Universitat Politècnica de Catalunya

Departament d'Enginyeria Elèctrica

Departament d'Enginyeria Elèctrica
UNIVERSITAT POLITÈCNICA DE CATALUNYA

CITCEA - Centre d'Innovació Tecnològica en Convertidors Estàtics i Accionaments

Doctoral Thesis

Energy management systems for smart homes and local energy communities based on optimization and artificial intelligence techniques

Sara Barja Martinez

Thesis advisors:
Dr. Mònica Aragüés Peñalba (UPC, Spain)
Dr. Roberto Villafáfila Robles (UPC, Spain)

Examination Committee:
Prof. Hossein Farahmand (NTNU, Norway)
Dr. Vincent Debusschere (Grenoble INP, France)
Dr. Daniel Montesinos Miracle (UPC, Spain)

Thesis reviewers:
Prof. Hossein Farahmand (NTNU, Norway)
Dr. Vincent Debusschere (Grenoble INP, France)

Barcelona, 7th July 2023

Universitat Politècnica de Catalunya
Departament d'Enginyeria Elèctrica
Centre d'Innovació Tecnològica en Convertidors Estàtics i Accionaments
ETSEIB, Building G
Av. Diagonal 647
08028, Barcelona

First printing, July 2023

“To be or not to be is not the question.
The vital question is how to be and how not to be.”
– Abraham Joshua Heschel

“If you get tired, learn to rest, not to quit”
– Banksy

“Á miña nai e ao meu pai,
Mar máis Alfonso;
e a Macià,
por ser e estar.”

Acknowledgements

Esta tese é dedicada aos meus pais e familia. Grazas por todo o apoio e estar sempre. A Macià, el meu company de vida indispensable i extraordinari. A todos mis amigos del colegio, universidad y a aquellos que he ido sumando en diferentes etapas y lugares de mi vida. A mis amigos del CITCEA-UPC y a mis compañeros de la EPS con los que me vine a Barcelona. Siempre se hace más llevadero el camino con tan buena compañía como vosotros. Sois la familia que he formado aquí, y os estoy agradecida por ello.

Quiero expresar también mi agradecimiento a mis directores de tesis, Mònica y Roberto, por el seguimiento, apoyo y aprendizaje durante esta etapa. Extiendo esta dedicatoria y recuerdo a todos los profesores y profesoras que he tenido a lo largo de mi vida, desde mi infancia hasta la actualidad, ya que gracias a ellos sé todo lo que sé, inculcándome siempre las ganas de aprender y fomentando en mí la curiosidad.

To Fei and Adrià, for their dedicated time and turning my doctoral stay in London into an unforgettable learning experience. To my colleagues on the *11th floor* of Imperial College London, for all the adventures, laughs and lifelong friendships that were formed. It was a privilege to have crossed paths with all of you in space and time.

Para concluir, me gustaría tener unas palabras de agradecimiento a mi yo del pasado. Recuerdo cuando en el último año de instituto se me desaconsejó estudiar ingeniería industrial porque era muy difícil. Ahora, unos cuantos años más tarde, quiero dedicar esta tesis a esa yo adolescente por haber creído y confiado en sí misma.

Financial support

This thesis has been supported by the Universitat Politècnica de Catalunya and CITCEA-UPC. The author has received funds from the following public entities:

- BD4OPEM H2020 project, which has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 872525
- Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya under the FI-AGAUR Pre-doctoral grants and scholarships.

Thesis dissertation peer-review

The following two external referees have reviewed the thesis dissertation:

- Prof. Hossein Farahmand (NTDU, Norway)
- Dr. Vincent Debusschere (Grenoble INP, France)

The author would like to thank the reviewers for their dedication and time reviewing this thesis. Their valuable comments have greatly contributed to improving the manuscript.

Abstract

The rapid advancement of digitization, combined with the integration of renewable generation and the development of information and communication technologies within the distribution network, is accelerating the transition towards distributed, digitized, and decarbonized smart energy systems. Despite the crucial role of data in this transformation, it is still underutilized. Artificial Intelligence (AI) technology has the potential to extract valuable insights from these data, enabling innovative energy services and improving the performance of existing ones.

This thesis first explores the potential of AI in data-driven energy services for distribution power systems through a comprehensive study. After reviewing the state of the art, the next step focuses on energy management systems at the household level. This thesis develops a multi-objective energy management system that simultaneously minimizes greenhouse gas emissions and electricity expenses, considering the entire life cycle of the generation assets used to provide energy, including the grid. The results demonstrate that this methodology can significantly reduce greenhouse gas emissions without incurring expensive electricity costs. The following chapter extends this innovative environmental-based approach to local energy communities with centralized PV and battery.

Finally, the last chapter focuses on federated learning technology applied to home energy management systems (HEMS). Due to the increasing digitization of the low-voltage network and the implementation of smart meters, data protection has become increasingly important. Therefore, this study seeks to preserve user privacy by training prediction models in a distributed manner. Moreover, the proposed personalized federated learning methodology for HEMS demand forecasting incorporates a cost-oriented loss function to minimize imbalance costs while preserving customer privacy. The study compares cost-oriented and traditional loss functions and reveals that the personalized federated learning approach with cost-oriented loss function obtains the

lowest imbalance cost for HEMS optimization. Moreover, the study demonstrates that new households without large historical consumption data can still achieve good load prediction outcomes through collaborative learning models.

Resum

L'avanç ràpid de la digitalització, combinat amb la implementació de generació renovable i el desenvolupament de tecnologies de la informació i la comunicació dins la xarxa de distribució, està accelerant la transició cap a sistemes energètics intel·ligents distribuïts, digitalitzats i descarbonitzats. Malgrat el paper crucial de les dades en aquesta transformació, encara es subtilitzen. La tecnologia d'intel·ligència artificial (IA) té el potencial d'extreure informació valuosa d'aquestes dades, permetent serveis energètics innovadors i millorant el rendiment dels existents.

Aquesta tesi explora primer el potencial de la IA en serveis energètics impulsats per dades per als sistemes de distribució de potència a través d'un estudi exhaustiu. Després de revisar l'estat de l'art, el següent pas se centra en els sistemes de gestió energètica en l'àmbit domèstic. Aquesta tesi desenvolupa un sistema de gestió energètica multiobjectiu que minimitza simultàniament les emissions de gasos d'efecte hivernacle i les despeses d'electricitat, considerant tot el cicle de vida dels actius de generació utilitzats per proporcionar energia, incloent-hi la xarxa. Els resultats demostren que aquesta metodologia pot reduir significativament les emissions de gasos d'efecte hivernacle sense incórrer en despeses d'electricitat elevades. El capítol següent amplia aquesta aproximació innovadora basada en l'entorn al nivell de les comunitats energètiques locals amb fotovoltaica i bateria centralitzades.

Finalment, l'últim capítol se centra en la tecnologia d'aprenentatge federat aplicada als sistemes de gestió energètica dels habitatges. A causa de la creixent digitalització de la xarxa de baixa tensió i la implementació de comptadors intel·ligents, la protecció de dades s'ha tornat cada vegada més important. Per tant, aquest estudi busca preservar la privacitat de l'usuari entrenant models de predicció de manera distribuïda. A més, la metodologia d'aprenentatge federat personalitzat proposada per a la previsió de la demanda d'HEMS incorpora una funció de pèrdua orientada al cost per minimitzar els costos d'equilibri mentre es preserva la privacitat del client. L'estudi compara les funcions

de pèrdua orientades al cost i les tradicionals i revela que l'aproximació d'aprenentatge federat personalitzat amb funció de pèrdua orientada al cost obté el cost d'equilibri més baix per a l'optimització d'HEMS. A més, l'estudi demostra que les noves llars sense dades de consum històriques encara poden assolir resultats satisfactoris.

Contents

List of Figures

List of Tables

Nomenclature

Sets

B	Set of storage units
G	Set of generation units
P	Set of participants of a local energy community
T	Set of periods in the optimization planning horizon

Parameters

α	Weighting factor for price-based objective function EMS approach [-]
β	Sharing coefficient among participants in a LEC [-]
η^{ch}	Efficiency parameter for charging battery unit [-]
η^{dis}	Efficiency parameter for discharging battery unit [-]
$E^{gwp,avg}$	Average emissions for each type of generation source [kg CO_{2eq}/kWh]
$E^{gwp,bat}$	Emissions related to a Li-Ion battery [kg CO_{2eq}/kWh]
$E^{gwp,grid}$	Emissions related to the power system grid [kg CO_{2eq}/kWh]
$E^{gwp,pv}$	Emissions related to a photovoltaic generation unit [kg CO_{2eq}/kWh]
K^{cal}	Cost depreciation of the battery [€/kWh]
K^{invest}	Acquisition cost of the Li-Ion battery [€]
L	Life time of a BESS [days]
N^{hour}	Number of periods per hour [-]

$O^{soc,init}$	State of charge of a BESS at the beginning and end of the optimization horizon [kWh]
O_b^{max}	Maximum state of charge allowed for battery unit [kWh]
O_b^{min}	Minimum state of charge allowed for battery unit [kWh]
P^{buy}	Price for purchasing electricity from the grid [€/kWh]
P^{buy}	Price for selling electricity back to the grid [€/kWh]
P^{VAT}	VAT tax [-]
Q^{ch}	Maximum charging power allowed for battery unit [kW]
Q^{dis}	Maximum discharging power allowed for battery unit [kW]
S^{ch}	Threshold in charging process for battery unit [-]
S^{dis}	Threshold in discharging process for battery unit [-]
T	Temperature for battery calendar aging calculation [K]
W^{load}	Baseline consumption of load unit [kWh]
W^{pv}	Baseline generation of a photovoltaic unit [kWh]
$X^{max,ep}$	Maximum power allowed to export to the grid [kW]
$X^{max,imp}$	Contracted power. Maximum power allowed by contract to import from the grid [kW]
C	Cost associated to imbalances [€]

Variables

χ^{buy}	Amount of electricity bought from the grid [kWh]
χ^{sell}	Amount of electricity sold to the grid [kWh]
δ^{buy}	Binary variable, if a participant is buying electricity from the grid is set to 1; else 0 [-]
δ^{sell}	Binary variable, if a participant is selling electricity to the grid is set to 1; else 0 [-]

ψ^{pv} Amount of electricity produced from generation unit after optimization and flexibility activation [kWh]

σ^{ch} Amount of electricity charged to battery [kWh]

σ^{dis} Amount of electricity energy discharged to battery unit [kWh]

σ^{soc} State of charge of battery unit [kWh]

θ Net electricity generation associated to each participant, according to the sharing coefficient [kWh]

θ^{lec} Net electricity generation produced by the photovoltaic and BESS smart meter. [kWh]

V^{cell} Voltage in battery cell [V]

Acronyms

ADMM Alternating Direction Method of Multipliers

AI Artificial Intelligence

AMI Advanced Metering Infrastructure

ANN Artificial Neural Networks

ARIMA Auto-regressive Integrated Moving Average

BEMS Building Energy Management System

BRP Balance Responsible Party

CDFL Centralized Federated Learning

CEC Citizen Energy Community

CEP Clean Energy Package

CNN Convolutional Neural Network

DBSCAN Density-Based Spatial Clustering of Applications with Noise

DER Distributed Energy Resources

DL Deep Learning

DLT	Distributed Ledger Technology
DM	Data Mining
DNN	Deep Neural Network
DoME	Detection of Measurement Errors
DR	Demand Response
DSO	Distribution System Operator
DT	Decision Tree
EB	Environmental-Based
ELM	Extreme Learning Machine
EMS	Energy Management Systems
ESS	Energy Storage Systems
EU	European Union
EV	Electric Vehicle
FL	Federated Learning
FTL	Federated Transfer Learning
GBM	Gradient Boosting Machine
GDPR	General Data Protection Regulation
GHG	Greenhouse Gases
GIS	Geographic Information System
GRU	Gated Recurrent Units
GS	Generation Source
GWP	Global Warming Potential
HB	Hybrid-Based
HEMS	Home Energy Management Systems

HFL	Horizontal Federated Learning
IB	Incentive-Based
ICT	Information and Communication Technologies
IDAE	Instituto para la Diversificación y Ahorro de la Energía
iid	Independent and Identically Distributed
IoT	Internet of Things
IRENA	International Renewable Energy Agency
IT	Information Technology
KNN	K-Nearest Neighbors
KPI	Key Performance Indicator
kWh	kilowatt-hour
LCA	Life Cycle Assessment
LECs	Local Energy Communities
LEMS	Local Energy Management System
LL	Local Learning
LogR	Logistic Regression
LP	Linear Programming
LR	Linear Regression
LSTM	Long Short Term Memory
LT	Long term
LV	Low Voltage
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error

MIP	Mixed-Integer Programming
ML	Machine Learning
MLP	Multi-layer Perceptron
MLR	Multiple Linear Regression
MOP	Multi-Objective Problem
MSE	Mean Square Error
MT	Mid-Term
NB	Naive Bayes
NILM	Non-Intrusive Load Monitoring
NTL	Non-Technical Losses
OCV	Open-Circuit Voltage
P2P	Peer to Peer
PB	Price-Based
PCA	Principal Component Analysis
PFL	Personalized Federated Learning
PMU	Phasor Measurement Unit
PV	Photo-Voltaic
REC	Renewable Energy Community
RES	Renewable Energy Systems
RF	Random Forest
RL	Reinforcement Learning
RMSE	Root Mean Square Error
RNN	Recurrent Neural Networks
SCADA	Supervisory Control And Data Acquisition

SOC	State of Charge
SOM	Self-Organizing Maps
ST	Short Term
SVM	Super Vector Machine
T-SNE	T-distributed Stochastic Neighbor Embedding
ToU	Time of Use
TSO	Transmission System Operators
VFL	Vertical Federated Learning

Chapter 1

Introduction

This first chapter introduces the current context of the energy transition and establishes the motivation of this thesis. The main research questions, objectives and scope of this investigation are identified, addressing the key challenges and opportunities arising from the ongoing climate situation and defining the author's contribution to the field. Next, the structure of the thesis is presented. Lastly, a chronological summary of the research activities and related work undertaken during the doctoral period concludes the chapter.

1.1 Current context and motivation

The emission of Greenhouse Gases (GHG) and their negative impact on climate have become a primary concern for society. To minimize their impact in the future, an international consensus was achieved in 2016 with the Paris Agreement [11]. It aims to limit the increase in global average temperature by the end of this century to below 2 °C compared to pre-industrial levels. Recently, this threshold has been reduced to 1.5 °C.

The European Union (EU) is committed to leading the sustainable energy transition by setting even more ambitious energy targets and regulatory frameworks. Figure 1.1 provides a visual representation of the European GHG historical emissions, future net emission targets and medium-term predictions. In this context, the EU climate strategy for 2020 established a target of reducing GHG emissions by 20% compared to 1990 levels. The EU had already achieved its 20% target before the pandemic lockdown began to impact emission levels, having reduced emissions by 26% in 2019, while the gross domestic product (GDP) increased by 58% [1]. However, the pandemic recovery and the increment of fossil-based generation sources in the second half of 2021 have resulted in emissions growth in 2021.

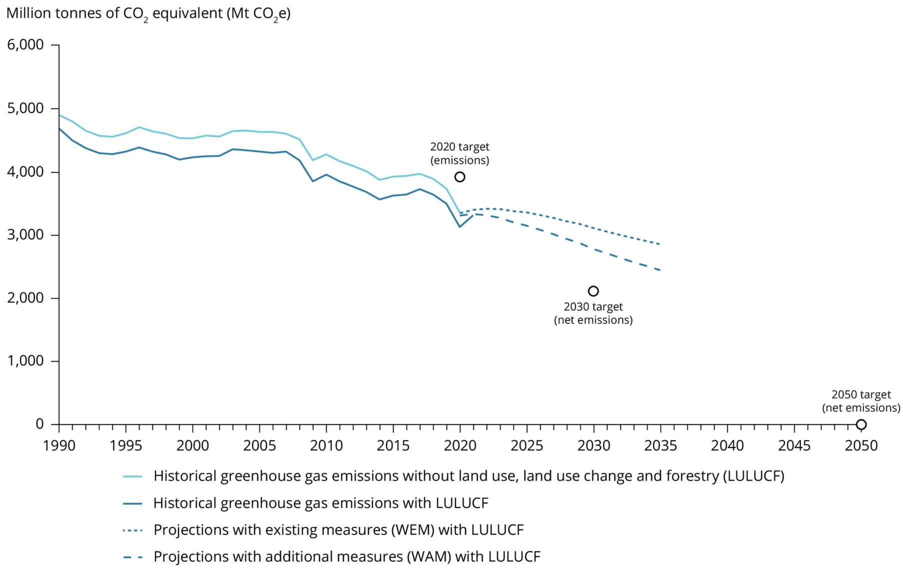

Figure 1.1: Historical trends and future projections of EU greenhouse gas emissions. Source: European Environment Agency.

In 2019, the EU presented the European Green Deal [12], a comprehensive set of policy initiatives covering almost all sectors, from building renovation, transport, energy, biodiversity, agriculture or innovation. This framework aims to achieve a minimum of 55% reduction in GHG emissions by 2030 and achieve a climate-neutral EU by 2050 with a net-zero pollutant emissions economy. This objective is even more challenging as electricity consumption is projected to increase by 20% to 40% by 2050 [13]. Figure 1.2 illustrates that thanks to the ambitious EU's climate policies, GHG emissions are expected to decouple from GDP growth. The EU's economy is projected to more than double by 2050 compared to 1990 levels while fully decarbonizing.

In order to reduce these emissions effectively, it is essential to understand their sources and the sectors that make the greatest contri-

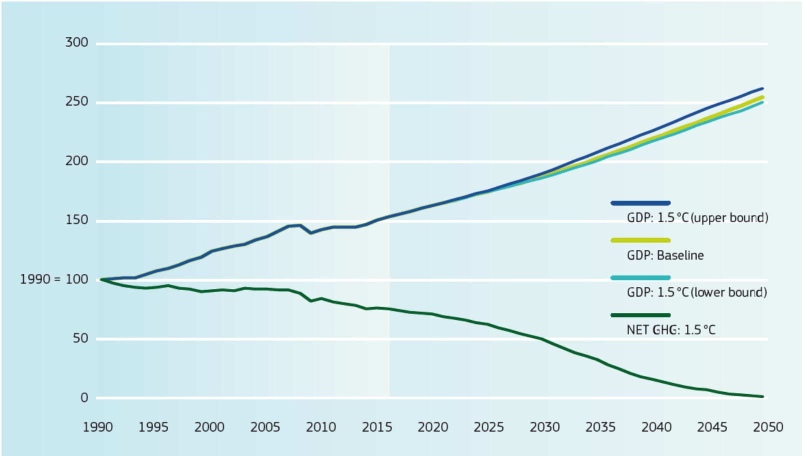

Figure 1.2: European climate policy decouples GHG emissions and GDP growth [1]. Source: European Commission.

bution to them. According to [2], energy (including electricity, heat, and transportation) is in charge of almost three-quarters of global emissions (73.2%). More in detail, the electricity and heat lead the annual emissions, as shown in Figure 1.3. Of the total energy-related CO_2 emissions, more than 40% are due to the combustion of fossil fuels for electricity generation. Therefore, the energy sector plays a critical role in mitigating these effects by improving energy efficiency and transitioning traditional fossil-based electricity generation to carbon-free alternatives.

Finally, Figure 1.4 depicts the historical and projected average annual investments required to achieve the aforementioned EU climate targets [3]. The residential sector has invested the most by far, and it is expected to increase its investments even further in the next 10 years. For these reasons, this thesis focuses on both, the residential and power grid sectors, aiming to help these categories become drivers of change towards a sustainable economy and energy democratization through the implementation of small-scale, distributed renewable energy and self-consumption.

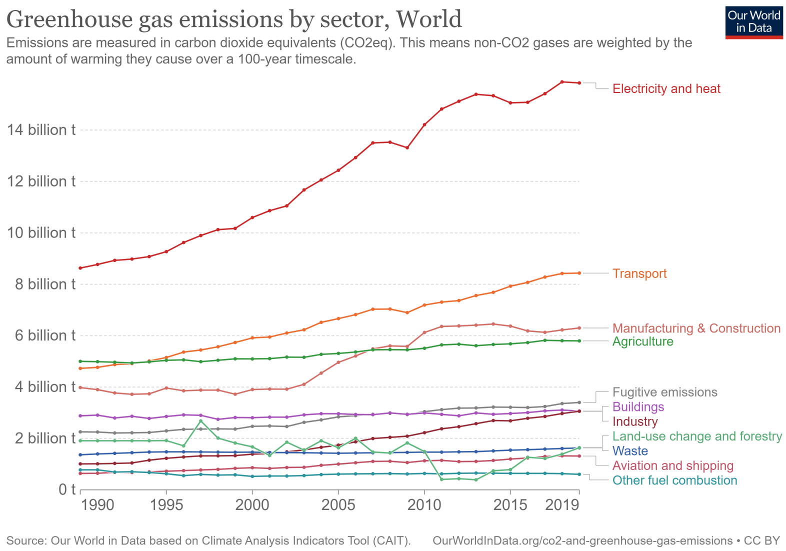

Figure 1.3: Annual greenhouse gas emissions by sector [2]. Source: Our World in Data.

1.1.1 Digitization and flexibility as energy transition enablers

Recent advances in Information and Communications Technology (ICT) together with the implementation of smart meters and sensors at low voltage levels are facilitating real-time monitoring, transforming the traditional power system into a Smart Grid. It is undeniable that digitization enhances customer engagement, providing customers with more detailed information and insights about their electricity consumption patterns, empowering them to make better decisions about their energy usage. Additionally, current artificial intelligence technology and data analytic tools are enabling the extraction of added value from the massive amount of data generated daily, leading to novel data-driven business models in the energy domain.

In this context, the *flexibility* term arises, which refers to the power system's ability to modify or adjust the electricity consumption and/or generation in response to variability or external factors such as electricity prices or flexibility requests from other electricity market

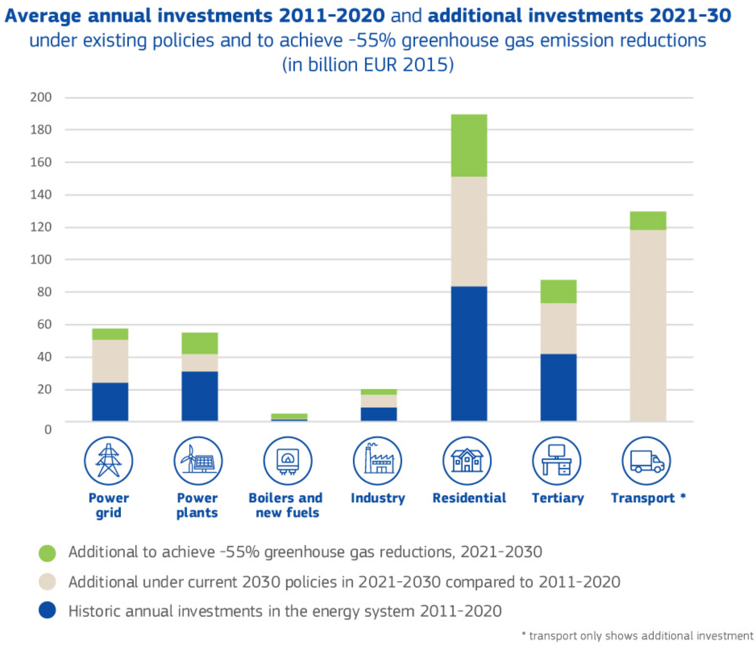

Figure 1.4: Average annual and additional investments to achieve 2030 EU climate targets [3]. Source: European Commission.

participants such as Transmission System Operators (TSOs), Distribution System Operators (DSOs) or Balance Responsible Parties (BRPs), with the purpose of maintaining network stability, minimizing congestions, and reducing or avoiding imbalances, among others.

