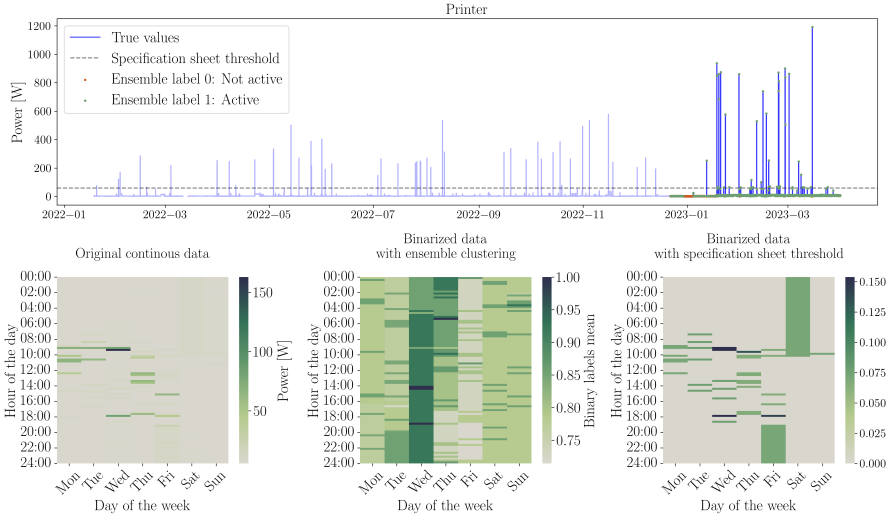
Figure 4.3. (Continued) Each figure consists of four plots. First, the complete time series is plotted, with the training set in light blue and the test set in dark blue. On the test set, the labels outputted by our mode detection algorithms are incorporated. Active power values are marked with green dots, while inactive values in orange dots. The first heatmap is build using the original test data. The second heatmap is generated with the binarized test data through our detection algorithm. The comparison of these two heatmaps is used to categorize our algorithm performance. The last heatmap is formed with the binarized test data, based on power thresholds specified in product sheets. The dotted line on the initial time series plot also depicts this data sheet threshold. (a), (b) and (c) are examples of mode detection results classified as high-fidelity. Each binarized dataset captures the underlying usage pattern of the true dataset accurately. Different device types and usage patterns are represented.



(a) Partial-fidelity mode detection of a printer.

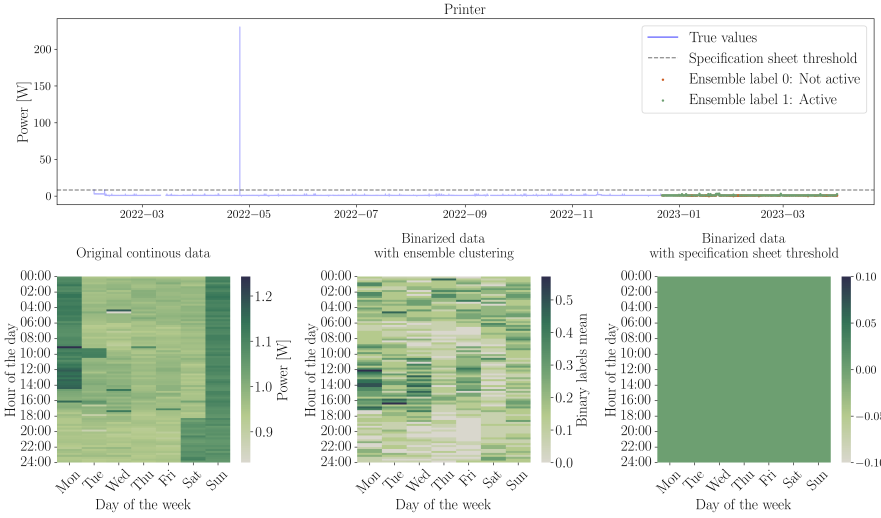


(b) Partial-fidelity mode detection of a printer.

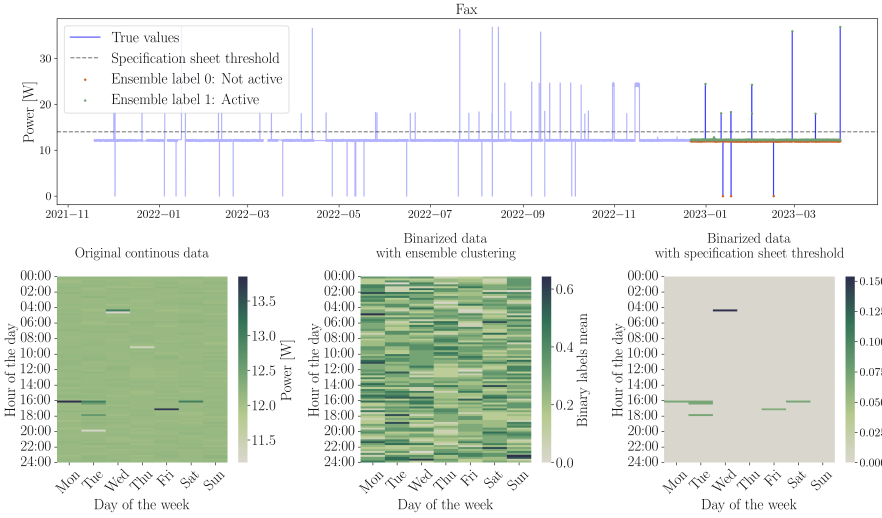


(c) Partial-fidelity mode detection of a printer.

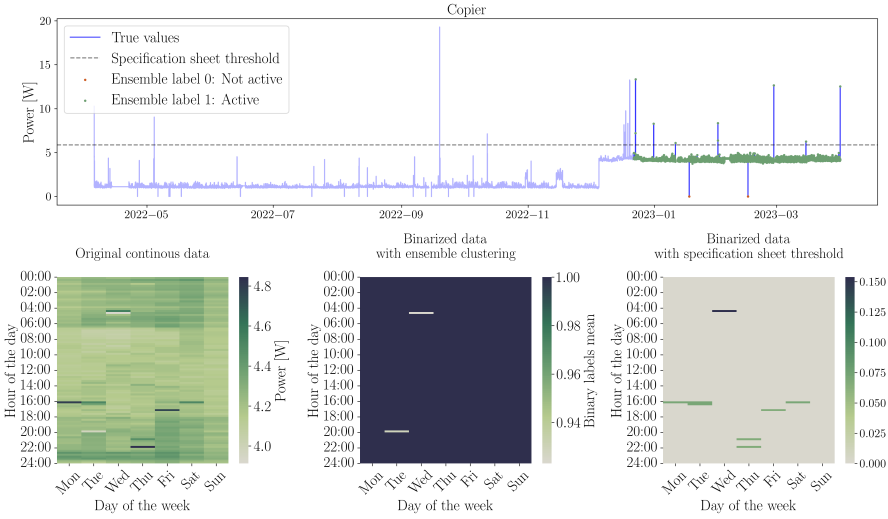
Figure 4.4. (Continued) Each figure consists of four plots. First, the complete time series is plotted, with the training set in light blue and the test set in dark blue. On the test set, the labels outputted by our mode detection algorithms are incorporated. Active power values are marked with green dots, while inactive values in orange dots. The first heatmap is build using the original test data. The second heatmap is generated with the binarized test data through our detection algorithm. The comparison of these two heatmaps is used to categorize our algorithm performance. The last heatmap is formed with the binarized test data, based on power thresholds specified in product sheets. The dotted line on the initial time series plot also depicts this data sheet threshold. (a), (b) and (c) are examples of mode detection results classified as partial-fidelity. Each binarized dataset captures the underlying usage pattern of the true power values, nevertheless it also wrongly labels some non-active values as active.



(a) Low-fidelity mode detection, which might be due to the small difference in power absolute values.



(b) Low-fidelity mode detection, which might be due to the three existing operating modes.



(c) Low-fidelity mode detection, which might be due to the occurrence of concept drift.