Focusing on consumers, energy management systems enable customers to monitor, control, and optimize their energy consumption and usage through demand-side management strategies, which aim to change consumption patterns by encouraging customers to shift their electricity consumption to different periods or reducing/curtailing the overall or some demand at specific periods. These changes are for the user's own benefit or due to external signals or flexibility requests by thirds of electricity market agents in exchange for economic compensation. Thus, indirectly, GHG emissions associated with electricity generation are reduced, the efficiency of the electricity system is increased, and congestion in the grid can be avoided by activating flexibility, thereby increasing the resilience and resistance of the electricity system. Additionally, the flexibility market provides a platform to aggregate the end-users' flexibility to participate in the energy transition actively and contribute to reducing greenhouse gas emissions while also potentially benefiting from cost savings and other incentives. Thus, the end-user becomes an active and indispensable part of the electricity market and energy transition.

Therefore, the energy management systems' flexibility service support and accelerate the ongoing energy transition and also promote the empowerment of ordinary citizens, allowing them to manage and monitor their energy consumption and production (in the case of being prosumers), involving them and making them key players in the essential energy and climate transition. Furthermore, thanks to self-consumption and the emergence of local energy communities among neighbors, electricity generation is decentralized, thus minimizing transmission losses and democratizing energy consumption, making it accessible to everyone. This new vision of the electricity system also generates direct economic benefits for users, providing substantial savings through self-consumption and the sale of excess electricity in exchange for economic compensation. This new scenario is highly beneficial for communities as it not only enriches them but also serves as a strong incentive for the creation of jobs in the field of energy transition.

1.1.2 Motivation

In summary, the urgent need to mitigate climate change and reduce greenhouse gas emissions has encouraged the acceleration of the transformation of energy systems, especially in the residential and building sector. Moreover, 90% of European citizens agree that reducing CO_2 emissions is necessary to achieve the 2050 emissions neutrality target [14]. Therefore, the main objective of this thesis is to contribute to this global effort by developing innovative environmental-based energy management system strategies that leverage the digitization of the distribution network through the implementation of sensors and smart meters. By enabling self-consumption and flexibility, end-users can become active participants in the energy transition and essential drivers of change. Furthermore, this thesis considers user data privacy preservation to enable the creation of forecasting models without compromising data security. This investigation creates an innovative and sustainable energy management solution that supports the transition towards a decarbonized economy while also protecting the privacy and security of user data.

1.2 Objectives and scope

This section outlines the Research Questions (RQ) that led to the objectives and scope of the thesis conducted by the author.

RQ1) What novel data-driven energy services are likely to emerge or could benefit from the daily vast amount of operational and non-operational data related to distribution networks?

RQ2) Which Artificial Intelligence techniques are utilized in the development of these data-driven energy services in the distribution network, and how do they contribute to enhancing the sustainability, efficiency, and reliability of the system?

RQ3) How can electricity consumers and local energy communities contribute to enhancing sustainability through emissions reduction and how can they be engaged and incentivized?

RQ4) Given the need to protect user data privacy, which machine learning distributed method can be applied to train energy management system prediction models without compromising personal data?

RQ5) How can cost-oriented loss functions in EMS prediction models help to minimize imbalance penalizations due to energy consumption estimation errors?

Figure 1.5 provides a conceptual overview of the objectives and contributions of this thesis. The image details the scope of this thesis, which is mainly focused on energy management systems, both at the user and energy community levels. Firstly, a comprehensive analysis of the state-of-the-art data-driven energy services and AI methods is conducted in (O1). Subsequently, this work focuses on energy management systems, developing an optimization model in (O2). In addition, a novel environmental-based optimization approach is developed and evaluated for Home Energy Management System (HEMS) in (O3), and (O4) applies this method to local energy communities. Furthermore, federated learning is included in HEMS to ensure customer data privacy and enhance collaboration between different energy stakeholders who may be hesitant to share their data otherwise. In this context, (O5) presents the personalized approach for end-users demand forecasting, while (O6) introduces a cost-oriented loss function for training the models developed in (O5).

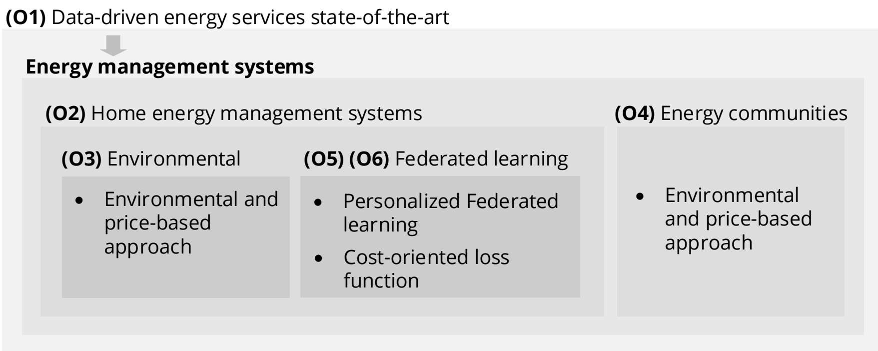

Figure 1.5: Global overview of the objectives and contributions of this thesis.

This thesis responds to the following objectives, presented in different chapters:

O1 Analyze the artificial intelligence methods applied in distribution networks to enable data-driven energy services.

First, a comprehensive analysis of artificial intelligence methods applied in distribution networks to enable data-driven energy services is conducted. The research involves identifying and classifying various data-driven techniques for power systems and the data sources involved in data acquisition. The study explores different energy services, including operation and monitoring, predictive maintenance, non-technical losses detection, forecasting, flexibility and planning of distribution grids. The research also maps the relationship between the distribution grid applications, proving that multiple services can be offered as a single clustered service to different energy-related stakeholders. Additionally, the analysis identifies the dependencies between the AI techniques with each energy service.

O2 Develop an energy management system optimization model capable of controlling flexible assets in order to minimize the electricity cost.

After comprehending the state of the art of the different energy services, this thesis focuses on energy management systems. Therefore, this objective seeks to develop an optimization model for a household energy management system capable of controlling flexible assets to minimize electricity costs while considering the end-user constraints. In this phase, different flexible asset models are designed, including a battery model taking into account the battery degradation parameters due to usage and calendar aging and flexible PV generation, among others.

O3 Develop an environmental-based objective function for energy management systems that, in addition to minimizing cost, incorporates the reduction of greenhouse gases associated with consumption.

After completing the previous objective, a novel strategy that simultaneously minimizes greenhouse gas emissions and electricity expenses is proposed. The approach considers the entire life cycle of the generation assets used to provide energy, including the electricity grid. This multi-objective function empowers end-users to determine their preferences at any time, enabling them to prioritize minimizing cost, emissions, or a balance of both.

O4 Transform and adapt the individual energy management system model to a local energy community optimization model, with centralized PV and battery.

The aim of this study is to transform and adapt the individual energy management system model and the environmental and price-based approach to local energy communities consisting of several buildings with centralized and shared PV and battery. The optimization seeks the overall neighborhood benefit.

O5 Apply federated learning for energy management systems forecasting models using personalization.

A personalized federated learning methodology is developed for home energy management systems demand forecasting. The approach addresses the challenges of customers' data privacy and security, as well as mitigates the challenges of data silos in the energy sector by enabling collaboration between stakeholders, such as energy providers and customers, reducing or eliminating cloud-computing costs. The methodology involves retraining a global centralized federated learning model using user-specific smart meter data to build a personalized federated learning model for each consumer. The new personalized model is kept locally, maintaining the preservation of personal data.

O6 Integrate cost-oriented loss functions with personalized federated learning models for HEMS load predictions

The final contribution of this thesis is to integrate a cost-oriented loss function with the personalized federated learning approach developed in the previous step for HEMS load predictions. The study demonstrates and validates the cost savings resulting from the use of cost-oriented personalized federated learning models in HEMS for end-users with varying historical data availability.

1.3 Thesis outline

The contents of the thesis are organized in the following chapters:

- **Chapter 2** provides a comprehensive literature review of artificial intelligence methods applied to data-driven energy services that have emerged with the recent digitization of the electrical distribution network.
- **Chapter 3** presents a novel multi-objective hybrid HEMS designed to minimize both electricity costs and greenhouse gas emissions resulting from end-user consumption. To assess the impact of each technology generation on the climate, a life cycle analysis methodology is used.
- **Chapter 4** extends the scope of the previous research by assessing different combinations of multi-objective objective function approaches for local energy communities.
- **Chapter 5** combines federated learning technology with a cost-oriented loss function for HEMS load prediction to enhance data privacy and collaborative learning.
- **Chapter 6** summarizes the conclusions of the work and introduces the future research lines for each of the research topics addressed.
- **Appendix A** enumerates the publications and research outcomes, both related and unrelated to the thesis manuscript.
- **Appendix B** briefly presents the detection and measurement errors tool developed to avoid data errors.

1.4 Thesis related work and activities

This section provides a chronological overview of the relevant activities and work developed by the author during the thesis period, derived from national and international projects that have directly or indirectly fed this thesis. This is summarized in Figure 1.6.

Doctoral activities started in October 2018, with the author collaborating from 2018 until the end of 2019 on the Innovation Action H2020 European project INVADE *Integrated electric vehicles and batteries to empower distributed and centralized storage in distribution grids*, under Grant Agreement No 731148. This work consisted on the development of a home energy management system optimization model that minimizes the energy bill while offering flexibility to energy market stakeholders to cope with grid congestion and imbalances, which led to a conference paper [C1]. Moreover, the development of shiftable and curtailable flexible source models, such as electric water heaters, led to a conference article [C5] and oral presentation [P-C1]. The author also collaborated with colleagues at CITCEA-UPC, NTNU, and companies like Anell and Smart Innovation Norway on behalf of the INVADE Project, which resulted in several outcomes not included in this thesis, such as journal article [J5], conference papers [C3], [C4], [C6], [C7], and oral presentation in international events [P-C1].

The research on the HEMS assessment in smart grids continued in 2020, under the BD4OPEM H2020 Project *Big Data for Open Innovation Energy Marketplace*, Grant Agreement No. 872525. This project involved 12 partners from eight different countries and five pilot sites. In this context, a review of the state-of-the-art artificial intelligence methods applied to data-driven energy services was conducted in [J1]. These new services aim to add value to the vast amount of data generated daily in the distribution network, thanks to its digitization.

The research conducted under BD4OPEM involved the development of four data-driven energy-related services, which are described below:

- Detection of measurement errors. This service involved the use of a data-cleaning pipeline that used artificial intelligence to detect anomalies and impute missing values. This data-cleaning algorithm was used prior to HEMS execution to ensure optimal performance and data quality. This work is described in **Appendix B**.

- Flexibility aggregated services for balanced responsible parties. This service consisted of two approaches. The first approach involved balance portfolio optimization to avoid penalties for deviating from the system. The second approach involved minimizing purchase costs in the day-ahead electricity market. The results of this work are presented in [J6].
- Building energy management systems for the central offices of a distribution system operator in Spain. This service involved developing an algorithm to manage a group of buildings within a local energy community with self-consumption and collective storage. The results of this work are presented in a journal article [J4].
- Energy forecasting. This service involved developing PV generation and demand forecasting at the LV level.

In parallel with the BD4OPEM project, the author also participated in the FLEXRED project *Flexibility of distributed energy resources to optimize the operation of distribution networks*, supported by Ministerio de Ciencia, Innovacion y Universidades under RTI2018-09954. As part of this project, the author developed an environmental-based approach for an energy management system, which resulted in a journal publication [J1], a conference article [C2], and an oral presentation at a congress [P-C2]. In addition, the author has also participated during this period in an industrial project with a Catalan distribution system operator company to develop an energy management system for a local energy community, which consists of the offices of this corporation. As a result of this work, article [J4] has been published.

In 2022, the author started working on the ATLAS project *Digitization using novel data analytic methods and toolboxes for secure, renewable, and flexible grids* under reference PID2021-128101OB-I00 funded by Ministerio de Ciencia, Innovación y Universidades and by ERDF A way of making Europe. The project topic is aligned with the focus of an international stay in Imperial College London (in the Electrical Engineering Department), from August to December 2022. During this stay, the author researched personalized federated learning for HEMS and combined it with a cost-oriented loss function approach to minimize imbalance costs. The research led to journal publication [J3]. Moreover, during 2022 I have served as the Scientific Coordinator of the Estabanell Chair in Smart Grids, where my responsibilities included fostering collaborations between industry and academia, as well as overseeing and leading research and innovation projects in the distribution area of the company.

In 2023, the author started participating in two projects related to the topic of this thesis: the MERIDIAN project (Flexible distribution grid management for maximum decarbonization using artificial intelligence) funded by Ministerio de Ciencia, Innovación y Universidades under TED2021-131753B-I00 and PLATON (PLAtaforma ONline integradora de datos de energía y de servicios IA para redes de distribución).

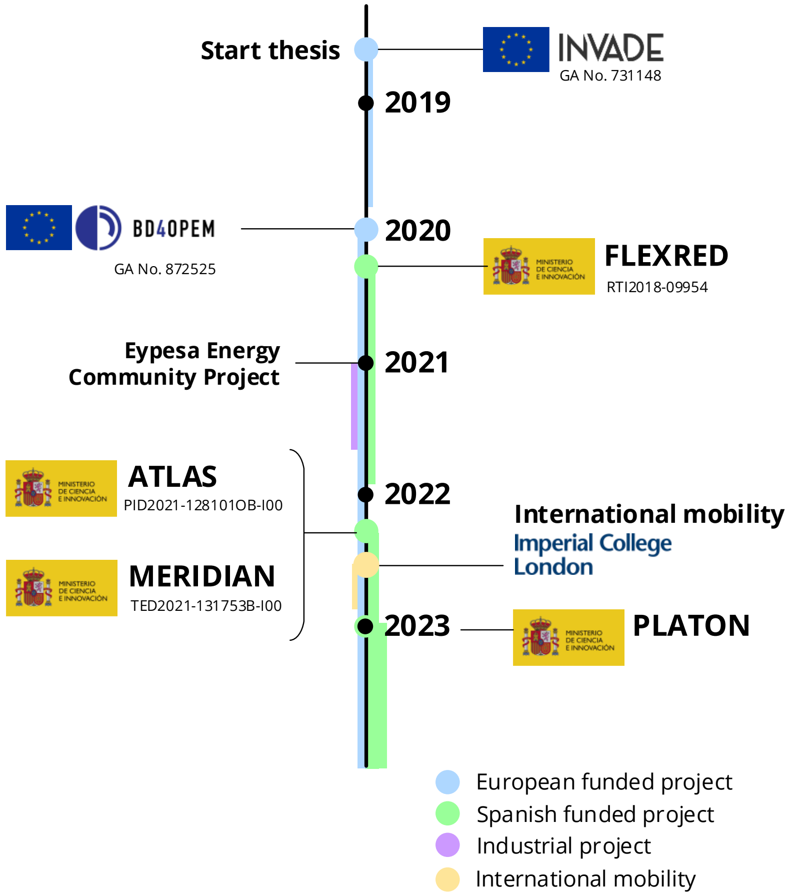

Figure 1.6: Timeline of the projects and main activities carried out during the thesis.

Chapter 2

Artificial intelligence for enabling data-driven energy services

In this chapter, a comprehensive literature review is conducted on Artificial Intelligence (AI) methods applied to data-driven energy services that have emerged with the recent expansion of digitization in electrical distribution networks. This exhaustive analysis includes energy services such as operation and monitoring, predictive maintenance, non-technical losses, forecasting, flexibility management and planning. The relationships and interactions between them are examined. Many of the applications identified lead to data outputs that can be input for other energy applications, enabling several groups of potential services for different stakeholders. Furthermore, future opportunities and challenges for the application of AI in distribution grids are identified.

2.1 Introduction

The progress of Information and Communication Technologies (ICT) and digitization are accelerating the transition towards smart energy systems [15–17], where data have a remarkable but still under-exploited role [17, 18]. These data, collected by different types of sources, need to be preprocessed [19, 20] before applying AI techniques that lead to several Big Data applications in power systems [21].

The operation and management of electrical grids are per se complex decision-making processes, even more challenging taking into account the increasing penetration of renewable energy sources, which are adding more variability and uncertainty in the power system operation [22]. To address the operation, maintenance, and planning of electrical grids, classical analysis tools can require large computational time and might not always find a feasible solution. In this sense, AI techniques can contribute to operating, maintaining and planning electrical networks by treating and extracting value from large volumes of data, dealing with its variety and velocity, through much faster computations [23].

Considering the potential of the data collected in electrical networks, the scientific community is applying and developing AI techniques for power system applications [23]. AI can be applied in all the power system domains, including generation, transmission, distribution and consumption. In particular, the International Renewable Energy Agency (IRENA) envisages its application to promote the grid integration of renewable energy sources in all the before mentioned power system chain through: forecasting for renewable power plants (like wind and solar large-scale power plants), grid stability and reliability at transmission and distribution level, demand forecasting, demand-side management, optimized energy storage operation and optimized market design and operation (the latter two as multi-domain applications) [24]. The present chapter focuses on AI applications in the distribution and consumption domains.

2.2 Data-driven techniques for power systems analysis

The massive amount of data currently being produced alongside the power system pipeline, from generation to demand side, has arisen the opportunity to understand the system better and create innovative services based on these data. For the sake of this investigation, the data-driven techniques classification is based on the most relevant publications in the energy field and based on Statistical, Neuroscience, Computer Science and Mathematical references. To help the reader to identify the different data-driven techniques, Figure 2.1 displays the most relevant data-driven related areas in the energy sector considered in this study.

Let us first define what is meant by artificial intelligence. IRENA defines AI as an area of computer science that focuses on creating intelligent machines that follow human behavior, according to the data collected [24]. This is considered the starting point for the development of data-driven services in the electricity sector. However, this concept can sometimes be misunderstood or considered too broad since some techniques inside Statistics and Data Mining (DM) are placed within the AI area.

ML is a sub-field of AI and computer sciences [4,24] that has evolved from pattern recognition to analyze the data structure and fit it into models that can be understood and replicated by users [4]. Figure 2.1 defines all the ML categories, methods and models applicable to energy-related projects, taking into account the standard approaches and definitions of different authors [25–28]. The classification provided in this thesis matches the one proposed in [28] which is also implemented within the power systems field. Furthermore, ML is classified into four categories: Supervised Learning, Unsupervised Learning, Reinforcement Learning (RL) and Deep Learning (DL). Supervised and Unsupervised Learning categories aim to predict or describe the existing relationships within the data set, being called supervised when the dependent variable is available and unsupervised when it is not. RL is a computational approach that learns from the interaction with the environment, which means defining how system agents can take actions in their environment to maximize the cumulative reward [29]. RL is implemented mainly in energy dispatch problems and building energy management scheduling [30]. Some bottlenecks expected in these applications are the complexity of the objective functions (non-linear and non-convex) plus the limitations of physical models. The main advantage of using RL instead of predictive model strategies is that RL operates a model-free approach and does not require convergence guarantees, thus enhancing its applicability. Moreover, RL needs scarce knowledge about the problem physics to be competitive with standard rule-based controllers [31]. However, more research and real-world testing are needed for RL technology to become more mature. DL belongs to the Artificial Neural Networks (ANN) field. They are a broad family of techniques in multiple domains which can be applied to both Supervised and Unsupervised Learning. ANNs are inspired by the function of the brain, with the primary objective of learning from unstructured or unlabeled data, using single or multiple layers (DL approach) to extract higher-level features from the raw input progressively. DL techniques can be applied to power systems in different scenarios such as fault detection in transformers [32–34], Photo-Voltaic (PV) forecasting [35,36] and day-ahead electricity market price forecasting [37].

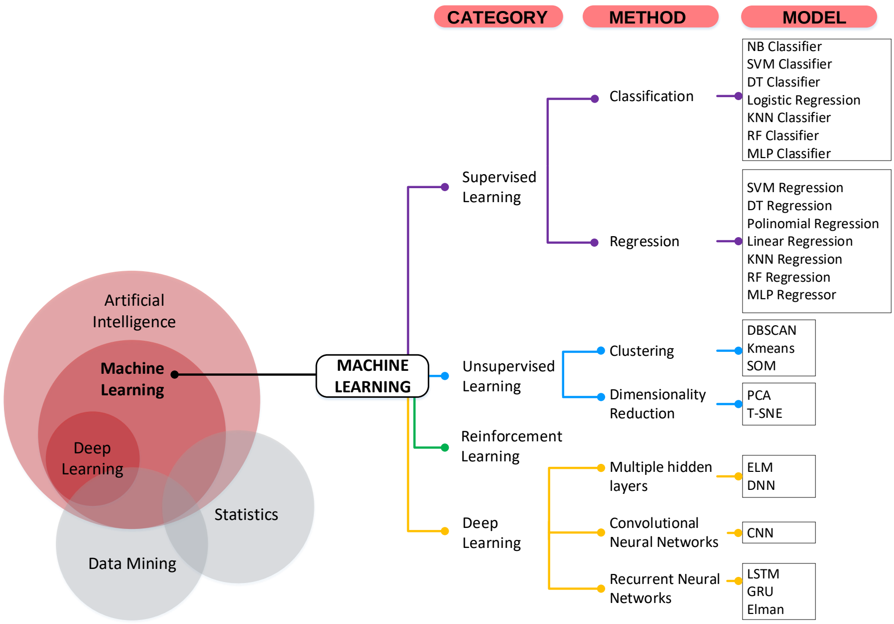

Figure 2.1: Data-driven techniques classification derived from [4,5] and Machine Learning categories for power systems analysis.

It is worth mentioning the role of Statistics in power systems. Statistics is an applied science concerned with the analysis and modeling of data [38]. Despite the similarity with ML, Statistics is the field of Mathematics that deals with the understanding and interpretation of data. For some references [39–41], Statistics aims to provide an overview of the data set, rather than forecasting or extracting relationships between the data, being generally applied in the preprocessing step of the data science pipeline. According to [38], learning problems that can be solved by applying statistical techniques can be roughly classified as either Supervised or Unsupervised. Defining the boundaries between Statistics and ML can be controversial, and often, some methods are considered both Statistics and ML, while other references classify the same method in a specific knowledge domain. According to [39], Statistics uses a population sample to draw population inferences, while ML determines generalizable predictive patterns from data.

In conclusion, the research combines inference and prediction, and frequently the classification method is more related to the scientific domain where the techniques are applied (i.e., Computer Science, Mathematics, Engineering) rather than the particularities of the technique itself.

2.3 Energy data sources

This section identifies and classifies the massive amount of heterogeneous data required for developing and operating the distribution grid energy services listed in the previous Section 2.2. With the objective of accelerating the development and deployment of the Smart Grid, a significant amount of sensor devices have been installed in the distribution network to increase its observability of dynamic and transient events and collect information about the actual state of the grid, thus achieving a higher level of monitoring, observability and control beyond substation level. Nevertheless, not all the needed data come from direct electrical grid measurements. For instance, [42] distinguishes between electrical and non-electrical information and identifies three categories: measurement data, business data and external data. On the other hand, authors from [22] divide the data sources into operational and non-operational data, whose criterion is used in this thesis. Diverse investigations have examined the data sources available in the grid; for instance, [43] compares eight advanced measurement devices in distribution networks and reviews their most recent Smart Grid applications.

The volume of data generated is expected to grow in the upcoming years [22]. As an example, according to [44], the generated data coming from a single Phasor Measurement Unit (PMU) can be estimated at around 100 GB per year. Therefore, new energy-related services and business models need to emerge to take advantage of this massive data that are still underused or unused. Therefore, Big Data analytics is essential for processing data whose size is beyond the capability of a typical database software tool.

The flow chart presented in Figure 2.2 depicts the steps that data follow from the time they are collected until an energy service requests it. The data sources, located at the bottom, are classified into two large groups: operational and non-operational data. The power system operational data include all the measurement assets that collect power and energy data, including voltage, current, active and reactive power and grid status signals. On the other hand, non-operational data provide essential information that has a crucial role in supporting the energy services performance, such as weather conditions, electricity market data, social media, Geographic Information System (GIS) and known parameters given by customers. The Big Data distribution grid services can request both real-time and historical data. Once the data has been collected, the next step is to harmonize the data to enhance its usefulness and provide a standard structure regardless of the measurement source. The data harmonized are ingested and saved in the data storage, also known as data lake. When an innovative service requests data from the data lake database, the information goes through a cleansing process to further increase data quality by eliminating duplicates and imputing values to the missing data through AI and statistical techniques.

Primary Big Data sources within the distribution grid are described next:

Operational data: information extracted from the distribution grid measurement devices.

- Advanced Metering Infrastructure (AMI): is an integrated system of smart meters, communications networks and data management systems that enables two-way communication between power utilities and customers [45].
- PMU: measures time-referenced magnitudes and phase angles of voltage and current phasors [46].
- Supervisory Control And Data Acquisition (SCADA): collects

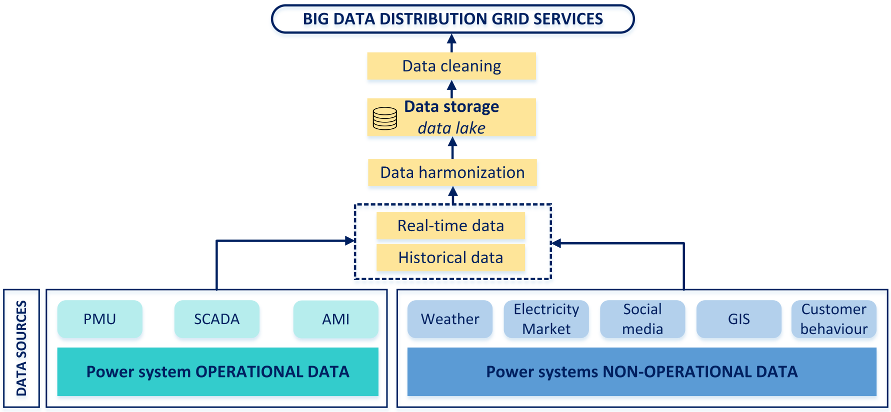

Figure 2.2: Big Data sources used in power system data-driven services.

data automatically from distribution grid components, thus facilitating remote monitoring and control.

Non-operational data: information that helps power system utilities gain deep insights into why some events occur in the grid.

- Weather data: refers to time-dependent meteorological conditions such as irradiance, temperature and wind speed. For instance, atmospheric information is vital for forecasting algorithms related to energy systems.
- Electricity market data: the results obtained from the matching up of the daily and intraday markets offer relevant information, such as the day-ahead electricity price and the amount of energy by generation technology type.
- Social media: through text mining methods, faults in the distribution grid or a fire that may harm the electrical infrastructure can be detected through social network comments.
- GIS: provides information about the grid components' location such as lines, transformers and feeders.
- Customer behavior data: known parameters related to customers. For example, the number of people living in the house, square meters, number of rooms and income level.

2.4 Data-driven services in distribution systems

In this section, a review of a wide range of innovative energy services within the distribution network is evaluated. The most relevant AI techniques used in each service are detailed. To obtain a comprehensive outlook, Figure 2.3 classifies these energy services into the following three categories:

- Distribution grid operation: responsible for ensuring the correct operation of the distribution network.
- Flexibility management: in charge of the flexibility market.
- Planning: responsible for optimal investment strategies that contribute to the long-term planning in the distribution grid.

The measurements error detection service is excluded in the flow chart since it is inherent to the rest of the services and is not offered directly to the electricity market stakeholders since its task is to identify, detect and solve anomalies, errors, or missing values from data sources to ensure the quality and usability of the distribution grid services. The primary objective of implementing AI methods in Big Data energy services is to accelerate and stimulate the existing power system toward an environment-friendly, cost-effective and reliable Smart Grid. The services are offered to a broad range of stakeholders involved in the energy domain, including DSOs, BRPs, prosumers and aggregators, among others. The purpose is to improve their performance and encourage the creation of novel business models in the energy sector in order to take advantage of the massive data that are being generated and underused.

2.4.1 Measurements error detection

The measurement error detection application identifies, detects and solves eventual anomalies, errors or missing records from data sources in order to ensure data quality and usability. Depending on the type of anomaly detected, a correction is automatically executed. The data cleaning step in Figure 2.2 is responsible for executing these tasks. The following anomalies have been distinguished in operating and non-operating data:

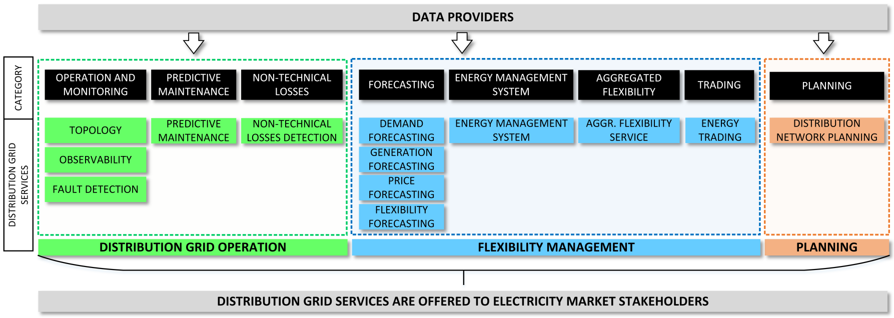

Figure 2.3: Scope of the Big Data services in the distribution network.

- Duplicate records: frequently happen during data collection due to communication channel problems or combined data sets from multiple sites.
- Structural errors: arise during measurement or data transfer.
- Unit inconsistencies: this happens when there is a change in the units and the past recorded data are not altered.
- Outliers: abrupt and short-duration changes in the consumption pattern that are not a valid representation of the actual consumption. The sources of spikes could be mechanical faults of the meter or storing multiple inconsistent readings for the same timestamp.
- Missing data: occurs when no value is received for an observed variable.

The measurement error detection sequential scheme is presented in Figure 2.4, which identifies possible anomalies (see the dark blue boxes) and proposes what techniques can be implemented to solve them (see the light blue boxes). Moreover, a logical order when preprocessing data need to be followed. The most suitable technique to address an anomaly might vary depending on each service requirement. AI methods are a powerful tool for assigning predictive values for missing data. Regarding missing data imputation, [47] implements a Long Short Term Memory (LSTM)-based method for bi-direction data imputation, [48] applies a Multi-layer Perceptron (MLP) ensemble, while [49] works with Graph-based ANN. The main advantage of AI and Big Data analysis is that it automates and improves the error detection process of the ever-increasing energy-related measured data. For instance, [50] has developed a smart meter data error recognition technology applying ANN and Super Vector Machine (SVM) classification algorithms.

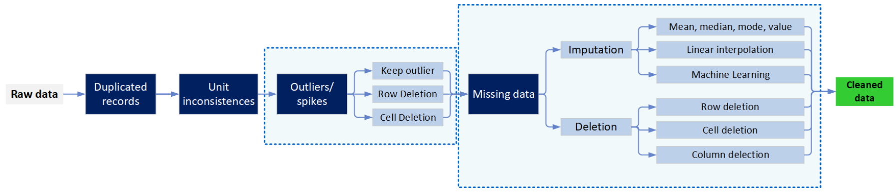

Figure 2.4: Measurements error detection service steps.

2.4.2 Operation and monitoring

The operation and monitoring category is responsible for improving the observability and performance of the distribution grid in nearly real-time. Data-driven services such as topology, observability and fault detection are included in this subsection.

New measurement devices with high granularity and power quality resolution data like PMUs (120 Hz to 30 kHz and beyond) [22], AMI and SCADA contribute to strengthening the LV grid monitoring by providing essential information that assists in comprehending the grid status and identifying possible congestions. The challenge is to monitor the distribution grid operating conditions in nearly real-time to check its status. Nevertheless, it is necessary first to know the distribution grid topology to identify branches and nodes with technical problems in order to generate a quick response to mitigate them. Therefore, the topology estimation is an important step to ensure the distribution grid operation and monitoring robustness.

2.4.2.1 Topology

The topology service aims to perform the complete retrieval of the entire LV network electrical scheme. For security and operational reasons, the transmission system is equipped with real-time measurement devices at each node (bus voltages, line flows, bus injections) to ensure a reliable, robust and accurate topology identification and observability of the power system [22]. However, topology is commonly unknown at distribution levels [51] due to the scarcity of real-time measuring and breaker status devices, which hinders its observability. Nevertheless, due to the ever-increasing presence of advanced ICT within the distribution grid domain, combined with the constant rise of smart meters deployment [52], an enhancement of the original topology structure is possible. For this reason, it is essential to implement some of the Big Data techniques explained in Section 2.3 to be capable of processing and analyzing all these amounts of data generated. In addition, the electrical grid topology is an essential input for other energy services like observability, non-technical losses detection, predictive maintenance and aggregated flexibility services in order to have outstanding performance and reliable operation. The literature distinguishes essentially between transmission and distribution network topology, although research efforts have recently concentrated on the latter. Depending on the frequency of data collection and the purpose of the service, topology can be estimated in real-time [53–55] or offline [56, 57]. The principal data sources for assessing low and medium voltage topology are SCADA, smart meters and PMUs, the latter being the most used in research. As [58] points out, topology estimation accuracy depends mainly on the availability and accessibility of the measurement instruments; however, PMU is capable of achieving satisfactory outcomes even with limited measurements. [59] calculates an equivalent grid applying a least-squares model-free approach by choosing PMU measurements at a limited number of buses, whereas [60, 61] use local electricity market prices as input data to obtain the distribution grid topology.

The Alternating Direction Method of Multipliers (ADMM) is a popular method for distributed convex optimization problems, which decomposes a large problem into smaller sub-problems, enhancing its robustness. The topology estimation problem is formulated and solved using ADMM in [60–63]. The main advantages of this method are that it allows handling and solving large-scale data problems and the implementation is straightforward. On the contrary, the convergence could be slow.

Regarding statistical methods, correlation is widely used in literature as a tool for topology estimation. For instance, [56] analyzes the correlation among voltage measurements to determine the grid topology, meanwhile [64] proposes a correlation-based algorithm to identify the transformer and phase to which a customer is connected. [65] reconstructs the topology given the voltage magnitude measured from smart metering devices, formulating the adaptive Lasso algorithm to obtain the correlation coefficient matrix. Another study [66] calculates the correlation coefficient among voltage measurements of smart meters under the same distribution transformer and is capable of grouping customers that belong to the same phase effectively. Ultimately, [67] uses a statistical learning framework for verifying single-phase grid structures using non-synchronized voltage data.

Regarding AI-based methods, a binary classifier based on ANN identifies the status of a distinct line [53]. A learning-to-infer variational method [57] considers three classifiers methods -Decision Tree (DT), MLP and Logistic Regression (LogR)- for predicting the state of the switch line, where MLP outperforms. On the other hand, [68] presents a data-driven topology estimation method that applies the Linear Regression (LR) algorithm and historical voltage measurements as input to the model, where the admittance matrix or switch location information is not required. In [69], smart meter voltage patterns enable topology identification by applying unsupervised learning clustering methods, but the article does not specify which algorithm is applied. The study carried out in [70] develops a Deep Neural Network (DNN) system that interprets the reflected signal from the impedance discontinuities in the network, which gives the possibility to determine the topology at the reflection site. The authors of [71] propose a Supervised Learning framework that first estimates the parameters and an Unsupervised Learning model to identify the topology. The proposed algorithm performs well in radial and loopy distribution networks. In [72], a LR method is proposed to evaluate the distribution network topology changes. Unfortunately, this method is accurate only if there is no noise in both input and output measurements [71].

Data-driven techniques found in the literature for topology estimation are listed in Table 2.1. The input data required for topology estimation is shown in Table 2.2.

Table 2.1: Data-driven techniques used in distribution network topology applications.

Data-driven technique	**Ref.**
Correlation	[56, 59, 64–66]
DNN	[70]
DT	[57]
LR	[65, 68, 72]
LogR	[57]
MLP	[53, 57]

Table 2.2: Data sources for distribution network topology estimation.

Data source	**Input measurement**	**Ref.**
Electricity market	Electricity prices	[60, 61]
PMU	Voltage phasor	[51, 53, 57–59, 63, 71]
SCADA	Power injections	[55, 62]
	Voltage magnitude	[51, 55]
	Current	[55]
	Switch status	[55]
Smart meters	Power injections	[64, 72]
	Voltage magnitude	[64–68, 71, 73]

2.4.2.2 Observability

The observability service assesses the most probable state of the distribution network state in nearly real-time. Potential applications like congestion management, optimal voltage/power control, fault detection and non-technical losses detection, among others, required instantaneous information regarding distribution system status to perform accurately.

For classic state estimators, power system measurements and switching device statuses are collected using SCADA systems, but the downside is that SCADA sampling rates are slow [74]. On the contrary, PMUs provide high power quality resolution data in real-time. Therefore, the innovation in the observability field attempts to include real-time operation data to provide a continuous and safe state estimation of the distribution grid and apply AI methods to develop reliable data-driven solutions. It should be noted that errors made in topology estimation might downgrade the performance of the observability service [58]. Smart meters are commonly applied for state estimation, for example, in [75,76]. The authors in [76] analyze the optimal positioning of smart meters for optimal cost-effective operation of the distribution grid to increase the state estimation accuracy, while [77] implements real-time PMU measurements to monitor the distribution grid status. A hybrid state estimation using AMI and SCADA measurements is formulated in [78].

Concerning data-driven methods, ML algorithms -DNN, SVMs, and Recurrent Neural Networks (RNN), among others- are used in [74] to develop a sophisticated power system status monitoring using a Big Data platform. According to [79], the distribution grid might be near-optimal observability shortly thanks to the improvement of real-time devices and AI-based technical solutions. A DNN approach for unobservable systems is presented in [80]. This work overcomes the computation complexity in Bayesian estimation, although it is less capable of adapting to outage changes, in addition to the fact that deep learning training algorithms are still under research.

A significant limitation of observability in distribution systems is the lack of sufficient real-time and high-granularity measurement devices such as PMUs. Although they are being deployed, their high cost prevents installing the required sensors to make the system fully observable.

2.4.2.3 Fault detection

The fault detection service intends to recognize and locate unusual electric currents within the distribution network. Two fault detection approaches are distinguished [81]. On one side, data-driven methods seek a pattern recognition of measurement readings gathered from sensors placed at different points of the network that indicate a fault. On the other, model-based approaches compare real data from sensors with prediction model results. High residual differences might indicate an electrical fault.

The leading causes of electrical faults are damaged equipment, environmental events, falling tree limbs and direct animal contacts [82]. Concerning natural phenomena-generated faults, [83, 84] propose AI-based approaches to predict blackouts due to convective storms [84] or ice-wind events [83]. Several works center their attention on the LV domain considering different data-driven methods [85–90]. Real-time anomaly detection is proposed in [91,92], where [91] defines an approach using smart mater large-scale consumers data, [92] formulates a Convolutional Neural Network (CNN) considering bus voltages. After a fault occurs, [93] proposes a DT approach to identify the power line-fault cause based on historical fault records.

Some studies focus on Microgrid faults detection [94–98]. A Random Forest (RF) classifier model is used to detect unexpected Microgrid islanding problems from normal operation conditions [94] that can be located by knowing the topology. [95] proposes RF, K-Nearest Neighbors (KNN) and SVM to detect faults in Microgrids, while [96] employs Extreme Learning Machine (ELM) for the classification and location of outages. An MLP classifier detects and isolates the fault in [97] and [98] applies the ensemble bagged DT classifier to detect dynamic events within the power system. Clustering techniques, such as Principal Component Analysis (PCA) [95] and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [84,99] are implemented. The following articles listed in 2.3 cover a fault diagnosis in PV systems [100–102], using different AI techniques and input data, as shown in Table 2.3.

Recently published articles that use AI techniques to detect and predict faults within the distribution network are listed in Table 2.3, considering the equipment of power lines [83, 84, 88] and underground cables [103], among others. In addition, the needed input data to perform the fault detection problem in each reference is specified.

2.4.3 Predictive maintenance

The predictive maintenance service intends to dig out the potentially valuable information from the collected sensor data located in the electrical equipment within the distribution network to help make de-

Table 2.3: Data-driven techniques for detecting faults in the distribution grid domain.

Data-driven technique	**Equipment**	**Data used for detecting faults**	**Ref.**
GBM	Distribution grid	Time, load, generation, current, voltage	[85]
DBSCAN, RF, DNN	Distribution grid	Area, lighting density, rain intensity, storm parameters, air temperature and pressure, wind speed, precipitations, snow depth	[84]
RF	Distribution grid	Fault data set	[86]
MLP, ELM	Distribution grid	Three-phase current signal	[90]
ELM, SVM, MLP	Distribution grid	Measurements of three-phase fault currents	[87]
SVM	Distribution grid	Traced current of distribution feeder	[88]
PCA, RF, KNN, SVM	Microgrid	Current and voltage signals at each endpoint of the line	[95]
ELM	Microgrid	Fault current signals	[96]
MLP	Microgrid	Current modules in the DC Microgrid	[97]
Ensemble DT	Microgrid	Distributed generator data	[98]
CNN	Power system	Bus voltage	[92]
MLP	Power line	Line impedance, reflection coefficient and the channel transfer function in the PLC signal band	[104]
CNN	PV systems	PV loop current	[100]
RF	PV systems	PV array voltage and string currents	[101]
DNN	PV systems	PV module parameters	[102]
DNN	Underground power cables	Power line modems	[103]

cisions on the scheduling maintenance actions to anticipate an imminent failure [105]. Furthermore, scheduled maintenance through predictive maintenance models is more cost-effective than repairing after failure [106]. Concerning the energy system field, the vast majority of the predictive maintenance research centers on high-power wind turbines as the aim is to reduce the high operating and maintenance costs. However, wind turbines are out of the scope of this investigation, which concentrates on the distribution level domain.

The following authors focus on power transformer predictive maintenance [107–113]. The study presented in [107] reviews and identifies the monitoring methods for predictive maintenance of electric power transformers and identifies the operational lifetime degradation factors. Another study conducted by [114] reviews recent articles that apply ML for predictive maintenance, including power transformers and PV panels. [109] analyses the different operating periods of an oil-immersed power transformer through dissolved gas concentrations data. The K-means clustering method groups the operation periods into different classes characterized by the production activities of several gases and the incipient failures. Reference [110] presents a predictive maintenance ML method for power transformers based on RF and Ada-boost. The results conclude that the Ada-boost algorithm provides better results than the RF. The main disadvantage of data-driven prediction maintenance research is that high-resolution power-quality data are mainly applied to validate their investigations. However, it is unlikely to have such data in real-life distribution grid scenarios.

Concerning real-world trials and companies, Enel Distribution utility tests its predictive transformer maintenance monitoring in [113]. These data are of great interest to the distribution utility, as it provides faster detection of anomalies, life loss and a more profound understanding of the grid for future expansions. The company Neuron Soundware [115] has developed a predictive maintenance solution powered by AI and IoT for power utilities, covering from transformers to motors, ensuring more than a 50% reduction in mechanical failures. Another company named Predictive Layer [116] offers a ML tool for selective predictive maintenance, among other energy-related use cases.

Recent studies containing AI techniques for developing predictive maintenance models are listed in Table 2.4. Besides, the equipment and data used to perform the prediction model are indicated.

Table 2.4: Data-driven techniques used for prediction maintenance in the distribution grid domain.

Data-driven technique	**Equipment**	**Data applied for predictive maintenance**	**Ref.**
Correlation analysis	Line conductors, cables, breakers and transformers	Equipment's component outage failure data	[108]
CNN	Photovoltaic panels	Electrical power signal	[117]
K-Means, PCA	Power transformer	Dissolved gasses concentrations	[109]
LSTM	Underground power cables	Voltage, active power and current.	[118]
MLP	Power transformer	Age transformer, loading, meteorological data	[111]
RF, Adaboost	Power transformer	Transformers' specification, loading, location and meteorological data	[110]
SVM	Power transformer	Prosumer data and infrastructure	[112]

2.4.4 Non-technical losses detection

Electricity losses at distribution levels encompass both technical and non-technical losses (NTL). The first one occurs due to Joule's effect, while NTL refers to the electricity consumed but not billed [119]. In other words, energy is illegally taken by unidentified end-users without the awareness of the energy utility. Detecting and addressing electricity theft is an essential task for power companies. For instance, Endesa has developed a fraud detection system currently in operation [120]. In 2018 the system was capable of detecting 65000 cases of electricity fraud, recovering 601 million stolen kWh. This number is equivalent to powering the Spanish city of Palma de Mallorca for a duration of six months [121].

Some articles review in-depth data-driven techniques applied within the NTL field. A comprehensive review is presented in [119], which compiles the primary techniques, including AI, and the data used to detect energy thefts, exposing the limitations of the current solutions. [122] focuses on Big Data oriented to anomaly detection, which is a powerful mechanism for fraud detection, while [123] examines the ML classifiers for electricity thief detection.

ML methods assist in improving the accuracy of fraud detection solutions. Unsupervised Learning clustering techniques are capable of grouping customers according to their consumption profiles, thus detecting suspect load curves of end-users. For instance, [124] calculates regular consumption behavior by clustering data collected from smart meters to identify NTL [124]. Moreover, distinguishing outliers in demand profiles aids to monitor and detect suspicious customers by identifying abnormalities in consumption patterns [125, 126]. The NTL classification algorithms achieve better performance results thanks to clustering techniques [127]. The most common evaluation metrics found in the literature for NTL classification approaches are accuracy, recall, precision, F-value and Matthews correlation coefficient. One of the main challenges found when building an NTL classification model is the lack of abnormal and irregular consumption data, a fact that is known as data imbalance. The following articles [128–130] take into consideration imbalanced data sets for NTL detection, while [131] proposes strategies for improving imbalanced data performance. Some events that could alter classification and clustering algorithms are the change of residents or the purchase of new devices, such as EVs [125].