Figure 4.5. (Continued) Each figure consists of four plots. First, the complete time series is plotted, with the training set in light blue and the test set in dark blue. On the test set, the labels outputted by our mode detection algorithms are incorporated. Active power values are marked with green dots, while inactive values in orange dots. The first heatmap is build using the original test data. The second heatmap is generated with the binarized test data through our detection algorithm. The comparison of these two heatmaps is used to categorize our algorithm performance. The last heatmap is formed with the binarized test data, based on power thresholds specified in product sheets. The dotted line on the initial time series plot also depicts this data sheet threshold. (a), (b) and (c) are examples of mode detection results classified as low-fidelity. (a) The proposed algorithm has difficulty grasping the usage patterns since the difference between active and non-active mode is very small in terms of absolute power values. (b) There seem to be three operating modes for this appliance, while the proposed algorithm is designed to detect two modes. (c) The large change in power dynamics between the training set and the test set, called concept drift, makes it challenging for the proposed algorithm to accurately discern the new usage pattern.

4.6 Future research directions

In this section, based on the reviewed literature, the presented research, and the result discussion, we outline potential research directions in smart plug utilization and building energy efficiency. These suggestions provide potential extensions and improvements to the existing study.

Occupancy data: The integration of occupancy data into plug load forecasting models is recommended to increase accuracy. Several recent studies [2, 30, 145] have underscored the significant role of occupancy data in plug load predictions.

Categorizing plugs: A recommendation by Wang [141] suggests categorizing plugs based on their operational modes: continuous versus intermittent. For instance, while water dispensers generally run continuously, devices like TVs operate intermittently. Adopting distinct mode detection and forecasting strategies for each category could yield more accurate predictions. Another approach could be to cluster the time series and then developing a forecasting model per cluster or device type, leveraging more historical data. This hinges toward the new trend of global forecasting models [60].

Eco-feedback: Promotion of occupants' awareness, also called eco-feedback, emerges as a promising element in energy reduction. Kamilaris et al. [74] conducted an illuminating study over 22 weeks, targeting the understanding of office workers' energy consumption behaviors. Results revealed the compelling impact of continuous personal feedback, with notable reductions in energy use observed in the third and fourth months. Moreover, the study highlighted the efficacy of group-level feedback and peer-education interventions in fostering energy-saving behaviors. Doherty et al. [38] and Gandhi et al. [47] further emphasize the potential of informational campaigns, rewards programs, gamification, and regular energy usage feedback in enhancing occupants' energy conservation habits.

Transfer learning: The idea of leveraging hand labeling from existing datasets for transfer learning on new datasets, given device type consistency, could be promising. The dataset by Kalluri et al. [73], which was visually inspected and hand labeled, could provide insights for similar applications.

Impact of each pipeline step: It would be interesting to assess numerically the performance at each stage of the proposed pipeline. Specifically, the impact of precise operating mode detection and the impact of improvement in forecasting on the subsequent scheduling tasks.

Return on investment: The economic implications of plug load monitoring cannot be overlooked. Comprehensive financial analyses, as demonstrated in the NREL report by Trenbath et al. [131] and Wang et al.'s [142] return in investment analysis, could shed light on the actual monetary benefits and potential payback periods of such initiatives.

Frame a global optimization problem: Optimizing the scheduling of energy consumption should be approached from a multifaceted perspective, encompassing occupants'

comfort, potential financial gains, and the overarching goal of energy reduction. This multi-objective optimization problem can pave the way for sustainable energy management solutions in the future [118].

Randomized control trial: In the context of smart plug scheduling, implementing a randomized controlled trial could offer valuable insights into the practical effectiveness of the proposed scheduling strategies. Specifically, we suggest applying plug scheduling for an 8-hour duration, refraining from scheduling in the subsequent 8 hours, and then reintroducing the scheduling for another 8 hours. This cycle would be maintained over several weeks. By comparing the scheduled hours against the non-scheduled hours, we could assess performance metrics such as energy conservation and user satisfaction.

4.7 Conclusion

This paper introduces a pipeline for scheduling plug loads in office buildings with the aim of energy reduction. The pipeline integrates three components: plug operating mode detection, plug load forecasting, and plug scheduling based on the two prior components. The objective is to reduce the energy consumed when appliances are not actively used.

Significant contributions of this research include the development of an extensive dataset, capturing over a year's worth of plug load data from multiple office buildings. This dataset is made publicly available to foster further research. The first literature review on plug operating mode detection methodologies and terminology is presented, providing clarity and a foundation for future work in this domain. An unsupervised approach for determining plug operating modes is proposed, which has been a challenge in past literature. Additionally, the research introduces initial efforts in the domain of individual plug load forecasting, an area previously unexplored. The resulting pipeline demonstrates potential energy savings of up to 20% on plug loads, marking a significant step towards smarter energy management in buildings. Furthermore, the developed code is made publically available for reproducibility and benchmarking in further research work.

The research also acknowledges certain limitations, such as the absence of a numerical metric for evaluating the plug mode detection algorithm accuracy and opportunities in improving forecasting models. A specific section is dedicated to future research in the smart plug field.

In conclusion, the research presented proposes a novel and promising framework for office plug load scheduling and underscores the importance of sharing data and methodologies. It lays a foundation for collaborative progress and sets a benchmark for plug mode detection, forecasting and plug scheduling towards energy efficient buildings.

Chapter 5

Conclusion and perspectives

In this conclusion, we restate the research objectives and how we have addressed those in the different chapters. We highlight how they align with the original goal of contributing to the forecasting literature and more broadly to the sustainable development goal #7 (SDG7) of the United Nations, i.e., the need to **ensure access to affordable, reliable, sustainable, and modern energy for all**. The SDG7 will mainly be attained through the acceleration of electrification and investment in renewable energy sources. This will subsequently lead to increased electricity consumption, higher volatility in demand, and higher uncertainty in production. These aforementioned challenges can be tackled by solving different forecasting problems and this is the core of the research.

5.1 Scientific conclusions

Chapter 2

Chapter 2 proposes a method for predicting **monthly and yearly electrical consumption of individual households one year ahead**. This approach combines pre-processing, data augmentation, clustering, and ensemble learning, to generate accurate point forecasts. A key contribution of this algorithm is its ability to make predictions for a full year ahead, even when dealing with incomplete data, such as households with only one month of historical records. Additionally, our approach has the benefit of low computational requirements, allowing it to scale up to millions of households, without the need for specific household attributes like the number of rooms or heating system type. This model has demonstrated state-of-the-art accuracy, as evidenced by its top ranking in the IEEE-CIS competition.

Together with the other competition submissions, this model stands out as the first to address the challenge of long-term electricity forecasting at the individual household level and is in the upcoming, yet promising, research field of global models.

The original application of this method, and also the reason for the contest [132], is to determine customers' advance payments. This method offers supplier companies the capability to predict the consumption of any customer for the next year, even when dealing with incomplete data, e.g., if a new client joins this utility company.

More broadly, this contributes to SDG7 by avoiding uncomfortable situations for customers if their bills are over- or underestimated, incurring unnecessary additional monthly costs or sudden regulation costs respectively. It also allows suppliers to handle pricing and long-term planning.

Chapter 3

Chapter 3 introduced a novel probabilistic global method for **short-term forecasting of residential electricity** consumption, based on similar days selection and empirical quantiles. We assume that household consumption patterns not only give explicit information on the load, but also encapsulate other implicit dynamics, related to the residents' behavior, weather conditions, and potentially other influential elements. When we observe similar consumption for two households over a sequence of days, it hints at similar underlying dynamics. Empirical quantiles are derived directly from observed historical data without making theoretical assumptions. They offer a way to express uncertainty in predictions. If one is interested in a worse-case scenario of an entire day of predictions, a coherent profile can be extracted by choosing the q-th quantile value for each time step. This method is able to generate forecasts with minimal historical data requirements, only seven days of the concerned household, making it adaptable to situations where data availability is limited. Through rigorous evaluation on three different datasets, of over 300 households each, the proposed method has consistently demonstrated superior robustness to changing consumption profiles.

This method contributes to the forecasting literature with its four attributes, i.e., (i) using similar day selection, (ii) computing empirical quantiles and being a (iii) global and (iv) probabilistic method, both currently under-researched and rarely combined in the field of low-voltage grid, even if proven successful in other applications. Another significant contribution is the commitment to reproducibility and transparency. This research uses several open datasets in addition to the private dataset from Fluvius. The code of the novel approach as well as the implemented benchmark models are shared publicly. This ensures that the research remains accessible to the scientific community, fostering collaboration and further advancements in the domain, as strongly recommended by Haben et al. [56].