The most common data source applied to identify NTL using AI techniques is the customer consumption data, followed in the distance by customer information (such as location, complaints made and overdue bills) and load, voltage and current measurements [119]. The authors in [132] prove that it is possible to use only a small data set of recent smart meter measures to define the customer consumption pattern. Besides, studies generally focus on residential consumption to detect electricity fraud, leaving aside industrial consumers. [133] justifies this event as industries do not have a fixed electricity consumption pattern.

To conclude, traditional theft detection methods are mainly based on on-site line inspections, which are highly expensive, time-consuming and inefficient. In contrast, AI-based NTL detection methods are superior to conventional methods in terms of accuracy, time-consuming and labor required, but more irregular and abnormal historical data are needed in order to train the models optimally. The most popular ML algorithm is SVM for classification tasks, while K-means is widely popular for clustering consumption patterns.

Recent data-driven related articles for detecting NTL are listed in Table 2.5.

Table 2.5: Data-driven techniques used for detection of non-technical losses.

Data-driven category	Data-driven technique	Ref.
Supervised Learning		
Classification	DT	[120]
	RF	[127, 134]
	SVM	[125, 128, 129, 131, 133–138]
	Ada-boost	[130]
	KNN	[133]
	GBM	[133]
	LogR	[134]
Unsupervised Learning		
Clustering	K-means	[125, 130, 132]
	SOM	[127, 132, 139]
	Fussy C-Means	[132]
	Gustafson-Kessel	[124]
Dimensional reduction	T-SNE	[133, 140]
Deep learning		
	CNN	[133, 140–142]
	LSTM	[141, 142]
Statistics		
	Bayesian network	[120]
	Pearson coefficient	[120]
	Outlier detection	[126]
Data Mining	Data mining	[120, 124, 126, 137, 139]

2.4.5 Forecasting

The forecasting service implements AI methods for demand, generation, electricity price and flexibility prediction, which are essential to deal with uncertainty and risk management within the distribution grid. Furthermore, energy-related forecasting provides essential input for demand response programs [143]. The purpose of this section is not to conduct an in-depth forecasting review but to expose recent studies that apply AI techniques within the forecasting domain. The most relevant and recent review articles are revealed in each subsection to allow the reader to delve deeper into the subject.

Forecasting horizons are classified in three categories [144–149]; short-term (ST), medium-term (MT) and long-term (LT); although some literature adds a fourth category: very short-term or real-time (RT) forecasting [147, 150–152]. Table 2.6 specifies what applications and time range covers each forecasting horizon. Most articles focus on short-term forecasting, as [148] also points out. Concerning long-term forecasting, they are influenced by economic growth, policy adjustment and technological advancement, making it a complicated task [149].

The following studies review different types of predictions related to the energy sector. For instance, [151, 153–155] investigate both load and price forecasting models. In addition, [149] reviews ML algorithms, ensemble-based approaches, and ANNs implemented for renewable energy generation, load demand and electricity price forecast.

Table 2.6: Forecasting horizon classification.

Forecasting horizon	**Time interval**	**Applications**
Real-time	t $\leq$ 1 hour	Keep the power system balanced [156]
Short-term	1 hour < t < 1 week	Deregulated electricity markets [148, 157–159] Real-time energy management systems [146, 147] Optimal management of power system [159]
Medium-term	1 week < t < 12 months	Asses environmental impacts, maintenance scheduling [148, 160]
Long-term	t > 1 year	Long-term investment and political decisions [147, 161] Grid expansion criteria. A decision tool for purchasing futures of the spot product [161]

2.4.5.1 Demand forecasting

The demand forecasting service predicts the consumption profiles of a single or several end-users. For instance, the day ahead aggregated demand forecasting of a particular zone is essential for the DSO to foresee possible congestion in the network. The emergence of prosumers -a consumer who produces and consumes energy- in the power system has caused an increment in the uncertainty of the demand profile, as consumption has become more unpredictable and volatile due to demand response programs and weather conditions that affect last stay the end-users self-consumption. The system operator does not have information regarding the self-consumption behind each smart meter; therefore, predicting consumption becomes even more complex.

Reviews focusing on AI load forecasting techniques are conducted in the literature. For instance, a study comparing conventional and AI-based models for energy forecasting is carried out in [162]. A systematic review is presented in [163] and it concludes that regression models are the most suitable for long-term scenarios, whereas ML algorithms outperform for the short-term forecasting horizon. The combination of different ML algorithms is analyzed in [164]. In a narrower framework, [162] presents several research papers that implement data-driven models for building scale forecastings. A survey of statistical and conventional methods for demand forecasting is presented in [165], concluding that Auto-regressive Integrated Moving Average (ARIMA) statistical model combined with ANN increases the accuracy of predictions. The authors in [166] review load forecasting methodologies based on previous literature, classifying them into four forecasting methods: similar patterns, variable selection, hierarchical forecasting and weather station selection.

The most common and relevant input data found in the literature for demand forecasting are mostly historical demand data along with seasonal factors like weather and calendar data [167, 168]. According to [143], essential parameters for system electricity demand are weather data and random effects, such as maintenance work or customer behavior, where the game theory approach helps predict erratic performance [169].

Three demand forecasting levels are distinguished: the aggregated demand within an area or zone, the aggregated demand in a building/-household through smart meters and the disaggregated demand, which predicts the consumption of electrical appliances behind the meter thanks to the implementation of sensors that store their consumption data (also known as sub-metering). Disaggregated demand forecasting plays a vital role in DR programs; hence, data are crucial to predict their consumption to make accurate and optimal load scheduling. On the contrary, in case of not having a sensor for each flexible load, the Non-Intrusive Load Monitoring (NILM) method identifies each asset's consumption curve in order to predict the power consumption of each appliance directly from the smart meter data. Recent literature covers different load identification methods that classify the assets behind-the-meter for NILM methods using AI techniques [170–175].

Unsupervised clustering techniques are implemented to classify buildings based on their energy efficiency [168], determine natural segmentation of customers [176], identify appliances usage patterns [175, 177], estimate electricity consumption behavior patterns in households [176, 178, 179], group households profiles patterns to achieve better forecasting outcomes [180–183] and to identify peak demand profiles or electricity theft detection [125]. Dimension reduction of NILM features is applied in [184]. DT and Naive Bayes (NB) are used to identify residential device loads [185].

Studies focus mainly on short-term prediction for network operation purposes. In contrast, there are barely any long-term studies. As for the latter, long-term forecastings require residential and non-residential inputs, such as historical gross domestic product or population, to estimate demand in the following years, applying data-driven approaches such as Multiple Linear Regression (MLR) analysis [186]. It is concluded that long-term prediction models combine energy, economic and environmental fields for planning the energy future in a sustainable manner [165].

Table 2.7 describes the AI techniques used for developing aggregated demand forecasting models for different time horizons and locations; meanwhile, Table 2.8 focuses on smart meter level.

Table 2.7: Aggregated demand forecasting models and evaluation metrics.

Data-driven technique	**Forecast horizon**	**Highlights**	**Evaluation metrics**				**Ref.**
			MAE	MAPE	RMSE	Other	
DNN, RF	ST	Electricity consumption for residential buildings for the next day.			•		[183]
MLP	ST	Wavelet decomposition to capture the various seasonal cycles in electricity load data.		•	•		[187]
DNN	ST	Advanced data preprocessing strategy. DBN has outstanding data learning and forecasting capabilities	•	•	•	MSE	[188]
ARIMA-WaveletNN	ST	The WNN has a strong ability to fit the nonlinear component of the electricity load	•	•	•		[189]
SVM	ST	The features extracted by the auto-encoders forecast day-ahead load forecasting more accurately		•			[190]
CNN-LSTM	ST	Day-ahead aggregated load forecasting based on two-terminal sparse coding and DNN	•	•		MSE	[182]
SVM, ELM	MT	ELM performs better than SVM for 1 week prediction	•	•	•	MSE	[191]
SVM	MT	Forecast next week electricity demand	•	•		R^2	[192]
MLP	MT	Optimal training algorithm composed of two-particle swarm optimization and ant lion optimization		•		MSE	[160]
MLR	LT	Hourly and annual electricity consumption estimation for 2030 in 14 different West African countries		•	•		[186]
Multiplicative error model	LT	Monthly aggregated load prediction for a horizon of four years		•		MSE	[193]

Table 2.8: Smart meter load forecasting models and evaluation metrics.

Data-driven technique	Forecast horizon	Highlights	Evaluation metrics				Ref.
			MAE	MAPE	RMSE	Other	
LSTM	RT	Demand forecasting from an industrial steel plant	•	•	•		[156]
LSTM	RT	Probabilistic household load forecasting under high uncertainty and volatility	•		•		[194]
DT, RF	RT/ST	Focuses on online environments where data are analyzed as they arrive	•	•	•		[195]
DNN	ST	Aggregated households load forecasting	•		•	R^2	[196]
LSTM, MLP	ST	Energy Big Data is used as data set for load and price forecasting	•		•		[153]
LSTM	ST	The model aims to learn the uncertainty by applying a pooling Deep RNN. It is tested in 920 smart meters	•		•		[197]
CNN	ST	Single residential load forecasting using CNN combined with data-augmentation technique		•	•		[198]
DNN	ST	Power load and probability density forecasting	•	•			[199]
LSTM	ST	Individual and aggregated residential load forecasting		•			[200]
MLP, SVM, MLR	Short/LT	SVM and ANN achieve the best outcomes			•		[201]
K-means-MLP	ST	Day-ahead office building cooling demand, grouped in seasons		•			[202]

2.4.5.2 Generation forecasting

The generation forecasting service aims to predict the electricity production of renewable sources within the distribution network. The continuing increase of renewable energy sources and demand-side flexibility programs in power systems has raised the need for more accurate Renewable Energy Systems (RES) predictions. Regarding recent studies that evaluate AI methods applied to RES prediction models, authors in [149] review multiple renewable generation sources, including distributed wind, solar and geothermal energy, considering various forecasting horizon ranges. Key findings state that benchmark ML models handle a large amount of data with accurate forecasting outcomes; however, ensemble ML models could achieve even further accuracy by combining different data-driven techniques. Concerning solar generation, [203] presents a review on PV forecasting based on ML and metaheuristic techniques while [204] focuses on time-series statistical, physical and ensemble methods. A review of the state of the art of SVM in the application of solar and wind forecasting is conducted by [205]. SVM regressor is simple-to-use and reliable, but on the contrary, it is not suitable for large data sets and it has a low performance for high noise data.

Multiple ML methods such as ANNs, SVMs and Gaussian Process Regression are studied in [206] for wind and solar power generation. The authors in [207] extensively compare simple and sophisticated PV forecasting methodologies and conclude that some methodologies are more suitable under different weather conditions. The work presented in [208] studies day-ahead PV forecasting models based on deep learning neural networks. Multi-site PV forecasting is examined in [209] using CNN. In [210], a review of the main ML methods for forecasting wind speed and power is carried out, including weighting-based, data preprocessing, parameter selection, optimization and error processing methods. These combined approaches generally outperform the single models approach. Authors in [211] present a review of ANN implemented in wind energy systems, combining the main methods applied in forecasting models, and identifying strengths and weaknesses. The Wavelet transform is used to decompose the raw data into different frequencies. It is applied to mitigate spikes and fluctuations in raw data [212]. This method has been implemented in several studies [212–215]. A problem faced by the DSO and BRPs is the lack of knowledge of the aggregated small-scaled solar generation of prosumers. To solve this problem, [216] estimates the aggregated power generation of small-scale rooftop solar sites that are not monitored by system operators.

Finally, Table 2.10 shows the AI techniques applied for developing distributed generation forecasting models for different time horizons and locations, identifying the main evaluation metrics.

2.4.5.3 Electricity price forecasting

This service aims to predict the electricity price, which is essential for minimizing the energy purchase invoice for BRP and retail companies in the short-term horizon. The most relevant and recent research regarding Big Data and AI methods for electricity price forecasting is conducted in [147, 148, 157, 158, 201, 230, 231]. ANN models for day-ahead market price forecasting are reviewed in [157, 158], and [158] concludes that simple ANN models do not perform properly when electricity price time-series present high volatility, sharp price spikes and chaotic and non-linear behavior. Therefore, more sophisticated techniques are required to handle complex predictions. Techniques based on univariate and multivariate forecasting models are compared and covered in [147], while [231] focuses on benchmark techniques, from statistical to ensembles. Authors in [148] classify electricity price predictions in market equilibrium, structural, statistical, intelligent and combination models, separating short, mid and long-term estimations. Lastly, feature engineering for linear, ensembles and deep ML models is studied in [232].

Most literature focuses on short-term electricity price forecasting while medium and long-term predictions are not covered in sufficient depth, as [158] also points out. However, [148, 161] overviews electricity price forecasting for the mid and long-term, which is essential for distribution network planning purposes.

Deep learning models are widely used in literature to estimate electricity prices [153, 154, 231, 233–238], along with SVM regression [234,

Table 2.9: Solar forecasting models and evaluation metrics.

Data-driven technique	Forecast horizon	Highlights	Evaluation metrics				Ref.
			MAE	MAPE	RMSE	Other	
LSTM	RT	Hourly day-ahead solar irradiance prediction by using weather forecasting data			•		[217]
LSTM	RT	Five-minute forecasting horizons. Model-based on short-term multivariate historical data sets		•	•	R^2	[218]
LSTM	RT	Predicts the PV power in the next hour			•		[219]
CNN	ST	Thanks to the CNN advanced feature extraction, more meteorological features are introduced in the prediction mode	•		•	MASE	[220]
DNN	ST	Based on particle swarm optimization and trained feed-forward neural network (FNN)		•			[215]
GRNN, ELM, ElmanNN	ST	Predicts also the PV output associated uncertainty at different confidence levels		•			[221]
MLP, DNN and LSTM	ST	PV prediction only with meteorological and calendar data. LSTM algorithm presents the best outcomes for all seasons	•		•		[222]
SVM	ST	Model based on SCADA and meteorological information		•			[213]
MLP	ST	Correlation analysis of main variables. High accuracy			•		[223]
LSTM	ST	Uses the attention mechanism to focus on the most significant input features in forecasting	•	•	•		[224]
CNN	ST	Multi-Site Photovoltaic Forecasting		•	•	MASE	[209]

Table 2.10: Wind power forecasting models and evaluation metrics.

Data-driven technique	Forecast horizon	Highlights	Evaluation metrics				Ref.
			MAE	MAPE	RMSE	Other	
CNN-GBM	RT	Ultra short-term wind power prediction. Light-GBM improves performance of single CNN	•			MSE	[225]
RF	RT	The spatial average of the wind speed, its direction and past power values are the inputs.	•	•	•	MASE	[226]
LSTM, SVM	RT	Performs ten-minutes and one-hour ahead forecasting with extremal optimization	•	•	•	R^2	[227]
LSTM-ElmanNN	ST	The wavelet transformation is used. Its performance is compared with nine models	•	•	•		[214]
K-Means-LSTM	ST	The K-Means forms clusters of wind power impact factors to generate a new LSTM sub-prediction model.	•	•	•		[228]
Ensemble DNN	Short/MT	Uses a deep sparse auto-encoder and transfer learning during the training phase of base-regressors	•		•		[229]

235,239,240] and tree-based models [195,241]. RNN has been proposed to address time-dependent learning problems. In particular, LSTM and Gated Recurrent Units (GRU) have an extraordinary performance for time series price estimation according to [237]. The most accepted inputs are historical electricity prices and calendar data. Electricity price has a strong correlation with other variables like oil and natural gas price [240] if the energy mix is highly carbon-dependent. ELM techniques improve the generalization performance and learn faster than ANN trained using back-propagation [155,242–244].

A summary of the recent AI techniques applied for electricity price forecasting in literature is presented in Table 2.12 for different forecasting horizons and locations, together with the main evaluation metrics used. Finally, Figure 2.5 displays a bar graph showing the most used evaluation metrics in recent literature within the forecasting field.

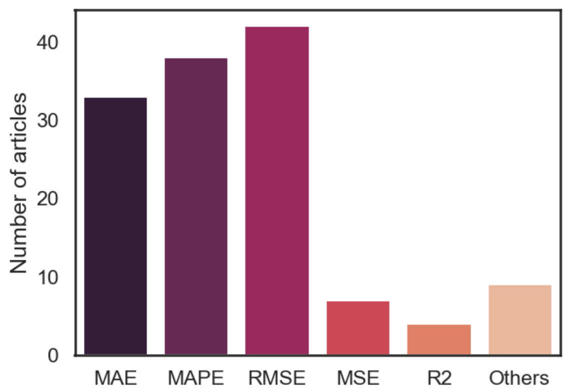

Figure 2.5: Forecasting evaluation metrics used in literature.

2.4.5.4 Flexibility forecasting

The development of aggregated flexibility forecasting services permits to delimit the accumulated feasible flexibility in a default area by aggregating flexible loads, distributed/centralized storage units and Distributed Energy Resources (DER) [248]. Using the aggregated flexibil-

Table 2.11: Price forecasting models inputs and evaluation metrics.

Data-driven technique	Forecast horizon	Highlights	Evaluation metrics MAE	MAPE	RMSE	Other	Ref.
CNN-LSTM	RT	Hybrid DNN performs better than traditional ML models	•		•		[233]
GBM	RT	Accurate and computationally inexpensive	•	•	•		[152]
SVM, MLP, DNN	RT	DNN obtains less error. Higher accuracy achieved by diversifying the data source			•		[235]
ELM	RT	Improves the forecast accuracy in real-time when an unexpected dynamic price change occurs.					[242]
Dynamic Trees	RT/ST	Dynamic Trees perform better than RF and are an adequate method for real-time and short-term	•	•	•		[195]
SVM, LSTM	ST	DL model outperforms the SVR			•		[234]
DNN	ST	Good performance for high volatility prices			•		[245]
DNN	ST	Inconsistencies were observed as layers were increasing when using a few input variables. Model performs better with more historical data		•			[236]
DNN, LSTM, GRU, CNN	ST	Compares the 4 proposed DNNs with 23 benchmark models for electricity price forecasting. DNN, LSTM, and GRU outperform literature models.		•			[231]
Hybrid outlier-ELM	ST	The model can be a reliable forecasting method in modeling time series with complex nonlinear characteristics and outliers	•	•	•		[243]
Neuro-fuzzy ANN	ST	This study reveals the efficiency of neuro-fuzzy models against MLP neural network and ARIMA statistic model	•	•	•		[246]
GBM	ST	Hour feature is the most relevant predictor in the model	•	•	•		[241]
ELM	ST	MKELM model provides better performance as compared to the ELM and KELM	•	•	•		[244]
Dimension reduction, DNN, SVM, LSTM	ST	Proposed method is recommended for studies with a large volume of input data. The feature extraction tool and rough neurons improve the forecasting results.	•	•	•		[159]

Table 2.12: Price forecasting models inputs and evaluation metrics.

Data-driven technique	**Forecast horizon**	**Highlights**	**Evaluation metrics**				**Ref.**
			MAE	MAPE	RMSE	Other	
MLP	ST	This study demonstrates the improvement in convergence speed with Tensorflow software	•				[44]
SVM	ST	Oil and natural gas prices are considered in the prediction model due to their high correlation with electricity prices				MSE	[240]
LSTM	Short/MT	Deep LSTM gives better results compared to ELM and NARX	•		•		[153]
GRU	Short/MT	The three-layered GRUs outperformed all other ANN structures and statistical techniques. Stacking multiple layers increases the performance.	•				[237]
Weighted KNN, DNN	MT	DNN outperforms Weighted KNN, a model based on autocorrelations in data, providing good accuracy forecasts even 29 days ahead		•	•		[238]
Jaya-LSTM	MT	Hyper-parameters tuned using Jaya optimization multivariate LSTM algorithm leads to better performance than SVM and univariate LSTM	•		•		[154]
Co-integration and vector error correction	LT	Brent crude oil spot and futures price along with the Spanish wind generation are the variables that yield the most accuracy		•			[247]

ity in the distribution grid reduces the need for grid extension [249] and enhances the technical and economic power system operations [250]. In [251], the aggregated flexibility calculation is discussed in more detail.

The aggregator is the service provider in charge of gathering and controlling its portfolio flexibility sources [252,253] in order to i) provide flexibility services to power system agents, ii) to minimize the end-users energy bill through Home Energy Management Systems and iii) to participate in electricity markets, by using optimal bidding strategies [254, 255]. New flexibility business model approaches are developed in [256]. Due to the increasing penetration of intermittent RES and the significant number of residential users with potential flexible sources, demand-side flexibility aggregation becomes essential for balancing the future power system [255, 256].

ML-based regression models are applied in [250] to forecast the flexibility of residential customers for real-time applications. A flexibility forecasting concept and its control from multiple energy domains and sources are presented in [249]. The GBM ensemble algorithm is selected by [257] to build a flexibility load forecasting model for DR capacity scheduling.

Intending to encourage a change in the demand-side consumption, the aggregator offers a monetary incentive signal in [258]. The flexibility potential of wet appliances in France (dishwasher and washer machine) is estimated in [259]. In [256], residential load flexibility forecasts are calculated using the NILM approach. Predefined customer preferences and loads and PV forecast uncertainty are considered in [260] to define a feasible flexibility space from controllable residential resources. [261] studies the required percentage of end-users with sub-metering capabilities needed to calculate the aggregated demand composition. The results state that only a 5% of sub-metering coverage is required to forecast the aggregated load composition at the substation level with high confidence. A scalable and non-intrusive model for identifying the flexibility of thermal loads is proposed in [262].

2.4.6 Energy management systems

The massive amount of data generated by the rapid deployment of smart meters in recent years supports the enhancement of building energy efficiency and DR programs (e.g., price-based, incentive-based and environmental-based [263]). This subsection focuses on the IA techniques used in EMS at the building level.

The application of AI methods allows for overcoming multiple challenges related to energy management, developing better tools for automatic decision-making to schedule and control multiple energy assets through the EMS. From 2013 onwards, there has been a perceptible increase in AI approaches across DR applications. These AI methods have been principally applied to price-based programs and residential consumer types, followed by small-scale industrial and commercial buildings [264].

Numerous papers have reviewed AI approaches for energy DR programs. In a more general context, [264] investigates the state of the art of DR applications and analyses the AI methods applied in different DR scheme categories and consumer types. An extended summary of companies, start-ups and European-funded industrial projects using AI for DR is also provided. Regarding home appliance schedule controllers for DR programs in smart households, [265] reviews various AI techniques based on ANN, fuzzy logic control and adaptive neural fuzzy inference system, which imitate human thinking behavior. More specifically, [266] reviews the existing AI-based methods for cloud EMS with the integration of blockchain technology. However, the high development cost and storage of blockchain and the lack of standardization and professional expertise in this topic represent a research challenge that should be addressed in the upcoming years.

The following studies apply Supervised and Unsupervised Learning techniques. A MLP deep learning model is used in [267] to optimize load consumption and storage management in response to dynamic pricing. A deep ANN and Genetic Algorithm reduces energy demand in peak periods, optimizing the residential appliances scheduling and RES generation [268]. Supervised Learning algorithms as DT and Naive Bayes identify loads through smart plugs [185] for EMS. [269] creates a control-oriented model for a heating system based on regression trees and RF. A steady price prediction model based on ANN deals with price uncertainty for EMS in [270]. [271] develops a residential scheduling controller using the hybrid lightning search algorithm ANN to predict the optimal ON/OFF status for home electrical appliances after a DR event imposed by the power utility to reduce peak consumption. [272] uses ANN in order to predict and schedule building appliances' energy consumption and genetic algorithms for task scheduling, while [273] develops a prediction method based on LSTM of the end-user response behavior to incentive-based DR program.

In recent years, RL has gained prominence in studying intelligent management and control of buildings and households. The main advantage of using RL algorithms instead of optimization techniques is that the RL algorithm can automatically learn the customer preferences imitating human behavior and determining optimal incentive rates that can maximize the profits of both energy service providers and customers fairly and efficiently. [274] narrows its research to a group of energy systems that use RL to control the assets that have the potential for DR applications. A DR price-based approach using deep RL in an industrial facility is conducted in [275]. It is noteworthy that the algorithm was tested in a real-world utility company, reducing energy costs and ensuring production targets as well. The authors in [276] perform a simulated-based followed by a lab experiment of an electric water heater cost of energy consumption minimization given an external price profile using RL techniques.

RL is used as a decision-making tool in EMS for scheduling and controlling flexible units such as EV [277, 278] and other flexible controllable loads [270,271,276,278–284], Energy Storage Systems (ESS) [282] and PV generation [278, 282, 284], in order to solve different problems like hour-ahead [270,278] and day-ahead [280,283] energy consumption scheduling [270, 278]. The Q-learning algorithm is the most common RL method applied in DR programs [264,270,278,280,282,284]. The Q-learning algorithm needs many iterations to converge, while the batch RL usually converges much faster [285]. Batch RL is applied in thermal controlled loads in order to find the day-ahead schedule [276, 283] to minimize costs.

In most of the articles reviewed, there is a lack of RL experimental results in real small and large-scale physical systems since many of the proposed methods are only tested in simulated environments. [264,274]. The scarcity of physical experimentation could be one reason that prevents buildings and households from adapting RL algorithms, as their reliability and performance in real-world scenarios are unproven yet. Moreover, the vast majority of the reviewed articles are single-agent system based, which means that they only focus on a single building, bypassing urban boundary conditions. The single-agent approach is correct when very few buildings participate in DR programs. However, if a large number of buildings use DR schemes, a Multi-agent approach is needed in order to address the computer limitations problems of centralized approaches by distributing the workload among the participating agents in order to make decisions for various buildings devices in a decentralized manner while maintaining data privacy of costumers. Thus, the multi-agent approach avoids shifting the peak demand to lower-cost periods, for instance.

The main drawback of the articles reviewed in this subsection is that they assume complete knowledge of the end-user environment, although this is unlikely to happen in reality. It is worth mentioning that the primary focus in literature is on price-based DR programs; nevertheless, a more comprehensive range of incentive-based DR schemes should be developed and tested, as it is indispensable for the optimal operation and balance of the distribution network. Only a few articles modeled the EMS controllable appliances with a high level of detail. Unlike the traditional model-based methods, the RL approach does not require any system model information. Finally, Table 2.13 classifies the data-driven technique, DR program and customer type for each reference.

2.4.7 Aggregated flexibility services

The aggregated flexibility service is responsible for gathering flexibility from different customers and offering flexibility services to potential energy agents such as residential and industrial clients, BRPs and DSOs. Benefits derived from the flexibility and DR programs include shifting or reducing peak demand, meeting the fluctuations of renewable generation, enabling higher penetration of renewable generation and customer bill reduction. Nevertheless, there are still many challenges

Table 2.13: AI methods and DR schemes used in EMS.

Data-driven method	**DR program**	**Consumer type**	**Ref.**
Supervised Learning			
ANN	Price-based	Residential	[267]
ANN	Price-based	Residential	[268]
ANN	Price and incentive-based	Residential	[271]
ANN	Price-based	Residential	[272]
RF	Price-based	Residential	[269]
Deep Learning			
LSTM	Incentive-based	Residential	[273]
Reinforcement Learning			
Deep RL (Single-agent)	Price-based	Residential	[276]
Deep RL (Single-agent)	Price-based	Residential	[281]
RL (Multi-agent)	Price-based	Residential	[270]
RL (Multi-agent)	Price-based	Residential	[278]
RL (Single-agent)	Price and incentive-based	Residential	[279]
RL (Single-agent)	Incentive-based	Residential	[280]
RL (Single-agent)	Price-based	EV management	[277]
RL (Single-agent)	Price-based	Industrial facility	[275]
RL (Multi-agent)	Price-based	Residential	[278]
RL (Multi-agent)	Price-based	Residential	[282]
RL (Single-agent)	Price-based	Residential	[283]
RL (Single-agent)	Price-based	Residential	[284]

to be addressed, such as improving the accuracy of the flexibility prediction models or selecting the best suitable customers for engaging DR programs.

As discussed earlier, it is necessary to add value to the massive data generated within the distribution network and create aggregated flexibility services for electricity market stakeholders using the most appropriate AI techniques. The aggregator is a relevant player contributing to flexibility aggregated services through DR incentive-based programs within its portfolio (a set of clients). The role of aggregators in the Smart Grid context is studied in [286]. Electricity market actors can request flexibility to avoid grid congestions (e.g., DSO) or imbalance penalizations (e.g., BRP). Thanks to the aggregation of individual customers, the total amount of flexibility available increases considerably. Thus, the end-users change their consumption pattern in exchange for economic compensation through incentive-based approaches [287].

Finding the best-suited customers to provide the flexibility requested by an electricity market agent is a computational challenge for the aggregator, especially with portfolios with a considerable amount of flexible resources. To cope with this challenge, [288] proposes a cluster-based (K-means) day-ahead bidding optimization approach that reduces the optimization execution time. Moreover, the ADMM technique is capable of solving large-scale optimization problems by breaking them into smaller pieces [289–291]. For instance, [289] formulates a cost minimization problem to provide flexibility services to DSO and BRP using ADMM to improve computational performance. Moreover, [290] applies ADMM for bidding optimization strategy in the day-ahead and secondary reserve markets. In [291], an ADMM-based market-clearing strategy is presented for day-ahead congestion management, using aggregated EVs and heat pumps as flexible sources.

The DSO can request demand-side flexibility from the aggregator to mitigate possible congestion in the distribution network. Day-ahead congestion management is proposed in several works [291–294], while [295] formulates real-time congestion management for the unforeseen events that might occur during operation. On the other hand, grid congestions can also be avoided with dynamic pricing strategies in order to encourage customers to flatter their demand curve [267,270,275–277, 279, 284].

Aggregators can attend both DSO and BRP flexibility requests for day-ahead and intra-day portfolio optimization. With the intention to address this issue, the traffic light system proposed in [296] is carried out in [289, 297] to coordinate the flexibility requests from DSO and BRP and establish a priority criterion for providing flexibility in the case of conflicting requests. Recent works use the aggregation of thermal loads [298, 299], Heating Ventilating and Air Conditioning (HVAC) [300, 301] and EVs along with heat pumps [291, 292, 295] in order to provide flexibility to energy agents. On the other hand, some studies focus on optimal bidding strategies for aggregators participating in the electricity market [255, 288, 290, 302], where [288, 302] participate in the day-ahead market, while [255, 290] are also involved in the reserve market.

In addition to the flexibility services mentioned above, electricity companies also desire to segment their large number of customers according to similar demand patterns to have insights into their energy usage behavior, enhancing the distribution network operation and management. Moreover, more customized products and services can be offered to each customer target group [303]. Unsupervised Learning techniques enable customer segmentation according to the consumption pattern, thanks to the smart meter measurements [264]. Diverse customer demand-based clustering studies are proposed in the literature; for instance, a robust comparative review of 11 clustering techniques applied to residential load time series profiles is carried out in [304]. The study concludes that centroid-based (Kmeans) and hierarchical algorithms are the best performers, whereas the density-based methods (such as DBSCAN) performed the worst for this kind of problem. On the other hand, [305] reviews the clustering methods for customers' consumption patterns. The K-means algorithm is the most widely used, followed by Fuzzy C-Means, hierarchical and Self-Organizing Map (SOM), being the latter the worst performer of a 4000 customers segmentation case study. A combination of SOM and K-means algorithms are used for analyzing industrial parks' energy consumption patterns [306]. The main drawbacks of using SOM are that the results are not intuitive at first glance and are computationally expensive compared to K-means.

2.4.8 Trading

The trading energy service focuses on one of the biggest challenges of the forthcoming energy transition: finding a reliable way to exchange energy between different customers, local energy communities and operators.

Based on the blockchain concept, Distributed Ledger Technology (DLT) is positioned as a benchmark technology in the P2P trading field, enabling smart contracts between prosumers and active users. The potentials of DLT for P2P transactive energy exchanges and its infrastructure in Local Energy Markets are detailed in [307], while [308] studies the DLT requirements and use cases. The authors in [309] carry out a detailed analysis of the concept, principles and types of blockchain and how this technology will revolutionize the green energy management of the future. The main advantage of the decentralization and automation of smart contracts is eliminating the human-based central authority, which implies lower settlement fees, simplified operational processes because of fewer intermediaries and a greater transparency level, thus avoiding corruption. However, the DLT is still immature and has not yet been tested on large-scale trials. In addition, the high computational cost of smart contracts raises the question of whether it is economically viable or not.

Decentralized blockchain mechanisms [310] enable reliable energy flexibility trading between the stakeholders involved in the flexibility market. DLT for P2P ancillary service markets in distribution networks is studied in [311,312], where the last article applies the ADMM technique to settle the P2P trading. Smart contracts for DR programs are formulated and the total incentives for an energy prosumer are calculated. Four different P2P and smart contract implementation approaches are conducted in [313], where producers and consumers send their offers and bids accordingly with smart contracts in the energy market. In terms of security, [314] proposes a RNN that detects network attacks and fraudulent transactions within blockchain-based energy transactions. Besides, a novel transactive controller is developed to manage the storage unit of a residential prosumer. The research [315] also studies P2P trading for managing the ESS and the surplus renewable energy in a smart energy community.

Focusing now on data-driven methods, a deep RL approach is adopted to address an energy trading decision-making problem for Microgrids [316]. The work in [317] implements LSTM for blockchain-based predictive energy analysis, intending to enable accurate short and long-term demand forecasting to minimize the cost of delivering electrical energy for the consumer and making effective policies. Similarly, [318] develops a P2P market based on deep learning (ELM), which learns the interaction between prosumer bidding actions and market responses from historical transaction data.

2.4.9 Asset and investment planning

The ever-increasing volume of stored measured data in the distribution network is expected to benefit the planning and operation of future power systems [281]. The investment planning service examines grid status and expansion criteria and selects the most appropriate technologies and optimal geographical locations. The objective is to contribute to the grid support during a settled planning horizon and estimate the associated costs for achieving a specific planning goal or criteria while meeting the forecasted demand.

ML forecasting methods play a critical role in mid and long-term renewable energy and demand predictions, which are essential inputs for the country's energy mix development and planning [149]. The authors from [319] propose a basic learning neural network to determine how Microgrids can be optimally planned and designed. Dimension reduction and correlation techniques are adopted for optimal planning for capacitors in [320]. Optimization techniques are applied to minimize total costs and investments for power distribution system planning in [321]. Flexibility is taken into account in [322] using a generic multi-stage distribution grid planning approach, while [323] studies in-depth network expansion under a DR scheme.

In recent years, some efforts have been made for long-term energy planning. For instance, long-term demand and renewable energy forecasting models are an energy planning tool [149]. However, there is a lack of research exploring AI techniques for medium and long-term distribution grid planning. Consequently, more efforts need to be taken in this field.

2.5 Data-driven services in distribution systems

2.5.1 Distribution grid services dependencies

Based on the comprehensive and meticulous analysis of the distribution grid services carried out in the previous section, it can be concluded that there are interdependencies among specific services. Consequently, the output of one service might be a fundamental piece of information for the execution of another. These interconnections among energy services are represented in Figure 2.6 through a flow chart. The green boxes represent the services related to the operation of the distribution network and the blue boxes outline the services related to flexibility management. Finally, the light orange box represents the planning services. In order to facilitate understanding of the flow chart, the outcome-dependence of each energy service is explained subsequently.

- **Topology**. This service offers the actual structure of the distribution network, which is often poorly known. The LV grid topology is a required input for the optimal performance of the following services. *Observability* -to calculate real-time network state estimation-, *fault detection* -to identify and locate outages-, *non-technical losses detection* -to detect possible frauds in the distribution network-, *distribution network planning* -to design optimal long-term investments and cost operations in the LV network- and last but not least, the *aggregated flexibility service* -to give the DSO a view of the network structure so that he can formulate flexible requests to minimize congestion-.
- **Observability**. The observability service outcome provides real-time knowledge of the LV network state, which is essential for grid operation. The *fault detection service* requires the LV network status to identify the precise location of a failure in the power grid in real-time.

- **Predictive maintenance**. This service is responsible for predicting the probability of failure of distribution grid components. Identifying the network components' state and health enables the *distribution network planning service* to optimally plan investments and operating costs in the long term.
- **Price forecasting**. This service predicts electricity price forecasting for short or long-term horizons. The *aggregated flexibility service* needs price forecasting to optimize the BRP portfolio purchase bids before buying energy in the day-ahead electricity market.
- **Generation forecasting**.The outcome is a RES forecast for short or long-term horizons. The *energy management system service* necessitates short-term RES generation in order to schedule the flexible resources optimally to minimize the electricity bill. In contrast, the *distribution network planning service* needs long-term RES forecasting as input -more than five years ahead- to minimize the LV network assets' investment and operating costs in the distribution grid.
- **Demand forecasting**. The outcome is a demand forecast for a short or long-term horizon. This service follows the same process as the last bullet point; the *energy management system service* needs short-term load prediction, while the *distribution network planning service* needs long-term demand forecasting as input data.
- **Flexibility forecasting forecasting**.This service determines the flexibility available in a zone or area of the distribution grid within a time horizon. The *aggregated flexibility service* makes use of this prediction, so the electricity market stakeholders - DSO, BRP, for instance- know in advance the flexibility available in the zone or area. The *energy trading service* also needs flexibility forecasting in order to schedule the energy trading optimally.
- **Energy management system**. The end-users consumption of a zone or area is aggregated and sent to the *aggregated flexibility services*. This information is useful for specific stakeholders like the DSO to detect possible congestions in the grid or the BRP to identify how much energy its portfolio will consume the fol-

lowing day after optimizing its consumption through the energy management service.

- **Aggregated flexibility services**. The outcome of this service is the flexibility requests of specific electricity market agents such as DSO or BRP. The flexibility requests of these agents are key input information for the *energy trading service*.

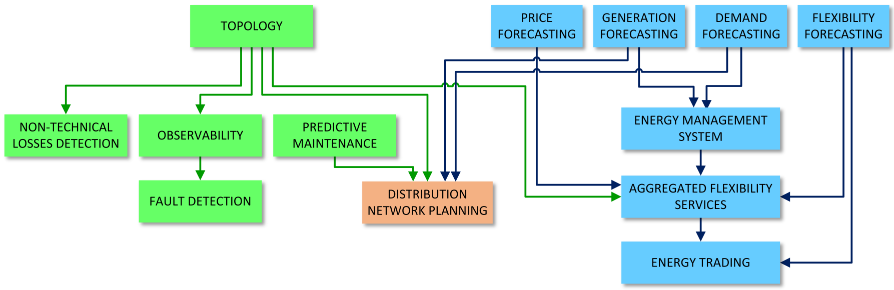

Figure 2.6: Distribution network services dependencies flow chart.

2.5.2 AI-techniques applied in distribution grid services

Certain services are more likely to implement AI techniques than others, particularly those that require prediction, classification or clustering tasks. To facilitate the understanding and interrelation among the AI techniques and the data-driven energy services, Figure 2.7 presents a chord diagram. This graphical representation displays the connections between the data-driven methods and the energy services, with the arc size corresponding to the flows' significance. Specifically, the arc sizes in this diagram indicate the number of publications in the literature that have applied AI methods to the respective energy services.

The creation of the chord diagram involved the following steps:

- Identification of distribution grid services and their associated AI techniques. This information has been progressively gathered and collected throughout the different sections comprising this chapter.

- Creation of a matrix, listing the distribution grid services vertically and the AI techniques horizontally (or vice-versa). This matrix is created by documenting the total number of papers presented throughout this chapter for each AI technique within each distribution grid service.
- Generation of the chord diagram using data visualization tools or programming libraries designed explicitly for chord diagrams. The matrix data completed in the previous step is used as input into the chosen tool. Visual elements such as colors, labels, and interconnections are customized to enhance the clarity and comprehension of the diagram.

The ribbons amplitude is equivalent to the number of articles that have adopted this technique. For a better visual perception, the data-driven methods that appear in less than three studies have been eliminated from the chord diagram. Figure 2.7 displays the AI methods first in a clockwise manner, sorted into the four categories proposed in Section 2.4, followed by the distribution grid services.

The services are analyzed in order of appearance. The Measurement Error Detection service (**MED**) is powered by ML regression-based algorithms for assigning predictive values for missing data. The topology service widely uses Correlation and ADMM methods (**TOP**). The Observability service (**OBS**) barely holds any articles that use data-driven methods. In the Fault Detection service (**FD**), there is a range of most used AI techniques for fault detection classification tasks: RF, SVM and MLP. The RF classifier outperforms the other ML techniques, as [86] also points out. Regarding the Predictive Maintenance service (**PM**), no technique predominates over another. The PM service mainly covers classification to clustering algorithms to classify or group potential failure events. Concerning the Non-Technical Losses service (**NTL**), using SVM to classify and detect future fault events in PV systems or power generation equipment is predominant. However, multiple data-driven techniques such as DM, CNN, K-means and SOM are also employed to detect these faults.

The Forecasting service (**FOR**) is the most AI-intensive since several investigations concentrate their research on predictive models applying AI procedures. The LSTM is the most utilized in the time series fore-

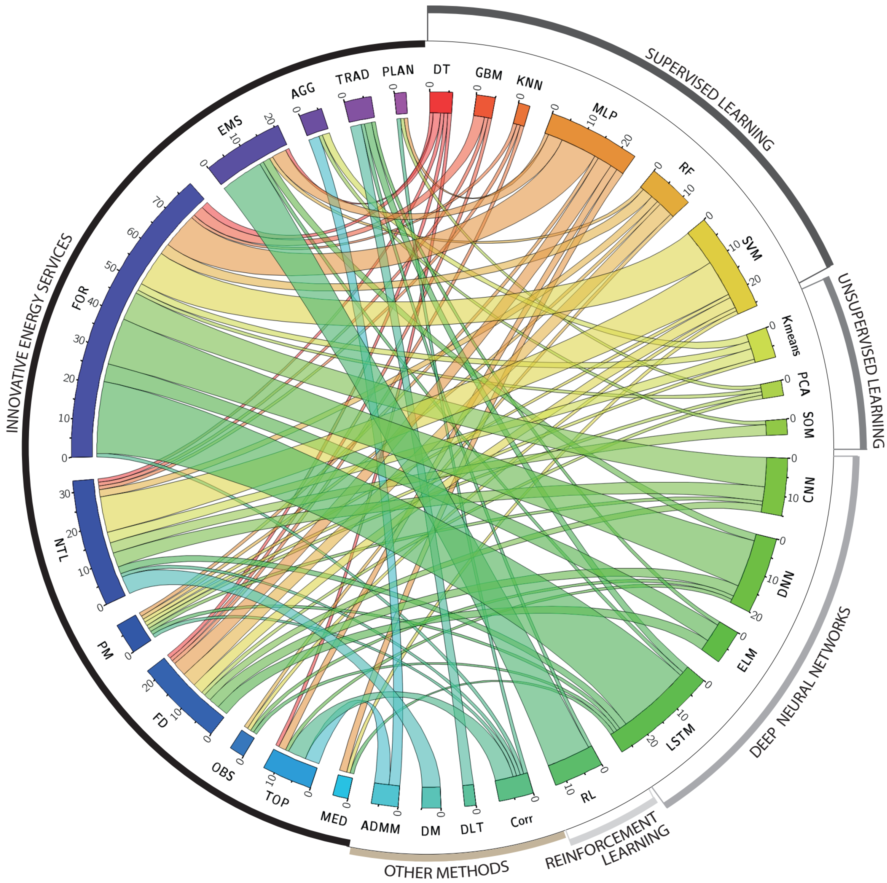

Figure 2.7: Chord diagram that represents the interrelation between services and their most-applied AI techniques in recent literature.

casting field in recent years, followed by a single hidden layer MLP algorithm. DNN with more than one hidden layer and CNN are utilized to a lesser extent. Therefore, RNNs are appropriate for time series data since they use temporal information from the input data. Thus, the LSTM is the best suited when predicting a times series outcome considering the model can associate the data of the previous time and the present time thanks to their recurrent architecture and memory units. The most common evaluation metric for regression tasks is the RMSE, followed closely by MAE and MAPE. Concerning the Energy Management System service (**EMS**), besides the optimization methods that are not in the scope of this thesis, RL is used as a decision-making tool for scheduling and controlling flexible assets. This flexibility can be used for the end-user benefit, such as reducing electricity costs considering the customer's comfort or being sold to a third party involved in the electricity market through an aggregator in exchange for monetary compensation.

The rest of the remaining services are not as dependent on IA techniques. The Aggregated Flexibility Service (**AGG**) uses mainly optimization techniques to minimize the cost of providing flexibility instead of AI methods. Clustering is applied to group customers with a similar consumption profile, thus detecting possible congestion or unusual behavior. Moreover, to cope with the massive amount of data, this service uses ADMM to relax computing complexity when aggregating flexible resources that provide flexibility, dividing the optimization problems into separated parts. Therefore, ADMM is a valuable and recurrent solution to deal with energy services that need a large amount of data as input. The Trading service (**TRAD**) relies mainly on the DLT method, which enables P2P trading and smart contracts between prosumers and active users. Lastly, the Planning service (**PLAN**) uses some data-driven techniques, such as dimension reduction and correlation techniques. Still, it is mainly a service that uses optimization procedures instead of AI methods. The main benefits of implementing AI methods within the distribution grid domain are addressed:

- Allow real-time distribution grid status estimation and gain observability, enhancing the monitoring and locating possible events in the network to provide a tool that enables the operator to react more rapidly when a fault or event occurs.

- Performing the predictive maintenance service increases the distribution network security and availability while diminishing the DSO costs.
- Use the aggregated flexibility available to avoid grid congestion.
- Optimal and reliable electricity trading among customers.
- Optimal medium and long-term distribution grid planning.

2.6 Opportunities and challenges

Energy sector organizations are increasingly interested in using data science and AI capabilities to solve their daily challenges. However, Big Data techniques applied to the energy sector are still in their early development phase and most of the related Big Data-driven applications are not mature yet. This brings new opportunities for this emerging and promising research area.

One of the primary triggers of this increasing interest is the availability of significant amounts of data from smart meters and the digitization of the distribution grid. Although used initially only for billing purposes, smart meters provide information about the grid end-point operation. If this information is combined with other systems related to the digitization of the distribution grid or other external data sources, it provides even more insight into how the system operates. This is a kind of information that utilities did not have before smart meters deployment and it has opened up opportunities for increasing operational efficiencies and enhancing the distribution grid reliability [324].