Several applications can benefit from this novel probabilistic forecasting algorithm. The development of Flanders low-voltage grid digital twin at VITO, and the subsequent objective of **forecasting congestion** on the network, initiated this research. It

also became clear along the way that **data imputation of incomplete smart meter** recordings could benefit from this method. Another key application area is the management of the hosting capacity of the low-voltage grid. The hosting capacity denotes the available capacity within a network [134]. It is increasingly vital as peak consumption and peak injection are expected to rise substantially. Accurate forecasting of electricity usage for each connection to the grid is essential for several applications, such as controlling battery operations, directing electric vehicle charging behaviors, and implementing demand-side responses. These responses may include strategies like temporarily shutting down heat pumps or reducing their maximum output [57]. Moreover, another significant use case is explored in a project involving VITO, KU Leuven, and the University of Gent, i.e., the study of **non-firm connection agreements**. Typically, a connection has a fixed maximum capacity, but under non-firm connection agreements, the distribution system operator (e.g., Fluvius) may impose a reduced maximum power capacity at times of potential grid congestion. Furthermore, at an individual household level, there are use cases for **reducing electricity costs** by synchronizing consumption with personal photovoltaic production or day-ahead electricity prices. This approach does not directly address grid capacity but rather aligns consumption with the variable production from wind and solar energy sources. The research presented in this chapter, including the study conducted in Chapter 4, is relevant in this context and can be integrated into home energy management systems for more cost-effective electricity usage and reduced peak consumption.

Within the framework of the United Nations Sustainability Goal #7, this study contributes to the reliable dimension, making sure all households have access to electricity at all times, especially during peak demand time, also allowing the distribution grid operator to prioritize updates or improvements on the grid. As discussed in the applications, it could also contribute to cost reductions in the electricity bills of households.

Chapter 4

Chapter 4 details a pipeline, focusing on three key components: **(i) smart plug load mode detection, i.e., idle vs. active, (ii) load forecasting, and (iii) on/off scheduling one day ahead**. The proposed methodology leverages unsupervised clustering methods to detect modes, followed by one-day-ahead load forecasting, and on/off scheduling leveraging the predictions and logic-based rules. The dataset is from a non-residential real-life setup at the University of California, San Diego (UCSD).

To the best of the authors' knowledge, the dataset used, and the pipeline proposed, are the first of their kind. The dataset includes over 150 plug recordings, over a year, from different appliances and different non-residential buildings. This dataset is made available to the scientific community, along with the code produced, aspiring for more research in this novel field.

By achieving electricity savings of up to 20% this work exemplifies the potential for

enhancing demand flexibility in order to optimize building energy efficiency. This pipeline is already in use, in the prototyping phase, at UCSD.

This research is contributing to SDG7, as it actively aims for a more sustainable energy ecosystem within non-residential settings, and effectively reduces electricity costs.

In conclusion, these three research chapters have addressed a series of challenges related to low-voltage load and office plug load forecasting, with the overarching goal of contributing to Sustainable Development Goal #7 of the United Nations. This dissertation underlines the crucial role of accurate and scalable forecasting in the energy sector and is a promising step toward facilitating the ongoing energy transition.

5.2 Practical takeaways

In the field of forecasting, applying theoretical knowledge to real-world scenarios presents unique challenges and opportunities. This section outlines key practical takeaways, focusing on strategies that improve forecasting accuracy and effectiveness.

Benchmarking against strong baselines. Benchmarking developed forecasting methods against strong baselines is a fundamental practice. It is not uncommon for these established benchmarks to perform surprisingly well. By comparing new methods against robust baselines, researchers can gauge the genuine improvement of their approaches. This practice helps establish a clear understanding of the progress and contributions made in the field. Any improvement should also be put in perspective with the (computational) complexity of the developed model. Although it might sound counterintuitive, the complexity and stochasticity of individual household (and plug) data usually make "simple" models win. This doctoral research contributes to benchmarking by providing open-source code and accessible datasets. This initiative not only upholds the principle of transparency but also sets a precedent and standard for subsequent work. With these resources readily available, there's a compelling argument for future researchers to benchmark their innovative models against our work, on the same open datasets or on new datasets, thus fostering a consistent and comparative evaluation landscape in the discipline.

Data preprocessing. The quality of input data significantly influences forecasting accuracy. Data preprocessing, including data cleaning, imputation, outlier removal, and feature selection, is a cornerstone of successful forecasting. It is of course to be linked with the previous recommendation, once the data is good, strong baselines are often enough to produce accurate forecasts. Robust preprocessing techniques are essential to ensure that models receive clean, relevant, and informative data. Investing in data preprocessing is an investment in the overall forecasting process. This can be

highlighted with an example from Chapter 2, indeed the approach proposed in the paper outperforms all other approaches once the outlier profiles are detected.

Adapted Evaluation. Effective evaluation methods are crucial for assessing the quality and reliability of forecasts. For example, as detailed in [58], traditional metrics do not grasp the true quality of point forecast when peaks are present in the data. Another example is the probabilistic methods still evaluated with point forecasting metrics, which limits the conclusions on the results. Additionally, depending on the end application, evaluating not only the forecasting improvements but also the end goal should become standard practice. What if a 5% forecasting improvement yields a 20% improvement in the end goal, but the next 10% improvement in forecasting only yields 1% improvement on the end goal? Then this last improvement step (and probably increased complexity) in the forecasting model might not be necessary.

Visualizations for deeper insights. Data visualizations are invaluable tools for gaining deeper insights into results. They provide a visual inspection, allowing researchers to identify trends, anomalies, and areas for improvement that may not be apparent through numerical metrics alone. Leveraging visualization techniques can lead to richer, more informed forecasting decisions. In Chapter 3 for example, visualizing the predictions with the largest CRPS difference, allowed us to understand that the global approach performed better on households with a large change in consumption patterns. Another example arose during the idle vs. active mode detection step in Chapter 4. According to the true threshold, which defined whether a plug was active or idle, some plugs were always idle or always active. This seemed surprising and visualizing the results in the form of heat maps, allowed us to understand that some true thresholds were erroneous and subsequently allowed us to tackle this issue. This would not have been possible with numerical results only.

Collaboration with Industry Experts. Close collaboration with industry professionals is crucial. These experts provide valuable insights into the actual needs and challenges faced in their respective fields, guiding researchers away from abstract theoretical models and towards practical, impactful solutions. For instance, in the context of low-voltage grids, industry experts stress the importance of scalability. Models that require extensive time and infrastructure for execution, despite their high accuracy, may not be practical if they cannot be efficiently scaled to numerous households. Engaging with industry specialists ensures that the forecasting efforts address relevant problems and contribute meaningfully to the field. This approach not only aligns research with practical demands but also opens avenues for innovative challenges that might otherwise be overlooked in purely academic settings.

5.3 Research perspectives

This section introduces research opportunities yet to be fully explored. Section 5.3.1 is focused on advances in the technical aspects of low voltage load forecasting. Section 5.3.2 broadens the scope to encompass wider considerations within the energy

landscape. It aims to bridge the gap between technical forecasting advancements and their real-world implications, ensuring that progress in forecasting is not only technically sound but also aligns with and supports societal values and privacy concerns.

5.3.1 Perspectives in load forecasting

Enhancing data quality and diversity

The foundation of accurate load forecasting lies in the quality and diversity of data. While the research in Chapters 2 and 3 leveraged multiple smart meter datasets from different countries and years, it is imperative to use more recent and diverse datasets that capture the evolving energy landscape. This includes data reflecting the increased adoption of heat pumps, solar panels, and electric vehicles in recent years, as well as the impact of COVID-19 on households habits.

Chapter 4 plug load dataset comes from several buildings and devices within a single campus. Although it is the first of its kind made publicly available, future research could benefit from greater diversity in data sources to demonstrate the generalizability of the results. Furthermore, combining plug load data with additional information, such as occupancy patterns and geolocation data, holds potential for improving prediction accuracy.