Although these opportunities encourage the development of Big Data solutions, utilities are still largely missing the opportunity to utilize those newly available data. As an example, American Council for an Energy-Efficient Economy surveyed 52 large American utilities in 2018 and found that most of them are significantly under-utilizing data from smart meters [325].

Some challenges must be correctly addressed to boost the distribution grid development and reach its full potential. According to [23] and [326], Big Data challenges in energy management focus on six main issues, which are IT infrastructure, data collection and governance, data integration and sharing, data processing and analysis, security and privacy and need for professionals of Big Data analytic and smart energy management. These challenges fit with the more generic main issues that Europe must tackle in creating and sustaining a robust Big Data ecosystem, identified by Big Data Value Association [327]. These issues are related to data, skills, legal, technology, application, business and societal aspects. Six main challenges regarding the distribution grid digitization are addressed below.

- **ICT infrastructure and technology**. Utilities have been forced to strengthen their ICT infrastructure in their back-end systems to deal with Big Data collection and storage. It may include new sensors, improved transmission and storage capacity and increased data processing or data exchange capability [23]. New applications can be developed using existing data, but even more would be available if larger energy-related data were accessible and as close to real-time as possible. An example of this issue is the information currently available from installed smart meters. Valuable knowledge can be discovered from the massive electricity consumption data collected near real-time by AMI devices. However, limited real-time data are available from part of first-generation deployed smart meters. As an example, by 2020, 14 European Union (EU) Member States have implemented a refreshment rate of at least 15 min, while only 8 Member States confirmed to be able to provide near real-time information on electricity every 10 seconds [328]. A second generation of smart meters with near real-time available data is needed, making AMI data actionable for more operation-related tools and long-term planning applications.
- **Data collection and governance**. The availability and access to high-quality data sets are key challenges for enabling AI techniques. In the energy sector, available data are not always sufficient or of good enough quality to develop systems that can handle complex scenarios [24]. In addition, the timeliness, integrity, accuracy and consistency of data for energy AI applications need to be improved [23]. As digital technologies evolve,

these problems can be more efficiently addressed with proper data management and governance strategies.

- **Data integration and sharing**. As well as having agreed approaches, the interoperability of data sets and data-driven solutions are essential to ensure wide adoption within and across applications. However, companies are reticent to share their data to avoid security risks and unlock competitive advantages. For example, great opportunities can arise if operational data from distribution and transmission grids are exchanged between DSOs and TSO fairly and transparently. For this reason, TSOs and DSOs need to determine what information they require, the quality of the information, who owns it and how to ensure both confidentiality and transparency [329]. On the other hand, open data sets are needed to develop and test new algorithms and solutions. Several initiatives worldwide support energy data sharing among stakeholders, such as Green Button and OpenEI in the USA or ENTSO-E Transparency Platform in the EU. However, open energy-related data should increase, as opening up public-sector data and establishing common data standards can also help to boost innovation [330].
- **Data processing, analysis and business models**. New data analysis techniques in DM, ML, statistical analysis, data management and data visualization are applied to the energy sector. Continuous and recently more frequent developments have led to advanced technologies that make significantly easier the use of Big Data, not only in energy applications. These innovative data analysis techniques open up new opportunities to provide solutions and create new businesses in this sector. Thus, it is crucial to identify new business opportunities with existing data and create new data-driven business models to make the most of these techniques and innovations. This study [331] reviews examples of these new business models by analyzing 40 data-driven start-ups in the energy sector.
- **Security, privacy and legal issues.** The power system digitization has converted cybersecurity into an essential concern due to the increasing number of incidents in recent times. Besides, pri-

vacy and security should be guaranteed along the Big Data value chain to protect the customer and the risks and possible impact on supply security. To deal with this problem, several initiatives have been carried out all around the globe. At the EU, one of the critical pieces of legislation in this regard is the Directive on the Security of Network and Information Systems [332], which boosts the level of cybersecurity in the Union through the development of national cybersecurity capabilities, the increase of EU level co-operation and the introduction of security and incident reporting obligations in critical sectors, like the energy. In addition, General Data Protection Regulation [333] aims at protecting individuals concerning the processing of their data and warranting the free movement of such data within the EU. Although these European directives could be considered a late response to an already well-known problem, these regulations can only be considered as a single part of an international chessboard where they should be followed, complemented and particularized by many others [334], as, at the same time, China and USA have introduced their cybersecurity laws and policies.

- **Professionals and skills.** There is a need for trained and educated employees in the energy sector that can use Big Data technologies and build on data expertise. This can be achieved by enrolling experts from other more mature sectors, like finance or marketing, or providing master-level students with specific energy-related Big Data and AI techniques solid background. Although these specialists should be combined with other energy domain knowledge experts, the first option can bring immediate results. Otherwise, energy sector stakeholders can also consider investing in re-skilling and training their employees to manage and operate digitally-enabled power assets and systems effectively [24].

2.7 Conclusions

Implementing Big Data solutions and AI techniques in the power system domain is a promising approach for extracting knowledge and high added value from the vast amount of high-granularity data stored by intelligent devices placed along the distribution grid, such as smart meters or PMUs.

This chapter proposes and interrelates a set of innovative energy services designed to be offered to different electricity domain agents, such as DSO, BRPs and prosumers. These services are fed with high granularity and massive stored data. Thanks to the application of data-driven techniques, they provide solutions to problems like congestion management, distribution grid equipment maintenance, forecasting, detection and prevention of faults and fraud detection.

Once the innovative services have been identified, an exhaustive review of the most recent studies implementing AI techniques in each of them is carried out. Key findings state that ensemble models present better results than single ML models by combining different data-driven algorithms. Deep learning algorithms have gained importance in recent years for time series prediction tasks and outperform most benchmark ML and statistical algorithms. Concerning classification tasks, traditional ML algorithms such as SVM or RF yet provide excellent results. For instance, the RF classifier outperforms when it comes to supervised classification tasks, while LSTM recurrent network is the predominant algorithm for time series forecasting. Unsupervised learning methods are mainly responsible for customer segmentation, building efficiency clustering and consumption profile grouping for non-technical losses detection. Finally, RL is widely applied in the literature to optimally schedule flexible assets in households, although the scarcity of physical experimentation in a realistic environment prevents its application in real-world buildings and households.

To conclude, it is essential to equip the distribution network with sensors to collect the data that feed the innovative services. Implementing data-driven techniques in energy services development is essential for developing a reliable, secure and efficient Smart Grid. Thanks to these methods, the services show better performance compared to statistical benchmark procedures. Nevertheless, there are still challenges to overcome to extend and improve the AI applications in power systems, mainly related to ICT infrastructure, data collection and governance, data integration and sharing, data processing and analysis, security and privacy and the need for professionals of Big Data analytics.

Chapter 3

Environment Strategies for Home Energy Management Systems

This chapter presents and studies different HEMS optimization strategies, ranging from minimizing costs to reducing emissions associated with consumption. The most innovative strategy, a novel multi-objective hybrid HEMS, represents an intermediate point between the two aforementioned approaches. This strategy is designed to minimize electricity costs and greenhouse gas emissions resulting from end-user consumption. To assess the impact of each technology generation on the climate, a life cycle analysis methodology is employed.

3.1 Introduction

Global carbon dioxide emissions reached an all-time high in 2019, despite the fading use of coal [335]. Decarbonization is vital; for this reason, the electricity sector has already started moving from fossil-based to net-zero greenhouse gas (GHG) emissions. This transition is possible thanks to the increasing number of renewable energy sources (RES) and the use of flexibility in the power system to enhance grid integration and maximize the potential of renewable energies. A long-term forecasting study about the evolution of the global energy transition has been conducted in [336]. Key findings predict a significant reduction in fossil-fuel use (around 75% by 2050) and warn that the Paris Agreement will not be accomplished if no further decarbonization measures are taken. Focusing on Europe, current policies will reduce GHG emissions by 60% in 2050 compared to 1990 emissions levels. Nevertheless, the European Commission has increased its climate ambition through the European Green Deal by enumerating key transformative economic policies and measures to put Europe on track to achieve the goal of net-zero global warming emissions by 2050.

Regarding the residential sector, buildings and households play a crucial role in the energy transition. In 2018, they accounted for 28% of the global energy-related carbon dioxide emissions [335]. As there is still a significant margin for improvement in the energy efficiency field, buildings are expected to be the fastest sector in reducing the CO_2 emissions [337]. Therefore, more robust strategic measures to decrease GHG emissions associated with residential electricity demand need to be implemented.

3.2 Related work

This thesis states that intelligent home energy management systems (HEMS) can contribute to achieving environmental targets. In the literature, there are primarily two HEMS approaches. Price-based (PB) -the focus of most current work- and incentive-based (IB). The PB program aims to minimize the end-user electricity bill by optimally rescheduling controllable flexible sources, considering a time-varying pricing tariff. Several studies have used multi-objective HEMS functions for optimal scheduling, considering the minimization of electricity cost and end-user discomfort [338–346]. For instance, [342] minimizes electricity cost and the power profile deviation at the point of standard coupling, while [343] proposes a cost-effective HEMS considering thermal and electricity comfort. The IB program offers flexibility to a third electricity agent to exchange economic compensation for changing its baseline consumption. [347] assures minimum energy cost and supports the upstream micro-grid operation by minimizing the load profile deviation. Electric vehicles and electric water heaters provide flexibility in [348] for PB and IB programs. A third category, environmental-based (EB), has been proposed in [349], which focuses on minimizing the GHG emissions produced by the generation units that provide electricity to the household. The study in [350] presents a multi-objective dispatching optimization model of an energy system focused on energy production, conversion, and storage to maximize the operating revenue and minimize operational risk and carbon emissions. Focusing on HEMS sustainability factors, the study [351] calculates climate effects by displaying carbon emissions at customers' premises to motivate them to diminish their consumption. In [352], the curtailment of on-site PV is penalized for maximizing green energy consumption.

Therefore, this chapter presents a hybrid approach to Home Energy Management Systems, referred to as hybrid-based (HB) HEMS, combining the previously mentioned price-based (PB) and environmental-based (EB) methods. Table 3.1 lists various HEMS programs from the literature. The vast majority focus on PB programs, as stated in [348]. Some studies propose multi-objective functions incorporating PB and incentive-based (IB) programs, such as those presented in [347,348,350].

Table 3.1: HEMS programs and their services.

HEMS program	**Description**	**HEMS service**
PB [338–341, 347, 348, 352–354]	The objective is to minimize the end-users electricity bill.	Time-of-use pricing, real-time pricing and peak shaving.
IB [347, 348, 355, 356]	Flexible sources are economically incentivized to be flexible by modifying their electricity use.	Providing flexibility to a third energy agent.
EB [349]	Flexibility is used to minimize the GHG emissions of buildings associated with generators that produce the electricity they consume.	Minimization of GHG.

3.3 LCA for electricity generation systems: a time-varying GWP approach

This section provides the actual environmental impact considering the entire life cycle of an electricity generation system using the Life Cycle Assessment (LCA) methodology. This process is described in Figure 3.1, following the steps noted in [8]. For each electricity generation technology, GHG emissions are evaluated and categorized according to contributions from the following three life cycle phases:

- Fuel provision: activities from fuel extraction to its delivery at the plant gate.
- Plant operation: operation and maintenance of the plant and the appropriate management of residues.
- Infrastructure: covers the commissioning and decommissioning related emissions of the electricity generation system.

The LCA impact category selected for this thesis as the reference measure to quantify and assess the potential environmental impact of an electricity generation source is the Global Warming Potential (GWP) indicator. To calculate the overall GWP for each generation source, the first step is identifying and quantifying the GHG emissions associated with each life cycle stage listed and shown in Figure 3.1. The most common GHGs are carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (NO_x), as they have significant contributions to global warming. Once the emissions are quantified, each GHG is multiplied by its respective GWP factor. These factors represent the relative warming potential of each gas compared to CO_2. For instance, methane is estimated to have a GWP of 27-30 over a 100-year period. CH_4 emitted today lasts about a decade on average, much less than CO_2, however, CH_4 absorbs much more energy than CO_2. After multiplying the emissions by the GWP factors in all the life cycle phases, the results are summed up to obtain the total GWP for the electricity generation source. The unit typically used for GWP is carbon dioxide equivalent (CO_{2-eq}), which represents the amount of CO_2 emissions that would have the same warming effect as the combined emissions of all GHGs. It's important to note that the specific methodology and data sources used to calculate the GWP may vary depending on the LCA study and regional concerns.

Table 3.2 summarizes and lists the GWP indicators for the electricity generation sources evaluated in this investigation, taken from [8] since the scope and objective of this chapter do not involve calculating the GWP of the generation sources taken into account in this analysis.

Power systems consisting of diverse generation sources have time-dependent GHG emissions. Consequently, the GWP performance changes

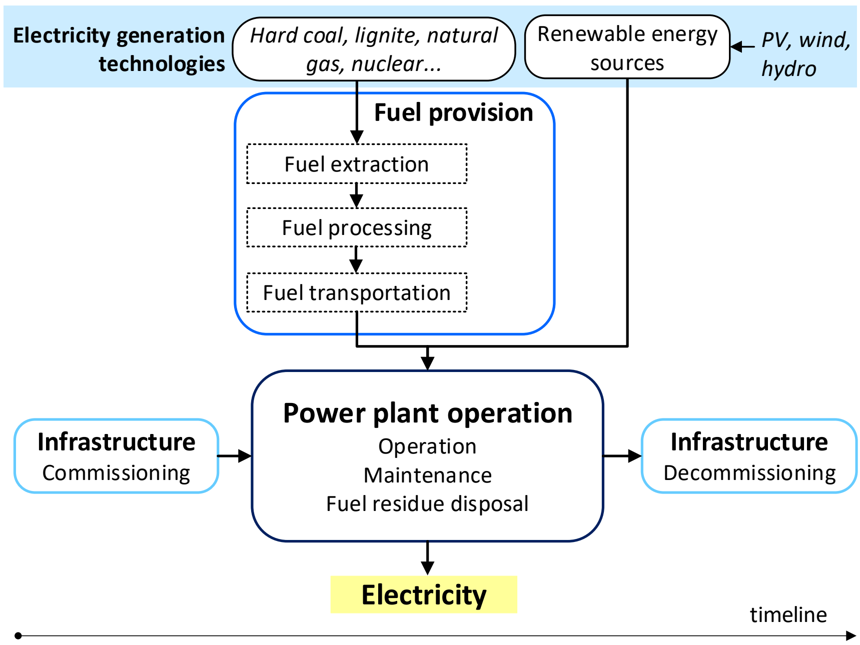

Figure 3.1: Electricity generation technologies LCA steps.

hourly along with the electricity mix of each country [357]. A primary objective is to determine the time-varying amount of kg CO_2 equivalent (CO_{2-eq}) emitted per kilowatt-hour (kWh) of the energy mix. The average hourly GWP impact of the electricity supply, denoted as $E_t^{gwp,grid}$, can be expressed as follows.

$$E_t^{gwp,grid} = \sum_{i \in I} E_i^{gwp,avg} \cdot GS_{t,i} \tag{3.1}$$

where $E_i^{gwp,avg}$ is the GWP average constant for each type of generation source i and $GS_{t,i}$ refers to the estimated generation in the day-ahead market at period t for each type of generation source i.

It should be mentioned that these GWP emission values are not static, as they are expected to vary over time, given that processes generally tend to become more efficient and improve performance in terms of associated emissions.

Table 3.2 shows the GWP indicator range for the overall life cycle stages for each electricity generation source type, according to [8]. This thesis uses the average GWP value.

Table 3.2: Lyfe cycle emission factors for electricity generation sources [8].

Generation source GS_i	**GWP range [kg** $\mathbf{CO_{2-eq}}$**/kWh]**	**Average GWP [kg** $\mathbf{CO_{2-eq}}$**/kWh]**
Hard coal	0.66-1.05	0.855
Lignite	0.8-1.3	1.050
Natural gas	0.38-1	0.69
Nuclear	0.003-0.035	0.019
Biomass	0.0085-0.13	0.0693
Hydro-power	0.002-0.02	0.011
Photo-voltaic	0.013-0.19	0.1015
Wind	0.003-0.041	0.022
Battery	-	0.0706

3.4 Mathematical formulation

This section covers the three HEMS objective functions approaches -PB, EB and HB- and the mathematical formulation of the optimization model for controlling and re-scheduling flexible household sources.

3.4.1 HEMS objective function

3.4.1.1 Price-based program

This program focuses exclusively on the economic aspect. It aims to minimize the electricity bill (4.1), considering the battery degradation cost K_t^{cal} due to calendar aging, where P_t^{buy} is the time-varying electricity price, χ_t^{buy} refers to the energy purchased to the grid, and P^{VAT} is the tax applied. Constraint (3.2c) ensures that the energy balance is always met, where ψ_t^{pv} is the optimized PV generation output and W_t^{load} stands for inflexible household consumption. Finally, constraint (3.2d) avoids exceeding the contracted power $X^{max,imp}$. To switch from power to energy units, N^{hour} is used, which refers to the number of periods per hour. The objective function is expressed as

$$\min_{\chi,V} \quad f_1 = \sum_{t\in T}(P_t^{buy}\chi_t^{buy}P^{VAT} + K_t^{cal}) \tag{3.2a}$$

s.t.

$$K_t^{cal} = 0.019 \cdot V_t - 0.0629, \tag{3.2b}$$

$$\psi_t^{pv} + \sigma_t^{dis} + \chi_t^{buy} = \sigma_t^{ch} + W_t^{load}, \tag{3.2c}$$

$$\chi_t^{buy} \leq X^{max,imp}/N^{hour} \tag{3.2d}$$

3.4.1.2 Environmental-based program

this program attempts to minimize the carbon footprint occasioned by the generation sources that provide electricity to the household. This HEMS is presented in [349]. The objective function is formulated as follows.

$$\min_{\chi,\psi,\sigma} \quad f_2 = \sum_{t \in T} E_t^{gwp,grid} \chi_t^{buy} + E^{gwp,pv} \psi_t^{pv} + \tag{3.3a}$$

$$+ E^{gwp,bat} \sigma_t^{dis} \tag{3.3b}$$

s.t.

$$\psi_t^{pv} + \sigma_t^{dis} + \chi_t^{buy} = \sigma_t^{ch} + W_t^{load}, \tag{3.3c}$$

$$\chi_t^{buy} \leq X^{max,imp} / N^{hour} \tag{3.3d}$$

where $E_t^{gwp,grid}$ indicates the kg CO_{2-eq}/kWh of the grid on average per period t. It is calculated with the hourly energy production mix, taking the values of scheduled generation in the day-ahead market for each technology described in Table 3.2.

3.4.1.3 Hybrid-based program

this multi-objective problem (MOP) approach combines the PB and the EB objective functions. It is a multiple-criteria decision-making problem with no unique optimal solution but a domain of feasible solutions that satisfy all constraints. Therefore, the result is a trade-off, a compromise between minimizing the energy costs and decreasing GHG emissions derived from the generation sources that provide electricity to the house.

The HB objective function f_3 is formulated in (3.4a). Normalization of the objectives is required so that both competing objectives can be equivalent and compared at the same level. f_1^* is the optimal solution of the PB objective function f_1 (4.1), and f_2^* is the optimal solution of the EB objective function f_2 (4.2). The linear MOP is formulated as

$$\min_{\chi,\psi,\sigma,V} \quad f_3 = \frac{\alpha \cdot f_1}{f_1^*} + \frac{(1-\alpha) \cdot f_2}{f_2^*} \tag{3.4a}$$

s.t.

$$K_t^{cal} = 0.019 \cdot V_t - 0.0629, \tag{3.4b}$$

$$\psi_t^{pv} + \sigma_t^{dis} + \chi_t^{buy} = \sigma_t^{ch} + W_t^{load}, \tag{3.4c}$$

$$\chi_t^{buy} \leq X^{max,imp}/N^{hour}, \tag{3.4d}$$

$$\alpha \leq 1 \tag{3.4e}$$

where α is the weighting factor for PB objective function, and (1-α) for the environment based. The value of α must be lower or equal to one (3.4e). To conclude, Table 3.3 shows the formulation of the objective functions of the three HEMS programs.

Table 3.3: HEMS programs' objective functions.

Objective function	Mathematical formulation
Price-based	$[MIN] f_1 = \sum_{t \in T} (P_t^{buy} \chi_t^{buy} P^{VAT} + K_t^{cal})$
Environment-based	$[MIN] f_2 = \sum_{t \in T} (E_t^{gwp,grid} \chi_t^{buy} + E^{gwp,pv} \psi_t^{pv} + E^{gwp,bat} \sigma_t^{dis})$
Hybrid-based	$[MIN] f_3 = \frac{\alpha f_1}{f_1^*} + \frac{(1-\alpha)_2}{f_2^*}$

3.4.2 Energy storage system

Battery aging is formed by calendar and cycling aging. Calendar aging happens during the battery rest time, whereas cycling aging is caused directly by charges and discharges. According to [358], the Li-ion battery degradation due to cycling shows minimal aging for low current rates [359] and also when the battery is not charged to its maximum state of charge (SOC) since there is a faster degradation when charging to 100% SOC. Therefore, the following constraints explained in Section 3.4.2.2 are added to the battery model to ensure that the storage unit works under conditions that minimize the cycle aging impact:

- Equation (3.13) ensures that the battery charges and discharges at low current rates.

- Equation (3.14) reduces large cycles at high SOC by limiting the maximum SOC allowed.

Given the above, the battery calendar aging is formulated.

3.4.2.1 Battery calendar aging

Battery operating conditions have a significant impact on their performance and lifetime. The storage model applied in this thesis considers calendar aging for a lithium-ion (Li-Ion) battery. This phenomenon leads to a decrease in usable battery capacity and an increase in the battery's inner resistance over time, resulting in a depreciation cost. The calendar aging model formulation applied in this study is parameterized in [359] through accelerated aging tests. The capacity defined in (3.5) is a phenomenon where the volume of energy that a battery can operate at the rated voltage diminishes over time [360]. Cell temperature and voltage are the variables that impact calendar aging, thus influencing battery lifetime. The loss of capacity is more prominent than the resistance increase in the calendar aging function, according to [359], as the end of the battery life is reached first due to the loss of capacity. For this reason, only the capacity is considered in the calendar aging formulation.

For a Li-Ion battery cell, the capacity C due to calendar aging is expressed as

$$C(t) = 1 - \psi(V,T) \cdot t^{0.75} \tag{3.5}$$

where ψ is an aging factor that describes the aging rate during period t and is formulated as

$$\psi(V,T) = (a \cdot V_t^{cell} - b) \cdot e^{-c/T} \tag{3.6}$$

where temperature is a constant parameter $T = 293\,K$ in this study, $a = 7.543 \cdot 10^6\; V^{-1} days^{-0.75}$, $b = 2.375 \cdot 10^7\; days^{-0.75}$ and $c = 6976\,K$ [359].

The relationship between the open-circuit voltage (OCV) and SOC is known and expressed by the non-linear equation shown in Figure 3.2.

To avoid non-linear constraints, the dependence between the OCV of the cell and SOC is linearized. The minimum SOC by restriction is limited to 25%.