Promising low-voltage load models

To advance fine granular load forecasting, several directions warrant exploration. **Global forecasting** methods have shown robustness in adapting to changing conditions. Additionally, investigating **clustering** techniques as a preliminary step before employing global modeling methods has demonstrated improved forecasting outcomes [60]. Furthermore, given the inherent stochasticity of fine granularity energy data, e.g., smart plug and household level data, exploring **probabilistic forecasting** methods is essential. These approaches provide valuable insights on the uncertainty, particularly when dealing with these highly variable profiles. **Ensemble models**, i.e., combining several models, each capturing different dynamics, has also proven useful. Research in these directions has started and is promising, it could only improve prediction accuracy and robustness.

Low-voltage load forecasting with exogenous inputs

There is no consensus yet in the literature on the effect of exogenous data on forecasting performances at the individual household level. A thorough forecasting study on different datasets, applying different benchmark models including statistical, machine learning, and hybrid approaches, to compare results with and without exogenous

data as input could provide valuable insights. Some research has been done in that direction [75] focusing on weather data. Other exogenous data could include traffic data, calendar information, geolocation, etc. The effect of exogenous data on forecasting could also be time series dependent, dataset dependent or even horizon dependent. Developing a framework to assess whether certain exogenous variables should be included in a specific forecasting problem statement would be very valuable.

Meta-model

While various forecasting methods may excel under specific circumstances, none consistently outperforms others across all scenarios. This challenge calls for innovative approaches to model selection that can adapt to individual forecasting tasks. One potential research branch for addressing this issue is the development of a meta-learning framework [60], which would equip the forecasting process with a "super model." This "super model" would be designed to select the most suitable forecasting model for each unique forecasting task, e.g., each unique household profile, or smart plug profile. It is expected that a meta-learning framework may entail additional computational costs, however, it would primarily rely on an offline procedure. This means that the computational overhead is not incurred during real-time forecasting but rather in the training phase.

Comprehensive evaluation of forecasting models

Beyond the traditional metrics of forecasting accuracy, future work should place a greater emphasis on evaluating forecasting methods with respect to their real-world impact. In Chapter 2, where the primary objective is improved payment plans and customer satisfaction, actually assessing the effect of enhanced consumption forecasting on customer satisfaction levels would be interesting. Similarly, in Chapter 3, which focuses on congestion forecasting, investigating the effect of consumption forecasting improvements on congestion outcomes is a pertinent avenue for research. Evaluating models based not only on their forecast performance but also their alignment with the ultimate objectives is under-researched yet essential in real-life applications. Chapter 4 sets foot in that direction by evaluating each step of the pipeline.

Evolution of household dynamics

The recent advancements in household electrification, characterized by the integration of heat pumps, electric vehicles, and solar panels, present a compelling need for updated data, as mentioned earlier. Additionally, it raises new challenges and new forecasting opportunities.

One critical aspect to consider is the consumption patterns of households, which are intrinsically linked to human behavior. However, the incorporation of heat pumps

and solar panels introduces an increased reliance on weather conditions and external environmental factors. This change might settle the consensus on the importance of exogenous data in enhancing the accuracy of consumption forecasts.

Presently, smart meters are capable of recording overall household energy consumption. To gain deeper insights, especially in identifying the presence and individual energy contributions of heat pumps, solar panels, or electric vehicles, the implementation of disaggregation techniques becomes imperative. An alternative, albeit more resource-intensive, approach would be the installation of dedicated meters for each of these installations, offering a granular view of energy usage.

An exciting area of exploration is the development and application of large-scale battery systems. These systems hold the potential to revolutionize household energy management by storing energy sourced from the grid during cost-effective periods. This approach would not only be influenced by human behavioral patterns but also critically by energy pricing, significantly altering the landscape of home energy forecasting. The creation of new forecasting models that can predict optimal times for energy draw, storage, and utilization, particularly for charging electric vehicles, is a horizon rich with opportunities.

Additionally, the concept of Vehicle-to-Grid (V2G) systems presents an innovative research avenue. V2G explores the use of electric vehicle batteries as an auxiliary home energy source, particularly during peak demand periods or grid instability. This area promises a future where electric vehicles play a dual role, not only as a mode of transport but also as a dynamic component in home energy systems.

Each of these developments marks a step forward in the field of energy forecasting, offering exciting research opportunities that could reshape our understanding and management of household energy consumption.

5.3.2 Perspectives in the energy landscape

Evolution low-voltage grid infrastructure

The changes in household energy dynamics and the advent of energy communities signify a crucial need for the adaptation of the low-voltage grid. A current issue occurs in scenarios where a street, densely populated with solar panels, generates excess energy. This overproduction can lead to temporary disconnection from the grid to prevent imbalances. Addressing these challenges requires strategic modifications to the grid infrastructure. Distribution system operators, like Fluvius, take measures in this direction. They are replacing and reinforcing their low-voltage grid infrastructure to accommodate a higher electricity throughput than originally designed.

As mentioned in this dissertation, a key tool in this transformative phase is the use of digital twins. These simulations offer invaluable insights into the impact of future changes on the grid and help prioritize the reinforcement of the low-voltage

grid. Future research opportunities lie in refining these models to more accurately simulate complex grid interactions, especially under varied scenarios of renewable energy integration and consumption patterns.

Evolution of billing

The transformation of energy consumption habits and patterns necessitates a corresponding evolution in billing structures. A notable aspect of this shift is the increasing cost implications of consumption peaks on the grid. These fluctuations in demand represent a significant challenge for **grid management and financial modeling**. Regions such as Flanders are pioneering in this respect, having already introduced capacity tariffs [41]. These tariffs are structured to reflect the cost of peak energy demands on the grid infrastructure, thus encouraging more balanced consumption patterns among users.

Another example is the initial phase of photovoltaic panel installations. These present an economic benefit to consumers, who can reduce their electricity bills by injecting surplus energy into the grid. However, this dynamic is changing. As the usage of the grid incurs operational costs for grid operators, future billing models will probably account for these expenses. This shift in policy underlines the need to reassess the economic framework surrounding renewable energy installations and grid usage.

This evolving landscape offers fertile ground for **research into innovative billing models**, and the subsequent changes in price forecasting. The goal is to develop systems that not only ensure the financial viability of grid operations but also promote sustainable energy usage and fair pricing for consumers.

Evolution of policies

The low-voltage grid is increasingly recognized as a vital component in enabling the ongoing energy transition, necessitating proactive engagement from both the state and distribution system operators. These entities are **implementing policies** to accommodate but also anticipate, shifts in energy consumption patterns.

A notable example of such policies is the state's provision of incentives for household energy upgrades, including the installation of solar panels (below 10 kVA), heat pumps, and batteries [43,140]. In 2021, for instance, households received a maximum incentive of 1500 euros for solar panel installations, a figure that was adjusted to 750 euros in 2022 and is projected to decrease further by 2024. While incentives for battery installations were available until April 2023, those for heat pump installations continue to be offered. With the objective to reduce dependence on fossil fuels, significant policy shifts have been made, such as the 2021 mandate that prohibits gas connections in large new constructions [140]. These initiatives represent just a fraction of the broader effort to transition to sustainable energy sources.

Moreover, the cyclical process of policy formulation, implementation, evaluation, and revision is reliant on research and empirical evidence. Research institutions, like VITO and Energyville, play an important role in this cycle, conducting studies and providing **data-driven policy recommendations**. Their expertise and findings inform decision-making, ensuring that policies are scientifically sound.

Privacy and data protection

In energy management, particularly with the increasing granularity of data collection from household and plug usage, privacy and data protection emerge as critical concerns. The detailed information gathered from such data sources can reveal intimate aspects of individual behaviors and lifestyles.

The primary challenge lies in ensuring that the systems responsible for recording, transmitting, and storing this data uphold the **highest standards of security**. It is imperative that access to this data is restricted exclusively to authorized organizations, and that robust safeguards are in place to prevent malicious use of the information. For instance, as outlined in Chapter 4, data from monitored plug usage, such as a screen monitor, could potentially be misinterpreted. If a monitor is used for only two hours a day, it might wrongly suggest the limited desk presence of an employee. Such misinterpretations, if used as a basis for employment decisions, could lead to unwarranted consequences. Similarly, household energy usage patterns can inadvertently lead to profiling, revealing details such as occupancy at certain times, routine activities, or even changes in the number of residents. The potential misuse of such sensitive information, through unauthorized sale or theft, poses a significant threat to privacy.

Moreover, the question of **legal boundaries** in accessing this data for state investigations parallels similar issues faced in telecommunications. An example of this could be law enforcement using energy consumption data to ascertain whether someone was present at home at a specific time, potentially as a part of establishing or refuting an alibi, which introduces nuanced ethical and privacy considerations.

Addressing these challenges requires not only technological solutions but also a **comprehensive legal framework**. The opportunity lies in research and policy development that effectively balances the benefits of advanced energy management systems with the imperative of protecting individual privacy.

Energy inequalities

An additional concern within the energy sector is the risk of energy inequalities. Recent research on the installation and use of smart meter technologies has shed light on how they may inadvertently exacerbate disparities in energy access and usage.

In the United Kingdom, studies [123] reveal research has shown that the **benefits of smart meter technologies** have not been evenly distributed across all demographics.

Particularly the elderly, the ill, those with lower educational backgrounds, residents in social housing, and rural communities often find themselves less equipped to engage with and benefit from smart meter technologies. For instance, elderly residents in rural Scotland displayed limited interaction with and understanding of smart energy monitors, indicating a gap in the effective utilization of these tools for energy management.

In the United States, **disparities in the roll-out** of smart meter programs have been observed [153]. Communities with larger Hispanic populations and lower education levels were less likely to be included in smart meter roll-outs. Furthermore, high-usage customers tend to benefit more from net metering programs compared to their low-usage counterparts, suggesting an unintentional favoring of more affluent households.

These findings point to a pressing need for **research focused on ensuring equitable access to and benefits from smart metering**. The goal is to develop policies and strategies that not only recognize but actively bridge these gaps. Future research could explore innovative ways to tailor smart metering systems and energy policies to be more inclusive and supportive of vulnerable and disadvantaged groups. This is not only vital for achieving energy justice but also crucial for the successful transition to a more sustainable and efficient energy future.

Bibliography

[1] Abdullah, A. N., Mutlag, A. H., and Ahmed, M. S. Neural Network Based Home Energy Management for Modelling and Controlling Home Appliances under Demand Response. *Journal of Physics: Conference Series 1963*, 1 (2021).

[2] Acker, B., Duarte, C., and Wymelenberg, K. V. D. Office Space Plug Load Profiles and Energy Saving Interventions. *ACEEE Summer Study on Energy Efficiency in Buildings* (2012), 1–13.

[3] Agnieszka, W. EU progress towards Sustainable Development Goal on energy (SDG 7). Tech. Rep. July, European Union, 2023. `https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2023)751402`.

[4] Ahmad, T., Zhang, H., and Yan, B. A review on renewable energy and electricity requirement forecasting models for smart grid and buildings. *Sustainable Cities and Society 55*, April 2019 (2020), 102052.

[5] Ahmed, M. S., Mohamed, A., Homod, R. Z., and Shareef, H. Hybrid LSA-ANN based home energy management scheduling controller for residential demand response strategy. *Energies 9*, 9 (2016).

[6] Ahmed, M. S., Mohamed, A., Homod, R. Z., Shareef, H., Sabry, A. H., and Bin Khalid, K. Smart plug prototype for monitoring electrical appliances in Home Energy Management System. *2015 IEEE Student Conference on Research and Development, SCOReD 2015* (2015), 32–36.

[7] Ahmed, M. S., Mohamed, A., Shareef, H., Homod, R. Z., and Ali, J. A. Artificial neural network based controller for home energy management considering demand response events. *2016 International Conference on Advances in Electrical, Electronic and Systems Engineering, ICAEES 2016* (2017), 506–509.

[8] Alhussein, M., Aurangzeb, K., and Haider, S. I. Hybrid CNN-LSTM Model for Short-Term Individual Household Load Forecasting. *IEEE Access 8* (2020), 180544–180557.

[9] Amarasinghe, K., Marino, D. L., and Manic, M. Deep neural networks for energy load forecasting. *IEEE International Symposium on Industrial Electronics* (2017), 1483–1488.

[10] Ambati, L. P. R., and Irwin, D. AutoPlug: An automated metadata service for smart outlets. *2016 7th International Green and Sustainable Computing Conference, IGSC 2016* (2017), 1–8.

[11] Arpogaus, M. Short-Term Probabilistic Load Forecasting using Conditioned Bernstein-Polynomial Normalizing Flows (STPLF-BNF), 2021. `https://github.com/MArpogaus/stplf-bnf`.

[12] Arpogaus, M., Voss, M., Sick, B., Nigge-Uricher, M., and Dürr, O. Probabilistic short-term low-voltage load forecasting using bernstein-polynomial normalizing flows. In *ICML 2021 Workshop on Tackling Climate Change with Machine Learning* (2021), pp. 1–6.

[13] Arpogaus, M., Voss, M., Sick, B., Nigge-Uricher, M., and Dürr, O. Short-Term Density Forecasting of Low-Voltage Load using Bernstein-Polynomial Normalizing Flows. *arXiv* (2022), 1–10.

[14] Bandara, K., Godahewa, R., and Hewamalage, H. IEEE CIS Technical Challenge 2020 Git Hub Repository, 2020. `https://github.com/rakshitha123/IEEE_CIS_Comp`.

[15] Bandara, K., Godahewa, R., and Hewamalage, H. A Scalable Ensemble of Global and Local Models for Long-term Energy Demand Forecasting. In *41st International Symposium on Forecasting* (2021).

[16] Barolli, L. *Short term load forcasting using XGBoost*, vol. 1. Springer International Publishing, 2019.

[17] Ben Taieb, S., and Hyndman, R. J. A gradient boosting approach to the Kaggle load forecasting competition. *International Journal of Forecasting 30*, 2 (2014), 382–394.

[18] Botman, L. Python implementation: a global probabilistic approach for short-term forecasting of individual households' electricity consumption, 2022. `https://gitlab.esat.kuleuven.be/Lola.Botman/gpsthl-forecasting/`.

[19] Botman, L., Lago, J., Becker, T., Agudelo, O. M., Vanthournout, K., and De Moor, B. A scalable method for probabilistic short-term forecasting of individual households consumption in low voltage grids. In *2023 IEEE PES Grid Edge Technologies Conference & Exposition (Grid Edge)* (San Diego, 4 2023), IEEE.

[20] Botman, L., Lago, J., Fu, X., Chia, K., Wolf, J., Kleissl, J., and De Moor, B. Python implementation: Smart Plug Pipeline. `https://gitlab.esat.kuleuven.be/Lola.Botman/smart-plug-pipeline`.

[21] Botman, L., Soenen, J., Theodorakos, K., Yurtman, A., Bekker, J., Vanthournout, K., Blockeel, H., Fellow, B. D. M., and Lago, J. A scalable ensemble approach to forecast the electricity consumption of households. *IEEE Transactions on Smart Grid 14*, 1 (2023), 757–768.

[22] Bouyeddou, B., Kadri, B., Harrou, F., and Sun, Y. DDOS-attacks detection using an efficient measurement-based statistical mechanism. *Engineering Science and Technology, an International Journal 23*, 4 (2020), 870–878.

[23] Boylan, J. E., and Syntetos, A. A. *Intermittent Demand Forecasting*. John Wiley & Sons, Ltd, 2021.

[24] Bracher, J., Ray, E. L., Gneiting, T., and Reich, N. G. Evaluating epidemic forecasts in an interval format. *PLoS Computational Biology 17*, 2 (2021), 1–15.

[25] Bramer, M. Ensemble Classification. In *Principles of Data Mining*. Springer London, London, 2013, pp. 209–220.

[26] Braun, M. R., Altan, H., and Beck, S. B. Using regression analysis to predict the future energy consumption of a supermarket in the UK. *Applied Energy 130* (2014), 305–313.

[27] Cetinkaya, N. Long-term electrical load forecasting based on economic and demographic data for Turkey. In *CINTI 2013 - 14th IEEE International Symposium on Computational Intelligence and Informatics, Proceedings* (2013), IEEE, pp. 219–223.

[28] Chen, T., and Guestrin, C. XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 13-17-Augu* (2016), 785–794.

[29] Chia, K., Lebar, A., Agarwal, V., Lee, M., Ikedo, J., Wolf, J., Trenbath, K., and Kleissl, J. Integration of a Smart Outlet-Based Plug Load Management System with a Building Automation System. *2023 IEEE PES Grid Edge Technologies Conference and Exposition, Grid Edge 2023* (2023).