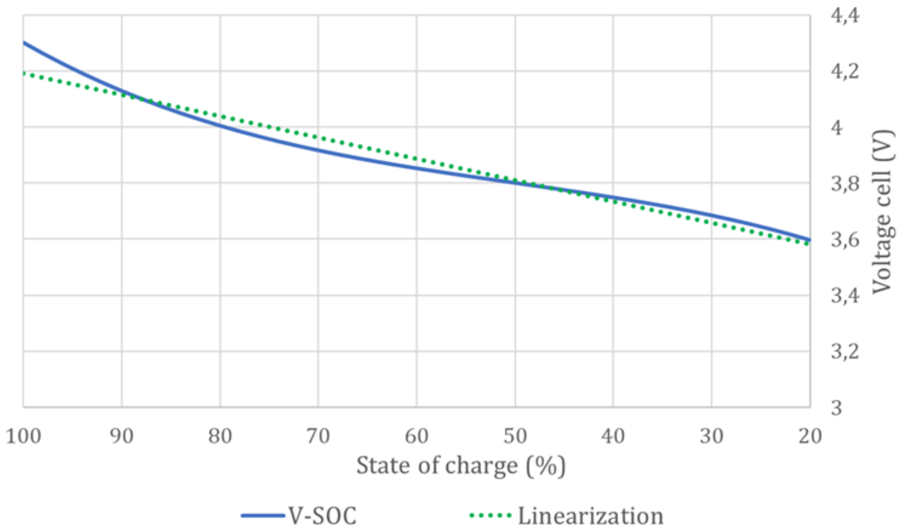

Figure 3.2: Linear and non-linear OCV and SOC dependence.

As a result, the linear correlation is represented as

$$V_t^{cell} = 0.0076 \cdot \sigma_t^{soc} + 3.4287 \tag{3.7}$$

where variable σ_t^{soc} indicates the percentage of energy stored per period t. The depreciation of the battery during each time step Δt leads to K_t^{cal} costs

$$K_t^{cal}(L, \Delta t) = \frac{K^{invest}}{L} \Delta t \tag{3.8}$$

where K^{invest} is the acquisition cost of the 8 kWh Li-Ion battery, and it is set in 7500 €, L is the lifetime of the battery and Δt is the time step. The end-of-life criterion is defined to be 80% of initial capacity C [360]. Therefore, the expected battery life can be calculated as $C = 0.8 = 1 - \psi L^{0.75}$, so equation (3.9) remains

$$K_t^{cal}(V,T,\Delta t) = \frac{K^{invest}}{\frac{0.2}{((a \cdot V_t^{cell} - b) \cdot e^{-c/T})^{1/0.75}}} \Delta t \tag{3.9}$$

A linear approximation to equation (3.9) is calculated to relax the constraint and implement a linear solving method. Figure 3.3 shows non-linear and linear equations, proving that the functions' behavior is practically identical.

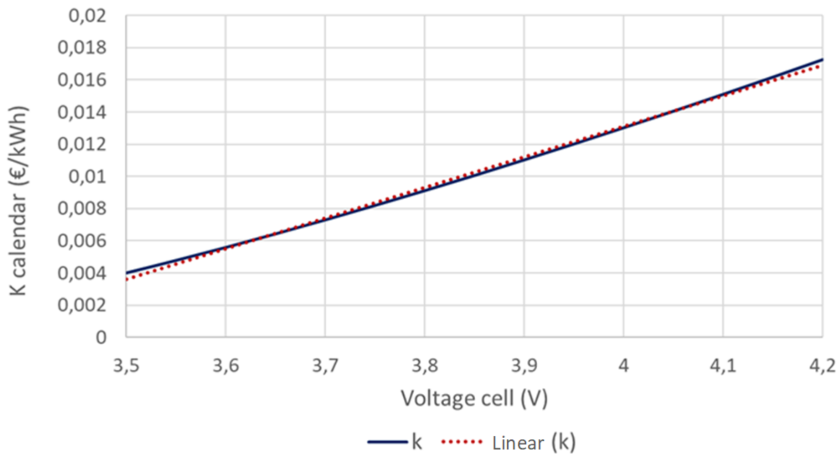

Figure 3.3: Dependence of cell voltage on SOC for a cell temperature of 293 K and time step $\Delta t = 15$ minutes.

Therefore, the linearized $K_t^{cal,linear}$ per time step t is formulated as

$$K_t^{cal,linear}(V) = 0.019 \cdot V_t^{cell} - 0.0629 \tag{3.10}$$

3.4.2.2 Battery constraints

The battery model constraints are formulated. The variable σ_t^{soc} in equation (3.11) represents the battery SOC for each period. The efficiency factors for storing η^{ch} and delivering electricity η^{dis} are considered to represent the actual behavior of the battery. The variables

σ_t^{ch} and σ_t^{dis} represent the amount of energy charged or discharged in each period.

$$\sigma_t^{soc} = \sigma_{t-1}^{soc} + \sigma_t^{ch} \cdot \eta^{ch} - \frac{\sigma_t^{dis}}{\eta^{dis}} \tag{3.11}$$

To avoid over-optimistic results, the battery SOC must be the same at the beginning and the end of the optimization horizon.

$$\sigma_{t=0}^{soc} = \sigma_{t=final}^{soc} \tag{3.12}$$

As mentioned before, it is essential to ensure the battery is not fully charged or discharged by limiting its maximum and minimum allowed SOC to a specific fixed value to avoid cycle aging. The equation (3.13) ensures that σ_t^{soc} is always between a minimum O^{min} and a maximum O^{max} to preserve and extend the battery lifetime:

$$O^{min} \leq \sigma_t^{soc} \leq O^{max} \tag{3.13}$$

Equations in (3.14) also help to minimize cycling aging by limiting the maximum power allowed for charging Q^{ch} and discharging Q^{dis}.

$$\sigma_t^{ch} \leq \frac{Q^{ch}}{N^{hour}}, \qquad \sigma_t^{dis} \leq \frac{Q^{dis}}{N^{hour}} \tag{3.14}$$

The following constraint makes sure that the energy charged $\sigma_{t,b}^{ch}$ to the battery unit b is linearly decreased from S_b^{ch} state of charge. This linear function typically goes from 80% SOC to 0 at 100% SOC. This constraint is represented in Figure 3.4.

$$\sigma_{t,b}^{ch} \leq \frac{-Q_b^{ch}}{1 - S_b^{ch}} \cdot \left(\frac{\sigma_{t,b}^{soc}}{O_b^{max}} - 1\right) \qquad \forall b \in B, t \in T \tag{3.15}$$

The same happens for discharging energy $\sigma_{t,b}^{dis}$ of battery b during period t. The lower threshold to limit the energy output is S_b^{dis}, typically from 10% SOC to 0 at 0% SOC. This constraint is shown in Figure 3.5.

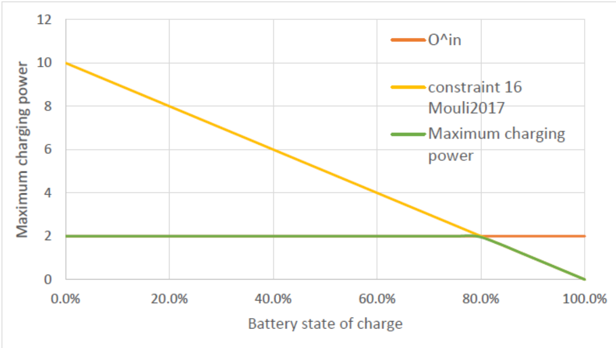

Figure 3.4: Battery SOC as a function of maximum charging power [6]. Source: INVADE H2020 Project.

$$\sigma_{t,b}^{dis} \leq \frac{Q_b^{dis}}{S_b^{dis}} \cdot \frac{\sigma_{t,b}^{soc}}{O_b^{max}} \qquad \forall b \in B, t \in T \tag{3.16}$$

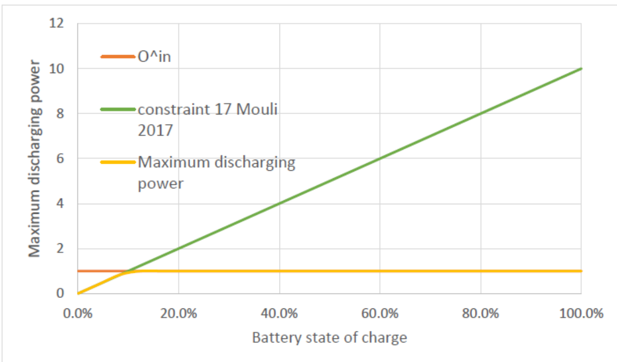

Figure 3.5: Battery SOC as a function of maximum discharging power [6]. Source: INVADE H2020 Project.

3.4.3 PV generation constraints

The formulation of a reducible PV generation model is presented. The optimization variable PV scheduled generation ψ_t^{pv} must be between 0 and the PV baseline electricity generation W_t^{pv}, which is the forecasted PV generation curve for the following day.

$$0 \leq \psi_t^{pv} \leq W_t^{pv} \tag{3.17}$$

3.5 Case study

The case studies presented in this section aim to analyze the three HEMS program's performance -PB, EB, and HB- to compare the electricity expenses and kg of CO_{2-eq} related to a single-family household. Actual consumption and PV generation data are taken from the Data Port database [361] and used as input to the HEMS programs. These case studies are located in Spain; therefore, the Spanish dynamic electricity tariff (Precio Voluntario Pequeño Consumidor tariff) and its electricity mix are used as input data. P^{VAT} is set to 21%. The optimization horizon is 24 hours, divided into 96 time periods of 15 minutes, starting at 00:00h. The household contracted maximum power is 6 kW and is equipped with a 4.8 kW PV and a 9 kWh battery, whose SOC must be at least 50% at the beginning and end of the optimization horizon. The value of the parameters applied for all the case studies are listed in Table 3.4. The end-user does not sell back electricity to the grid; therefore, the PV is exclusively for self-consumption, and the excess of production can be stored in the battery for later usage. The HB multi-objective function weights are set to $\alpha = 0.3$ and $\beta = 0.7$, according to the end-user preferences that emphasize environmental aspects. The HEMS has been implemented in Python, using the Pyomo optimization library and the Gurobi solver. The optimal solution of the HB HEMS program was obtained with a computational time of 0.44 seconds on a Laptop with a processor core i7 at 2,60 GHz and 8 GB of RAM.

Two opposite Spanish energy mix generation scenarios are proposed to examine the HEMS program's performance. On the one hand, low penetration of renewables in the electricity mix and, on the other, high participation of sustainable generation sources. Figure 3.6 presents a scheme of these case studies. The PB HEMS is run separately since its performance only depends on the dynamic pricing tariff, regardless of

Table 3.4: HEMS optimization strategies: the case studies parameters.

Parameters	Value	Units
Battery maximum allowed SOC	8	kWh
Battery minimum allowed SOC	2	kWh
Battery SOC initial/final	4.5	kWh
Battery maximum power charge/discharge	3	kW
Battery efficiency charge/discharge	0.95	-
Household maximum import capacity	6	kW
PV maximum output power	4.8	kW

the energy mix composition, since its objective is to minimize the cost, not GHG emissions. Case studies have identical inflexible demand, PV generation, battery parameters, and hourly electricity prices to compare the results of the three HEMS programs. The only input parameter that changes is the electricity generation mix of the grid.

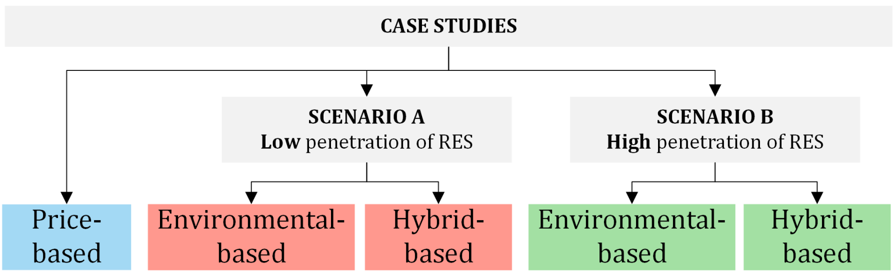

Figure 3.6: Scheme of the case studies proposed to test the proposed HEMS optimization strategies performance.

For Scenario A with low RES penetration, data from November 20th 2017 is used, in which the percentage of non-fossil generation penetration is 29.18%. For Scenario B with high RES penetration, March 6th, 2020 has been selected, with a daily average of 83.07% of electricity generation sources with zero emissions during their electrical grid operation. It should be noted that nuclear power is incorporated within zero-emissions energy sources. The hourly share of each generation source -listed in Table 3.2- in the Spanish energy mix for both scenarios is illustrated in Figure 3.7. The total generation curve is also represented to demonstrate that the selected generation types are primarily responsible for the overall generation and cover 96% of the total demand. The higher the RES penetration in the energy mix is, the lower the GWP grid value per energy unit. Combined cycle and coal generation dominate in Figure 3.7(a), whereas wind power takes priority in Figure 3.7(b).

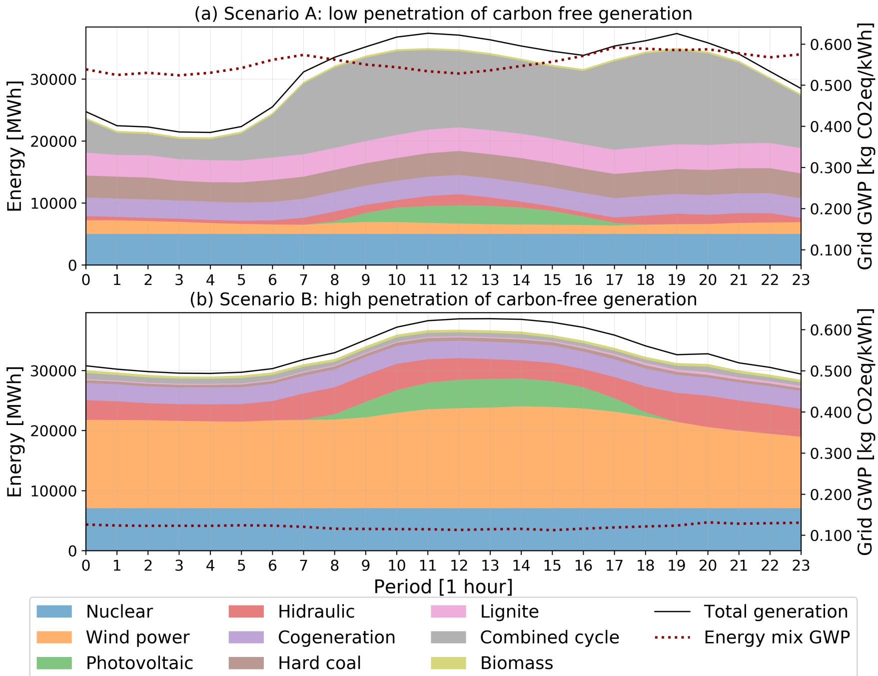

Figure 3.7: Generation sources grid penetration share for both scenarios.

3.6 Results

In this section, the case studies' results are presented and discussed. For a better understanding of the graphical outcomes obtained (see Figure 3.8, for instance), it should be noted that negative energy values represent a generation source like PV and battery discharging. In contrast, positive values represent energy consumption, such as inflexible loads or battery charging. Hence, following the energy balance equation (3.2c), in which generation must match consumption, the resulting plot exhibits symmetry when the total generation and consumption sources are aggregated individually.

3.6.1 Price-based scenario

The PB case study analyzes the HEMS behavior under dynamic price tariffs to minimize the end-user electricity cost. The PB objective function result is identical for high and low RES penetration scenarios since it only depends on the electricity tariff variation. However, the GHG emissions indirectly caused by PB optimization vary depending on the scenario. For clarification, the PB program is executed if the parameter α is set to 1 in the HB MOP, as equation (3.4a) indicates.

The PB HEMS optimization results are displayed in Figure 3.8. The upper graph shows the electricity price, while the lower displays the PB optimization result. Focusing in Figure 3.8(b), during periods of low prices (4-27), the consumed electricity is purchased directly from the grid, taking advantage as well to charge the battery (19-23, 91-94) to discharge it later during time intervals with more expensive costs (28-33, 66-81). PV allows self-consumption during most daylight periods, and the surplus energy is used to charge the batteries, reaching the maximum capacity of 8 kWh in period 66. The battery charges again in the last low-priced periods (91-94) to meet the restriction of ending at least half of its SOC. The total cost of the objective function is 2.79 €. If the result is broken down, 52% belongs to battery degradation cost, while 48% corresponds to the price of buying electricity from the grid, including taxes.

3.6.2 Scenario A: low penetration of RES in the energy mix

The outcomes obtained in Scenario A for EB and HB HEMS programs are displayed in Figure 3.9. The hourly-varying GWP of the grid, along with electricity prices, are represented in Figure 3.9(a). Figure 3.9(b) displays the EB and Figure 3.9(c) shows the HB results.

The EB program is fed from the grid during periods with moderate kg CO_{2-eq} levels (0-27) compared to the daily GWP average. In high

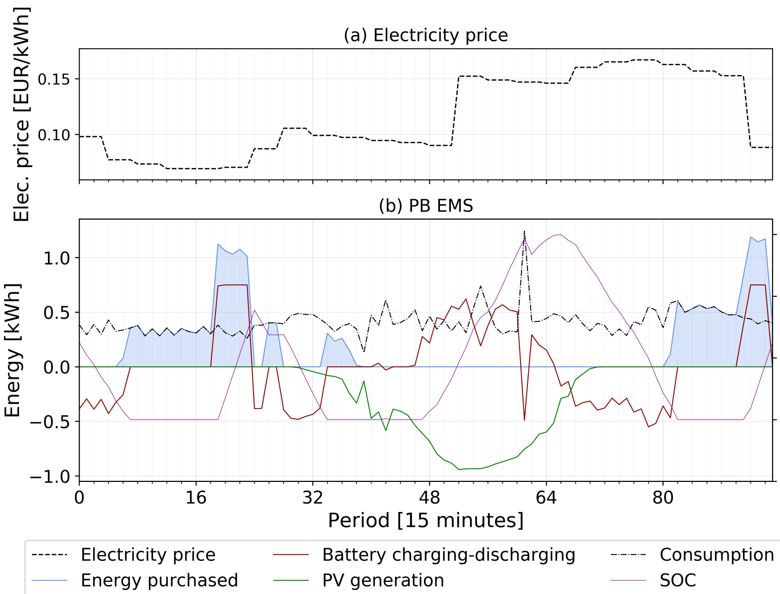

Figure 3.8: Results of price-based HEMS under a Spanish price scheme.

grid GWP indicators periods, the battery is discharged (28-35) to avoid consuming from the network. PV generation is diminished (43, 46, 52, 54) as it is not feasible to store the battery's surplus energy due to its maximum SOC limit. Besides, the sum of the kg of CO_{2-eq} per kWh of PV and the battery charge has a higher environmental impact than purchasing straight from the grid in specific periods. To comply with the battery SOC restriction at the end of the optimization horizon, electricity is bought from the grid to charge the battery when the grid GWP levels are low (see periods 38-60).

Concerning the HB HEMS program in Figure 3.9(c), it is discerned that compared with the EB HEMS, the battery discharges at moderately high prices compared to the subsequent periods (0-3). In (28-30), the battery discharges due to the high GWP values in the power system, although less energy than the EB, as the prices for that period are more high-priced. Solar energy does not reduce its production and is used for self-consumption. Meanwhile, the surplus energy is used for charging the battery. The HB avoids buying during the most expensive intervals of the day (76-80). The energy needed to charge the battery until the SOC imposed (4.5 kWh) at the end of the optimization horizon is purchased from the grid at affordable prices (91-96).

3.6.3 Scenario B: high penetration of RES in the energy mix

The electricity generation of the grid is composed of 83% on average by CO_2 free generation sources, including nuclear energy and RES. Grid GWP values are 4.5 times lower than in the previous Scenario A, while the price signals, inflexible household consumption, and PV remain the same. EB HEMS results are shown in Figure 3.9(b). Due to the high penetration of renewables in the electrical system, the EB purchases energy from the grid practically in its entirety, excluding daylight hours when the household is self-supplied, generating just the electricity required to meet the consumption. The EB program outcomes confirm that if a nation's energy mix is highly renewable as in Scenario B, the usage of batteries is more polluting than purchasing directly from the grid.

The HB HEMS outcomes are represented in Figure 3.9(c). It avoids

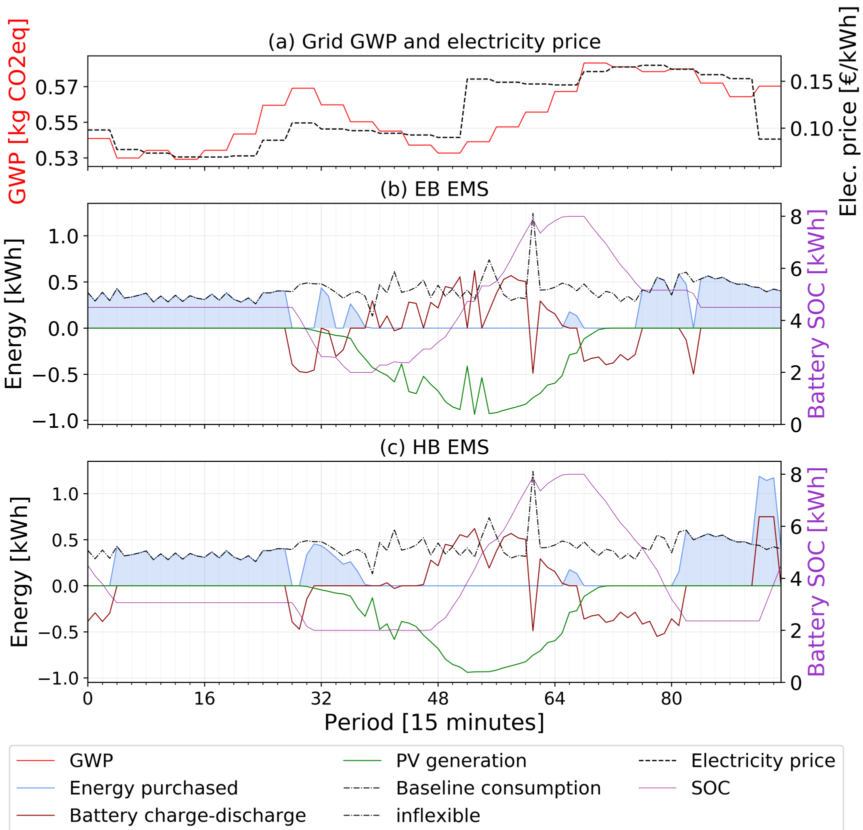

Figure 3.9: Results of an environmental-based and hybrid-based HEMS in Scenario A under a Spanish price scheme.

purchasing from the grid during expensive periods (0-6) and intermittent periods (68-82). It uses the PV surplus (46-65) to charge the battery and discharge it later at high prices (68-72).