[30] Choi, M., Park, W. K., and Lee, I. Smart Office Energy-Saving Service Using Bluetooth Low Energy Beacons and Smart Plugs. *Proceedings - 2015 IEEE International Conference on Data Science and Data Intensive Systems; 8th IEEE International Conference Cyber, Physical and Social Computing; 11th IEEE International Conference on Green Computing and Communications and 8th IEEE Inte* (2015), 247–251.

[31] Chujai, P., Kerdprasop, N., and Kerdprasop, K. Time series analysis of household electric consumption with ARIMA and ARMA models. *Lecture Notes in Engineering and Computer Science 2202* (2013), 295–300.

[32] Commission for Energy Regulation (CER). CER smart metering project - electrical customer behaviour trial, 2012. `https://www.ucd.ie/issda/data/commissionforenergyregulationcer/`.

[33] Cover, T. M., and Hart, P. E. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory 13*, 1 (1967), 21–27.

[34] CREG. Commission de Régulation de l'Électricité et du Gaz. `https://www.creg.be/en`.

[35] Dang, S., Peng, L., Zhao, J., Li, J., and Kong, Z. A Quantile Regression Random Forest-Based Short-Term Load Probabilistic Forecasting Method. *Energies 15*, 2 (2022), 1–20.

[36] De Gooijer, J. G., and Hyndman, R. J. 25 Years of Time Series Forecasting. *International Journal of Forecasting 22*, 3 (2006), 443–473.

[37] Di Gesù, V., and Starovoitov, V. Distance-based functions for image comparison. *Pattern Recognition Letters 20*, 2 (1999), 207–214.

[38] Doherty, B., and Trenbath, K. Device-level plug load disaggregation in a zero energy office building and opportunities for energy savings. *Energy and Buildings 204* (2019).

[39] Ebrahim, A. F., and Mohammed, O. Household load forecasting based on a pre-processing non-intrusive load monitoring techniques. *IEEE Green Technologies Conference 2018-April* (2018), 107–114.

[40] Elia. Elia Website (Belgian Transmission System Operator). `https://www.elia.be/en/company`.

[41] Engie. Capacity Tariffs. `https://www.engie.be/nl/blog/allerlei/capaciteitstarief-Vlaanderen/`.

[42] European Parliament. Energy performance of buildings 2010/31/EU Directive. Tech. rep., European Parliament, 2021. `https://data.europa.eu/doi/10.2861/05615`.

[43] Fluvius. Premie voor zonnepanelen, 0. `https://www.fluvius.be/nl/premies/premies-voor-huishoudelijke-klanten/premie-zonnepanelen`.

[44] Fluvius. Honderd geanonimiseerde 15' verbruiksprofielen elektriciteit, 2021. `https://opendata.fluvius.be/explore/dataset/1_50-aanlevering-dmk-e-kwartierwaarden/information/`.

[45] Fu, C. W., and Nguyen, T. T. Models for Long-Term Energy Forecasting. In *2003 IEEE Power Engineering Society General Meeting, Conference Proceedings* (2003), pp. 235–239.

[46] Fumo, N., and Rafe Biswas, M. A. Regression analysis for prediction of residential energy consumption. *Renewable and Sustainable Energy Reviews 47* (2015), 332–343.

[47] Gandhi, P., and Brager, G. S. Commercial office plug load energy consumption trends and the role of occupant behavior. *Energy and Buildings 125* (2016), 1–8.

[48] Gerossier, A., Girard, R., Bocquet, A., and Kariniotakis, G. Robust day-ahead forecasting of household electricity demand and operational challenges. *Energies 11*, 12 (2018).

[49] Ghatikar, G., Cheung, I., and Lanzisera, S. Miscellaneous and electronic loads energy efficiency opportunities for commercial buildings : a collaborative study by the United States and India. Tech. rep., Lawrence Berkeley National Laboratory, 2013. `https://escholarship.org/uc/item/80b1c401`.

[50] Gneiting, T., Balabdaoui, F., and Raftery, A. E. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society. Series B: Statistical Methodology 69*, 2 (2007), 243–268.

[51] Gneiting, T., and Katzfuss, M. Probabilistic forecasting. *Annual Review of Statistics and Its Application 1* (2014), 125–151.

[52] Gneiting, T., and Raftery, A. E. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*, 477 (2007), 359–378.

[53] Goshtasby, A. A. Similarity and Dissimilarity Measures. In *Image Registration: Principles, Tools and Methods*. Springer London, London, 2012, pp. 7–66.

[54] Grabner, M., Wang, Y., Wen, Q., Blažič, B., and Štruc, V. A global modeling approach for load forecasting in distribution networks. *IEEE Transactions on Smart Grid PP* (2022), 1.

[55] Gunay, H. B., O'Brien, W., Beausoleil-Morrison, I., and Gilani, S. Modeling plug-in equipment load patterns in private office spaces. *Energy and Buildings 121* (2016), 234–249.

[56] Haben, S., Arora, S., Giasemidis, G., Voss, M., and Vukadinović Greetham, D. Review of low voltage load forecasting: Methods, applications, and recommendations. *Applied Energy 304* (2021), 1–33.

[57] Haben, S., Voss, M., and Holderbaum, W. *Core concepts and methods in load forecasting*, 1 ed. Springer International Publishing, 2023.

[58] Haben, S., Ward, J., Vukadinovic Greetham, D., Singleton, C., and Grindrod, P. A new error measure for forecasts of household-level, high resolution electrical energy consumption. *International Journal of Forecasting 30*, 2 (2014), 246–256.

[59] Hafer, M. Quantity and electricity consumption of plug load equipment on a university campus. *Energy Efficiency 10*, 4 (2017), 1013–1039.

[60] Hamoudia, M., Makridakis, P., and Spiliotis, E., Eds. *Forecasting With Artificial Neural Theory and Applications*, 1 ed. Springer Nature, 2023.

[61] Han, J., Lee, H., and Park, K. R. Remote-Controllable and Energy-Saving Room Architecture based on ZigBee Communication. *IEEE Transactions on Consumer Electronics 55*, 1 (2009), 264–268.

[62] Henriot, P. Unleashing the benefits of data for energy systems, 2023. `https://www.iea.org/commentaries/unleashing-the-benefits-of-data-for-energy-systems`.

[63] Heo, S., Song, S., Kim, B., and Kim, H. Sharing-aware Data Acquisition Scheduling for Multiple Rules in the IoT. *Proceedings of the IEEE Real-Time and Embedded Technology and Applications Symposium, RTAS 2020-April* (2020), 43–55.

[64] Hong, T., and Fan, S. Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting 32*, 3 (2016), 914–938.

[65] Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., and Hyndman, R. J. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *International Journal of Forecasting 32*, 3 (2016), 896–913.

[66] Hong, T., and Wang, P. Artificial Intelligence for Load Forecasting: History, Illusions, and Opportunities. *IEEE Power and Energy Magazine 20*, 3 (2022), 14–23.

[67] Hussain, R. H., and Shujaa, M. I. IoT residential appliance monitoring and controlling system using MQTT protocol. *the Fourth Scientific Conference for Electrical Engineering Techniques Research (Eetr2022) 2804*, February (2023), 020027.

[68] Hyndman, R. J., and Fan, Y. Sample quantiles in statistical packages. *The American Statistician 50*, 4 (11 1996), 361.

[69] IEA - International Energy Agency. Belgium 2022: Energy Policy Review. Tech. rep., 2021. `iea.blob.core.windows.net/assets/638cb377-ca57-4c16-847d-ea4d96218d35/Belgium2022_EnergyPolicyReview.pdf`.

[70] Jain, R. K., Smith, K. M., Culligan, P. J., and Taylor, J. E. Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Applied Energy 123* (2014), 168–178.

[71] Kaack, L. H., Apt, J., Morgan, M. G., and McSharry, P. Empirical prediction intervals improve energy forecasting. *Proceedings of the National Academy of Sciences of the United States of America 114*, 33 (2017), 8752–8757.

[72] Kalluri, B., Kamilaris, A., Kondepudi, S., and Wei, K. H. Classifying Office Plug Load Appliance Events in the context of NILM using Time-series Data Mining. *Proceedings of the 12th REHVA World Congress and CLImA, Aalborg, DenmarN 6* (2016), 1–11.

[73] Kalluri, B., Kondepudi, S., Wei, K. H., Wai, T. K., and Kamilaris, A. OPLD: Towards improved non-intrusive office plug load disaggregation. *2015 IEEE International Conference on Building Energy Efficiency and Sustainable Technologies, ICBEST 2015 5* (2015), 56–61.

[74] Kamilaris, A., Kalluri, B., Kondepudi, S., and Kwok Wai, T. A literature survey on measuring energy usage for miscellaneous electric loads in offices and commercial buildings. *Renewable and Sustainable Energy Reviews 34* (2014), 536–550.

[75] Kang, J., and Reiner, D. M. What is the effect of weather on household electricity consumption? Empirical evidence from Ireland. *Energy Economics 111*, March (2022), 106023.

[76] Kawamoto, K., Koomey, J. G., Nordman, B., Brown, R. E., Piette, M. A., Ting, M., and Meier, A. K. Electricity used by office equipment and network equipment in the US. *Energy 27*, 3 (2002), 255–269.

[77] Kim, T. Y., and Cho, S. B. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy 182* (2019), 72–81.

[78] Koenker, R., and Hallock, K. F. Quantile regression. *Journal of Economic Perspectives 15*, 4 (2001), 143–156.

[79] Kong, W., Dong, Z. Y., Hill, D. J., Luo, F., and Xu, Y. Short-term residential load forecasting based on resident behaviour learning. *IEEE Transactions on Power Systems 33*, 1 (2018), 2017–2018.

[80] Krakov, D., and Feitelson, D. G. Comparing performance heatmaps. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8429 LNCS* (2014), 42–61.

[81] Kriechbaumer, T., and Jacobsen, H. A. BLOND, a building-level office environment dataset of typical electrical appliances. *Scientific Data 5* (2018), 1–14.

[82] Kuster, C., Rezgui, Y., and Mourshed, M. Electrical load forecasting models: A critical systematic review. *Sustainable Cities and Society 35*, June (2017), 257–270.

[83] Lago, J., De Ridder, F., and De Schutter, B. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy 221*, November 2017 (2018), 386–405.

[84] Langner, R., and Trenbath, K. Integrating Smart Plug and Process Load Controls into Energy Management Information System Platforms: A Landscaping Study. Tech. rep., National Renewable Energy Laboratory NREL, 2019. `www.nrel.gov/publications.`

[85] Lanzisera, S., Dawson-Haggerty, S., Cheung, H. Y., Taneja, J., Culler, D., and Brown, R. Methods for detailed energy data collection of miscellaneous and electronic loads in a commercial office building. *Building and Environment 65* (2013), 170–177.

[86] Leonardi, A., Ziekow, H., Konchalenkov, D., and Rogozina, A. Detecting Smart Plug Configuration Changes in Smart Homes. *Smart SysTech 2014; European Conference on Smart Objects, Systems and Technologies* (2014), 1–7.

[87] Lin, G. Y., Lee, S. C., Hsu, J. Y. J., and Jih, W. R. Applying power meters for appliance recognition on the electric panel. *Proceedings of the 2010 5th IEEE Conference on Industrial Electronics and Applications, ICIEA 2010* (2010), 2254–2259.

[88] Lindberg, K. B., Seljom, P., Madsen, H., Fischer, D., and Korpås, M. Long-term electricity load forecasting: Current and future trends. *Utilities Policy 58*, May (2019), 102–119.

[89] Liu, B., Nowotarski, J., Hong, T., and Weron, R. Probabilistic load forecasting via quantile regression averaging on sister forecasts. *IEEE Transactions on Smart Grid 8*, 2 (2017), 730–737.

[90] Liu, G., Reda, F. A., Shih, K. J., Wang, T. C., Tao, A., and Catanzaro, B. Image Inpainting for Irregular Holes Using Partial Convolutions. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11215 LNCS* (2018), 89–105.

[91] Lloyd, S. P. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory 28*, 2 (1982), 129–137.

[92] Ma, Y., Yu, J. Q., Yang, C. Y., and Wang, L. Study on power energy consumption model for large-scale public building. In *Proceedings - 2010 2nd International Workshop on Intelligent Systems and Applications, ISA 2010* (2010), IEEE, pp. 10–13.

[93] MacQueen, J., et al. Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability 1*, 14 (1967), 281–297.

[94] Maltais, L. G., and Gosselin, L. Forecasting of short-term lighting and plug load electricity consumption in single residential units: Development and assessment of data-driven models for different horizons. *Applied Energy 307*, June 2021 (2022), 118229.

[95] Masoso, O. T., and Grobler, L. J. The dark side of occupants' behaviour on building energy use. *Energy and Buildings 42*, 2 (2010), 173–177.

[96] McLachlan, G. J., and Basford, K. E. *Mixture models: Inference and applications to clustering*. JSTOR, 1988.

[97] McSharry, P. E., Bouwman, S., and Bloemhof, G. Probabilistic forecasts of the magnitude and timing of peak electricity demand. *IEEE Transactions on Power Systems 20*, 2 (2005), 1166–1172.

[98] Menezes, A. C., Cripps, A., Buswell, R. A., Wright, J., and Bouchlaghem, D. Estimating the energy consumption and power demand of small power equipment in office buildings. *Energy and Buildings 75* (2014), 199–209.

[99] Moreno Jaramillo, A. F., Laverty, D. M., Morrow, D. J., Martinez del Rincon, J., and Foley, A. M. Load modelling and non-intrusive load monitoring to integrate distributed energy resources in low and medium voltage networks. *Renewable Energy 179* (2021), 445–466.

[100] Morsali, H., Shekarabi, S. M., Ardekani, K., Khayami, H., Fereidunian, A., Ghassemian, M., and Lesani, H. Smart plugs for building energy management systems. *2012 2nd Iranian Conference on Smart Grids, ICSG 2012* (2012), 1–5.

[101] Nowotarski, J., and Weron, R. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews 81*, September 2016 (2018), 1548–1568.

[102] Olofsson, T., and Andersson, S. Long-term energy demand predictions based on short-term measured data. *Energy and Buildings 33*, 2 (2001), 85–91.

[103] Parti, M., and Parti, C. Total and Appliance-Specific Conditional Demand for Electricity in the Household Sector. *Bell Journal of Economics 11*, 1 (1980), 309–321.

[104] Pedregosa, F., Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., and Mueller, A. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research 12*, 1 (2011), 2825–2830.

[105] Pessanha, J. F. M., and Leon, N. Forecasting long-term electricity demand in the residential sector. *Procedia Computer Science 55*, Itqm (2015), 529–538.

[106] Petneházi, G. Recurrent neural networks for time-series forecasting. *arXiv* (2019).

[107] R. E. Abdel-Aal;, and Al-Garni, A. Z. Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate time-series analysis. *Energy 22*, I (1997), 1059–1069.

[108] Rice, A., Hay, S., and Ryder-Cook, D. A limited-data model of building energy consumption. In *BuildSys'10 - Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings* (2010), pp. 67–72.

[109] Rinaldi, G., Garcia, F. C., Agudelo, O. M., Becker, T., Vanthournout, K., Mestdagh, W., and De Moor, B. A framework for a data quality module in decision support systems: an application with smart grid time series. *International Conference on Enterprise Information Systems, ICEIS - Proceedings 1*, Iceis (2023), 443–452.

[110] Ruppert, D., and Matteson, D. S. *Statistics and data analysis for financial engineering*. Springer Texts in Statistics. Springer New York, New York, NY, 2015.

[111] Ruzzelli, A. G., Nicolas, C., Schoofs, A., and O'Hare, G. M. Real-time recognition and profiling of appliances through a single electricity sensor. *SECON 2010 - 2010 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks* (2010), 1–9.

[112] Sandys, L., Dobson, R., Brown, E., Graham, G., Lane, R., and Verma, J. Energy Data Taskforce: A Strategy for a Modern Digitalised Energy System - Energy Systems Catapult. *Energy Data Taskforce report* (2019), 1–56.

[113] Santos, A., Duggan, G. P., Davis, J., and Zimmerle, D. A Cautionary Note on Using Smart Plugs for Research Data Acquisition. *e-Prime - Advances in Electrical Engineering, Electronics and Energy 4*, December 2022 (2023), 100137.

[114] Schwermer, R., Buchberger, J., Mayer, R., and Jacobsen, H. A. Federated office plug-load identification for building management systems. *e-Energy 2022 - Proceedings of the 2022 13th ACM International Conference on Future Energy Systems* (2022), 114–126.

[115] Selim, S. Z., and Ismail, M. A. K-Means-Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6*, 1 (1984), 81–87.

[116] Shao, X., Kim, C. S., and Sontakke, P. Accurate deep model for electricity consumption forecasting using multi-channel and multi-scale feature fusion CNN-LSTM. *Energies 13*, 8 (2020).

[117] Sharda, S., Sharma, K., and Singh, M. A real-time automated scheduling algorithm with PV integration for smart home prosumers. *Journal of Building Engineering 44*, June (2021), 102828.

[118] Shareef, H., Ahmed, M. S., Mohamed, A., and Al Hassan, E. Review on Home Energy Management System Considering Demand Responses, Smart Technologies, and Intelligent Controllers. *IEEE Access 6* (2018), 24498–24509.

[119] Shepero, M., van der Meer, D., Munkhammar, J., and Widén, J. Residential probabilistic load forecasting: A method using Gaussian process designed for electric load data. *Applied Energy 218*, March (2018), 159–172.

[120] Siebert, L. C., Ferreira, L. R., Yamakawa, E. K., Custódio, E. S., Aoki, A. R., Fernandes, T. S., and Cardoso, K. H. Centralized and decentralized approaches to demand response using smart plugs. *Proceedings of the IEEE Power Engineering Society Transmission and Distribution Conference* (2014), 1–5.

[121] Sinaga, K. P., and Yang, M. S. Unsupervised K-means clustering algorithm. *IEEE Access 8* (2020), 80716–80727.

[122] Soenen, J., Yurtman, A., Becker, T., D'hulst, R., Vanthournout, K., Meert, W., and Blockeel, H. Scenario generation of residential electricity consumption through sampling of historical data. *Sustainable Energy, Grids and Networks 34* (2023), 100985.

[123] Sovacool, B. K., Kivimaa, P., Hielscher, S., and Jenkins, K. Vulnerability and resistance in the United Kingdom's smart meter transition. *Energy Policy 109*, July (2017), 767–781.

[124] Spiliotis, E., Makridakis, S., Semenoglou, A. A., and Assimakopoulos, V. Comparison of statistical and machine learning methods for daily SKU demand forecasting. *Operational Research 22*, 3 (2022), 3037–3061.

[125] Srinivasan, A., Baskaran, K., and Yann, G. IoT Based Smart Plug-Load Energy Conservation and Management System. *2019 IEEE 2nd International Conference on Power and Energy Applications, ICPEA 2019* (2019), 155–158.

[126] Student. The Probable Error of a Mean. *Biometrika 6*, 1 (1908), 1–25.

[127] Tekler, Z. D., Low, R., Zhou, Y., Yuen, C., Blessing, L., and Spanos, C. Near-real-time plug load identification using low-frequency power data in office spaces: Experiments and applications. *Applied Energy 275*, May (2020).

[128] Theodorakos, K., Agudelo, O. M., Espinoza, M., and De Moor, B. Decomposition-residuals neural networks: hybrid system identification applied to electricity demand forecasting. *IEEE Open Access Journal of Power and Energy 9*, October 2021 (2022), 241–253.

[129] Thorndike, R. L. Who belongs in the family? *Psychometrika 18*, 4 (1953), 267–276.

[130] Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological) 58*, 1 (1 1996), 267–288.

[131] Trenbath, K., Tuttle, R., Maisha, K., and LeBar, A. Assessing and Reducing Plug and Process Loads in Office Buildings. Tech. rep., National Renewable Energy Laboratory, 2020. `https://www.osti.gov/biblio/1530714/`.

[132] Triguero, I. IEEE-CIS technical challenge on energy prediction from smart meter data, 2020. `https://ieee-dataport.org/competitions/ieee-cis-technical-challenge-energy-prediction-smart-meter-data`.

[133] UK Power Networks. Smartmeter energy consumption data in london households, 2013. https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households.

[134] Ul Abideen, M. Z., Ellabban, O., and Al-Fagih, L. A review of the tools and methods for distribution networks' hosting capacity calculation. *Energies 13*, 11 (2020), 1–25.

[135] Uniejewski, B., Nowotarski, J., and Weron, R. Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies 9*, 8 (8 2016), 621.

[136] United Nations. Ensure access to affordable, reliable, sustainable and modern energy for all, 2015. https://sdgs.un.org/goals/goal7.

[137] Uribe-Pérez, N., Hernández, L., de la Vega, D., and Angulo, I. State of the Art and Trends Review of Smart Metering in Electricity Grids. *Applied Sciences 6*, 3 (2016), 1–24.

[138] van Wyk, A. Encoding cyclical features for deep learning. https://www.avanwyk.com/encoding-cyclical-features-for-deep-learning/%0A.

[139] Venkat, K., Gautam, T., Yadav, M., and Singh, M. An XGBoost Ensemble Model for Residential Load Forecasting. In *Proceedings of International Conference on Intelligent Computing, Information and Control Systems* (2021), pp. 321–334.

[140] Vlaanderen.be. Bouwen, wonen en energie. https://www.vlaanderen.be/bouwen-wonen-en-energie.

[141] Wang, S., Deng, X., Chen, H., Shi, Q., and Xu, D. A bottom-up short-term residential load forecasting approach based on appliance characteristic analysis and multi-task learning. *Electric Power Systems Research 196*, April (2021).

[142] Wang, S., Kim, A. A., and Johnson, E. M. Understanding the deterministic and probabilistic business cases for occupant based plug load management strategies in commercial office buildings. *Applied Energy 191* (2017), 398–413.

[143] Wang, Y., Gan, D., Sun, M., Zhang, N., Lu, Z., and Kang, C. Probabilistic individual load forecasting using pinball loss guided LSTM. *Applied Energy 235*, November 2018 (2019), 10–20.

[144] Wang, Y., Zhang, N., Tan, Y., Hong, T., Kirschen, D. S., and Kang, C. Combining probabilistic load forecasts. *IEEE Transactions on Smart Grid 10*, 4 (2019), 3664–3674.

[145] Wang, Z., Hong, T., and Piette, M. A. Predicting plug loads with occupant count data through a deep learning approach. *Energy 181* (2019), 29–42.

[146] Webber, C. a., Roberson, J. a., Brown, R. E., Payne, C. T., Nordman, B., and Koomey, J. G. Field Surveys of Office Equipment Operating Patterns. Tech. rep., Lawrence Berkeley National Laboratory, 2001. https://escholarship.org/uc/item/3r7480zm.

[147] Wu, W., Bandara, K., Godahewa, R., and Hewamalage, H. IEEE CIS Technical Challenge 2020 Git Hub Repository, 2020.

https://github.com/rakshitha123/IEEE_CIS_Comphttps://github.com/waylongo/cis-challenge-energy-prediction.

[148] Yang, D., Wu, E., and Kleissl, J. Operational solar forecasting for the real-time market. *International Journal of Forecasting 35*, 4 (2019), 1499–1519.

[149] Yao, R., and Steemers, K. A method of formulating energy load profile for domestic buildings in the UK. *Energy and Buildings 37*, 6 (2005), 663–671.

[150] Zhang, X. M., Grolinger, K., Capretz, M. A., and Seewald, L. Forecasting residential energy consumption: single household perspective. In *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018* (2019), IEEE, pp. 110–117.

[151] Zhao, H., Gallo, O., Frosio, I., and Kautz, J. Loss Functions for Image Restoration With Neural Networks. *IEEE Transactions on Computational Imaging 3*, 1 (2017), 47–57.

[152] Zhao, H. X., and Magoulès, F. A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews 16*, 6 (2012), 3586–3592.

[153] Zhou, S., and Noonan, D. S. Justice implications of clean energy policies and programs in the United States: A theoretical and empirical exploration. *Sustainability (Switzerland) 11*, 3 (2019).

[154] Zhou, Z. H. *Ensemble methods: Foundations and algorithms*. CRC Press, 2012.



FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF ELECTRICAL ENGINEERING
STADIUS
Kasteelpark Arenberg 10, 3001 Leuven
B-3001 Leuven
https://www.esat.kuleuven.be/stadius