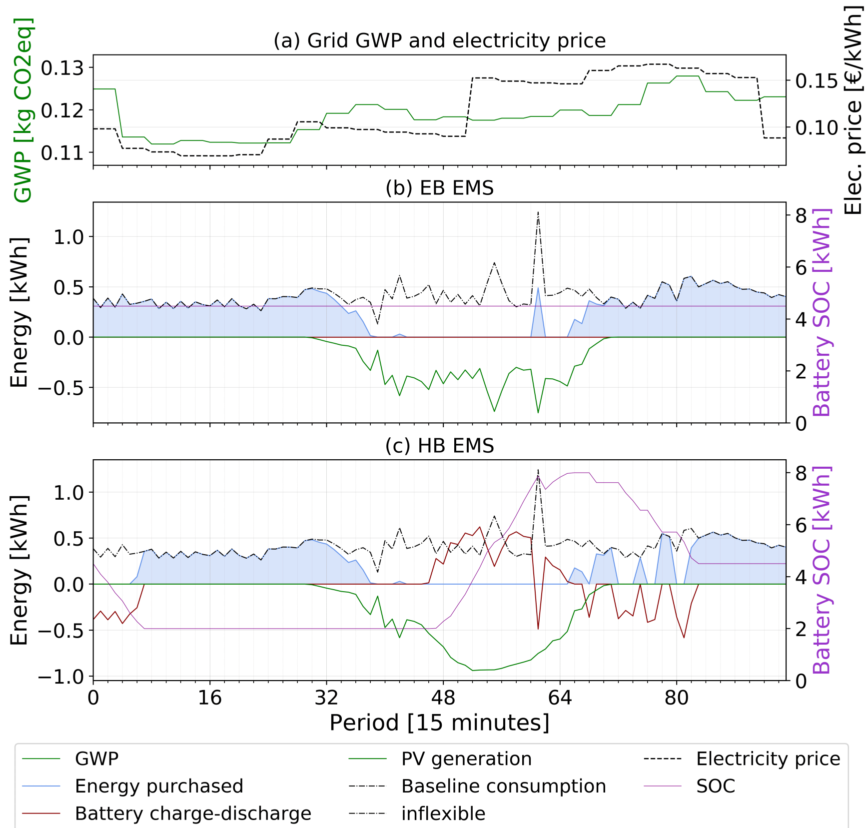

Figure 3.10: Results of an environmental-based and hybrid-based HEMS in Scenario B under a Spanish price scheme.

Comparison among scenarios

The results obtained from scenarios A and B emphasize the crucial role of the energy mix in influencing the environmental impact of energy management systems. With the growing penetration of renewable energy sources in the overall energy system, the environmental benefits of using batteries for home energy storage and photovoltaic generation might diminish. It is possible that, in some instances, direct purchases from the grid are a more favorable choice from a climate perspective.

In Scenario A, characterized by low renewable energy penetration, the EB program emphasizes self-consumption and storage during high GHG emissions peaks to minimize environmental impact. Contrarily, in Scenario B, with high renewable energy penetration, the EB program predominantly relies on purchasing clean and renewable energy from the grid, since using the storage energy has more pollutants associated. In some periods, even the PV generation is curtailed. The HB approach follows a similar basis but considers the cost factor by avoiding energy purchases during high-cost periods, regardless of the emissions during those specific hours.

These findings emphasize the importance of considering the energy mix composition and the effectiveness of HEMS programs in optimizing energy usage and minimizing environmental impact. The results highlight the potential trade-offs and the need for careful evaluation and adaptation of HEMS strategies based on the specific characteristics of the energy system, considering also the end-user preferences.

3.6.4 Sensitivity analysis

The Pareto optimal front for the two HB normalized objective components is represented in Figure 3.11(a), choosing as weighting factors α varying from 0 to 1 in steps of 0.02. The blue and purple points sequences represent the border of the feasible solution region that satisfies all the restrictions imposed for each GWP scenario. Figure 3.11(b) shows the z_3 optimal solutions for each α for high and low GWP scenarios.

A sensitivity analysis is performed in Figure 3.12 to show how the HB objective function z_3 is affected based on changes in the following input variables: maximum allowed battery SOC in (a)-(b), PV generation output in (c)-(d), inflexible household consumption in (e)-(f), and electricity price average in (g)-(h) for low and high GWP scenarios. The legend shows the maximum allowed SOC, the total daily PV generation and consumption, and the average daily electricity price. Thanks

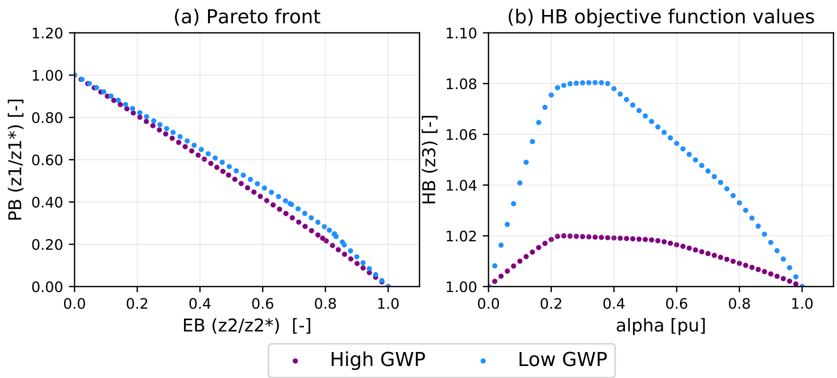

Figure 3.11: (a) Pareto front and (b) HB solution values for high and low GWP scenarios.

to this sensitivity study, it is known how the variation of one parameter affects the HB HEMS optimal solution, reducing uncertainty. It is reminded that $\alpha = 0$ corresponds to the EB HEMS program and $\alpha = 1$ to the PB HEMS program. The continuous grey line refers to the case study optimal solution described in Section 5.5 and displayed in Figure 3.11(b). The start and end of α take the value one due to the normalization of HB objective function z_3. Consumption is the variable that most affect the HB program. The lower the consumption, the lower the energy cost and emissions (see *consum 14 kWh*). On the contrary, the higher the consumption, the greater the cost and environmental impact (see *59 kWh*). Consumption is followed by the electricity price, although when α equals to zero, this variable is insignificant since the EB HEMS does not consider the electricity price in its objective. The PV generation is more sensitive when it produces less (3 kWh) as this implies an increment in the electricity cost ($\alpha = 1$) because more electricity needs to be bought. Still, it has barely any impact for high amounts of generation (see *PV 25 kWh* and *PV 32kWh*) because household consumption is minimal compared with the PV generation. Finally, the input parameter with the most minor influence is the maximum SOC allowed for the battery.

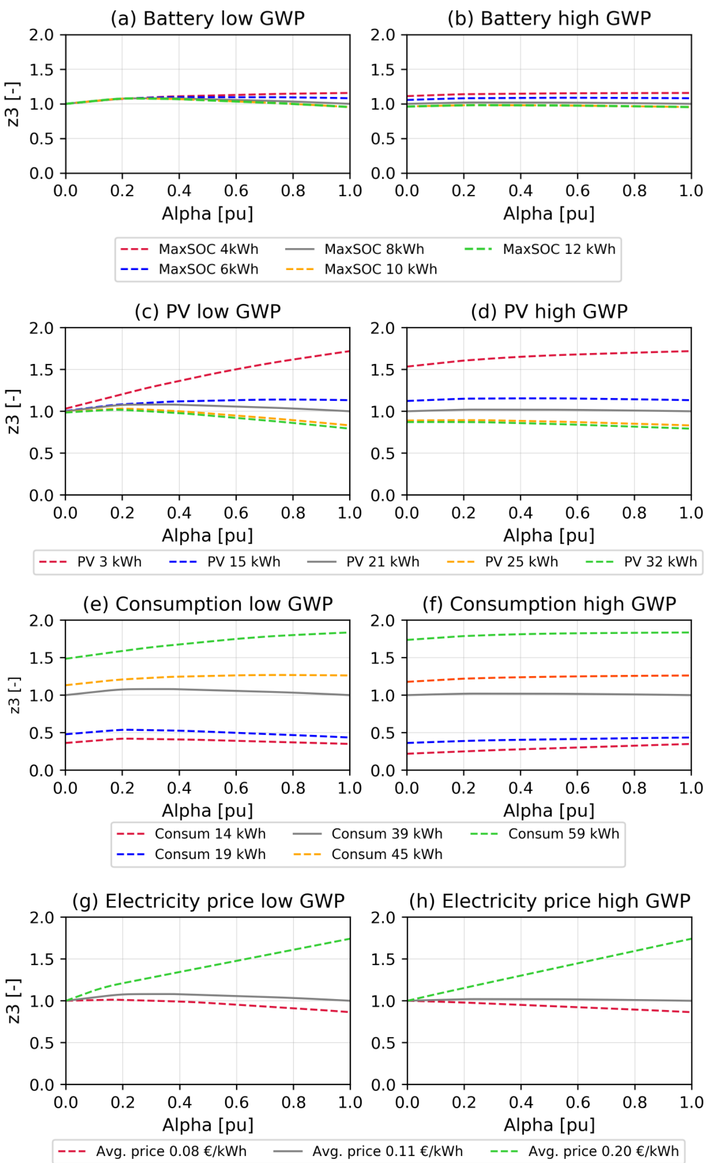

Figure 3.12: Sensibility analysis of different HEMS input parameters.

3.6.5 Results comparison and discussion

A comparative overview of the HEMS program results for both high and low RES penetration scenarios is presented in Table 3.5, showing the total GHG emissions and electricity costs. Moreover, these HEMS programs are compared with the household baseline, which is the energy exchanged with the grid if no flexible resources are activated. In other words, the end-user buys all the power from the grid to meet the inflexible consumption. It is concluded that PB HEMS achieves the lowest electricity costs for both scenarios, as expected, due to avoiding the purchase from the grid during expensive periods. However, in return, the PB HEMS has the most polluting emissions associated with its consumption, specifically releasing 15.36% more emissions than the less polluting program -the EB- in the high RES scenario. On the other hand, the PB produces 5.19% more emissions than the EB in the low RES scenario. On the economic side, the EB program pays 2.81 times more in the electricity bill than the PB in high RES situation and almost twice as much in a low RES scenario.

As can be appreciated in Table 3.5, the HB program has a balance between the PB and EB programs. For the high presence of renewables in the energy mix, the difference between the HEMS emissions does not exceed a range of 15%. In contrast, the boundary is narrowed to approximately 5% in the opposite RES scenario. This is because during periods of high renewable generation in the energy mix, the EB purchases from the grid most of the periods, so the battery is not utilized for self-supply nor is charged by PV surplus, resulting in a higher electricity bill when buying more energy from the grid.

The baseline case only emits 2% more pollutant emissions than the EB program in the high RES scenario. The explanation is the same as in the previous paragraph: energy from the grid is cleaner than discharging batteries previously charged with solar PV surplus. On the contrary, on days with a high percentage of fossil-type generation in the grid, the kg of CO_{2-eq} soared compared to the rest of the programs: 61% more emissions than EB, 58.16% compared to HB, and 53.44% compared to PB. The explanation is that the rest of the HEMS programs use the surplus PV generation to charge batteries for later use, avoiding to a great extent buying from the grid, which is much more polluting than using flexible resources.

Table 3.5: Results of the case study α=0.3. GHG emission and energy cost comparison between the different energy management system programs.

HEMS program	**GHG [kg CO_{2-eq}]**		**Electricity cost [€]**	
	High RES	Low RES	High RES	Low RES
Price-based program	5.33	14.39	1.33	1.33
Hybrid-based program	4.85	13.96	2.74	2.54
Environmental-based program	4.62	13.68	3.74	2.70
Baseline consumption	4.72	22.08	5.73	5.73

3.7 Conclusions

The chapter introduces a novel hybrid-based HEMS formulation that optimizes the operation of PV generators and distributed storage units behind-the-meter in order to achieve the best trade-off between electricity cost and GHG emissions minimization, considering a life cycle analysis of the generation sources used to meet the household demand. Two facing energy mix scenarios are proposed: high renewable energy participation and high fossil-type generation participation. The results confirm the reduction of GHG emissions in the HEMS containing the environmental component. The EB program achieves the lowest emissions, while the HB seeks a compromise between economic and environmental factors. By assigning weights to the HB multi-objective function, the end-user can modify its priority in a fast and flexible manner.

The more renewable generation is in the energy mix, the lesser the difference between the HEMS programs' emissions and the baseline case. However, electricity costs increase the more renewable energy is in the energy mix. Therefore, if the objective is strictly to minimize the environmental impact produced by household consumption, it is concluded that if a nation regularly holds a very high penetration of renewable generation in its energy mix as it occurs in scenario B, it is more sustainable from the household point of view to buy electricity directly from the grid than using self-consumption with batteries, previously charged with PV surplus, for instance. In return for prioritizing and considering the polluting emissions minimization, the HB HEMS electricity expenses can be two times higher than the PB. On the other hand, for countries with low penetration of renewable generation sources in the energy mix, flexible resources such as batteries and PV panels significantly reduce GHG emissions and costs. Therefore, HB HEMS is an excellent option to encourage end-users to participate in the fight against climate change without causing high economic expenses. It should be noted that, if second-life batteries were used for this purpose in the future, the value of the GWP parameter would be reduced, and the conclusions obtained in this study could vary. Finally, the sensitivity analysis carried out for a set of input variables indicates that household consumption is the input variable that most affect the HB objective function's results, followed by PV generation, electricity price, and maximum allowed SOC, respectively.

Chapter 4

Local Energy Communities Optimization

This chapter provides an overview of the current state-of-the-art in Energy Communities regarding regulation and explores the potential benefits and challenges of this technology. The price and environment-based optimization strategies presented in Chapter 3 are extended and tested for local energy communities.

4.1 Introduction

The generation and use of energy account for more than 75% of the EU's greenhouse gas emissions [362]. To achieve the goal of decarbonizing the economy and reaching net-zero emissions by 2050, the European Green Deal [12] proposes various objectives related to buildings and the residential sector:

- Enhancing the energy efficiency of buildings.
- Ensuring a secure and affordable EU energy supply.
- Prioritizing energy efficiency, improving the energy performance of our buildings and developing a power sector based largely on renewable sources.

These objectives can be achieved through a digitized energy sector based largely on renewable distributed energy resources.

Local Energy Communities (LECs) encourage the participation of consumers in electricity generation and distribution at local level, providing an opportunity for society to be involved in the energy transition. According to the European Commission, a LEC is an open and voluntary association that combines non-commercial aims with environmental and social community objectives. The aim is to promote community-driven and decentralized electricity generation, rather than central generation managed by a small number of large power plants, as has been the case until now.

To take advantage of user participation in local energy communities, electrification at the user's premises is essential; for instance, replacing a conventional natural gas boiler with heat pumps for space heating. Additionally, by increasing electrification, consumers and energy communities can offer flexibility to the energy system through demand response and storage programs.

The IDAE outlines the benefits of local energy communities profiling, including: [363]

- Providing citizens with fair and easy access to local renewable energy resources and the opportunity to benefit from these investments.
- Empowering users to take control and greater responsibility for meeting their energy needs.
- Creating investment opportunities for citizens and local businesses.
- Offering communities to generate income, increasing the acceptance of local renewable energy development.
- Enabling the integration of renewable energy into the system through demand-side management.
- Environmental benefits.
- Social benefits. The creation of local employment and promotion of social cohesion and equity.

The increasing growth and interest in LECs are primarily due to rising electricity prices and an increasing awareness of climate change in society, combined with the growing availability of affordable, small-scale distributed energy resources (DERs). Legislative changes and government subsidies have also helped to accelerate the creation of energy communities and self-consumption generation.

It is a fact that prosumers and their collective forms will play a key role in the forthcoming years by empowering consumers, boosting energy efficiency, and building interconnected energy systems that allow peer-to-peer energy trading and better-integrated grids to support renewable energy sources. This contributes to a fairer transition to climate neutrality that allows citizens to take ownership of energy consumption and production.

4.2 Energy communities regulation

The Clean Energy Package (CEP) [364] introduced by the European Commission has established a legislative framework for the operation of Local Energy Communities across Europe, which aims to facilitate citizens' participation in energy markets, evolving from traditional passive consumers to prosumers. A report made by the European Commission about community renewable energy in Europe confirms this transition, stating that by 2030 energy communities could own 17% of installed wind capacity and 21% of solar Europe-wide. By 2050, almost half of EU households are expected to produce clean energy [365].

The CEP introduces two types of energy communities:

- Citizen Energy Community (CEC). The *Internal Electricity Market Directive (EU) 2019/944* [366] introduces CEC as a legal entity that is based on voluntary and open participation and is effectively controlled by members or shareholders that are natural persons, local authorities, including municipalities, or small enterprises. CEC constitute a new type of entity due to their membership structure, governance requirements and purpose.
- Renewable Energy Community (REC). The *Renewable Energy Directive (EU) 2018/2001* [367] defines a REC as a legal entity that, in accordance with the applicable national law, is based on open and voluntary participation, is autonomous, and is effectively controlled by shareholders or members that are located in the proximity of the renewable energy projects that are owned and developed by that legal entity.

These two EU directives establish a legal framework for collective citizen participation in the energy system. The definition of CEC is similar to REC, but there are some fundamental distinctions. RECs have a specific focus on renewable sources and should be located close to renewable energy projects, while CEC has no such restriction. Another difference is that an energy community can only be called a REC if its activity is based on renewable energy sources, while a CEC may use renewable or conventional sources [368]. For this chapter, a REC is presented. The main differences between the two types of energy communities are summarized in Table 4.1.

	Citizen Energy Community	Renewable Energy Community
Members	Natural persons, local authorities, small micro-enterprises	Natural persons, local authorities, small/micro-enterprises; with the condition that the main professional or commercial activity of the members is not defined by their membership to the REC
Location	No limitations on location, even cross-border Citizen Energy Communities can be established	Members/shareholders need to be in a specific location, close to the associated project of the REC
Activities	Activities in the energy sector targeted exclusively for members; and activities exclusively in the electricity sector for the whole market	All areas of the energy market involving renewable energy
Technology	No limitation on technologies	Only renewable energy technologies

Table 4.1: Comparison of Citizen Energy Community and Renewable Energy Community [9]. Source: The Council of European Energy Regulators.

However, there are still several challenges confronting the proliferation of LECs. A significant challenge is the regulatory barriers due to the complex framework required to obtain the necessary permits and approvals for installation and operation. This can make it difficult for LECs to secure financing and access to the energy grid, as conventional investors may be uncertain about investing in short-time tested business models. This is reflected in Figure 4.1, which demonstrates that the transposition for enabling frameworks and support schemes for energy communities is not consistent across European member states. This map provides a comparative assessment of this progress, using a traffic light grading system to represent how far each country has progressed towards transposing EU regulations on energy communities.

4.2.1 Spanish regulatory framework

This chapter focuses on optimizing a renewable energy community located in Spain. To understand the context in which the proposed case study is developed, the fundamental regulatory aspects are explained below.

The Spanish Government has introduced the definition of REC and implemented policies to promote renewable energy development and encourage citizens' participation in the energy system. Most existing renewable energy communities in Spain use the legal framework provided in Real Decreto (RD) 244/2019 [369] for individual and collective electricity self-consumption and the use of renewable energy sources. It defines collective self-consumption as a group of owners sharing one or several solar panel installations. This limits the scope of energy sharing, particularly excluding other renewable technologies such as wind and small-hydro. As shown in Figure 4.1, Spain is in "average progress" in implementing the European directives in its national law framework; therefore, there is still a way to go.

According to RD 244/2019, a REC installation must comply with at least one of the following requirements:

- Self-consuming owners must be connected to the same LV transformation center.
- The distance between the self-consumer and the energy production center should be no more than 500 meters. The radius was

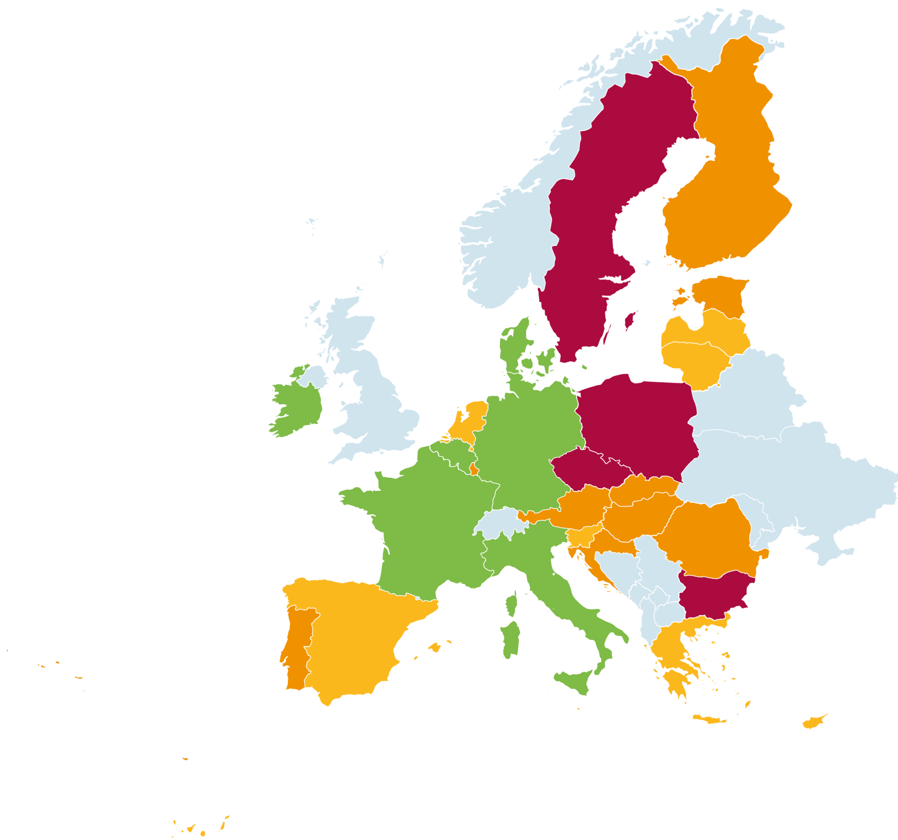

Figure 4.1: Comparative assessment of the progress for enabling regulation frameworks for RECs in the diverse European countries following a traffic light grading system in which red stands for bad transposition and green for best practices [7]. Source: REScoop.

recently expanded to 1km; however, this additional distance is only available to PV self-consumption if located on buildings.
- The photovoltaic production facility and the self-consumers must share the same cadastral reference.

There are two connection modalities available:

- Collective self-consumption with a connection through the public grid: The PV production is shared via the public grid. It is connected to the LV grid through a bi-directional smart meter, and the retailer compensates the end-user(s).
- Collective self-consumption with direct connection to the internal grid: In this case, the photovoltaic installation does not connect to the public grid, but the photovoltaic production is distributed directly to each of the internal grids of the self-consumers. This connection is typically used in large industrial customer installations.

The following modalities specified in RD 244/2019 require all consumers to belong to the same self-consumption modality. These modalities are:

- Collective self-consumption without surpluses: an anti-spill system is used to avoid injecting surplus energy into the electricity grid.
- Collective self-consumption with surpluses not subject to compensation: the owner of the generation facility sells the surplus energy to the electricity market.
- Collective self-consumption with surpluses subject to compensation: consumers receive financial compensation for the surpluses they inject into the electricity grid. The retailer is responsible for compensating the surplus energy cost at the end of each billing period. Surpluses that exceed imported consumption are not compensated.

Sharing coefficient strategies

The Spanish regulation, RD 244/2019, establishes guidelines for distributing the generation among customers, assigning a fixed sharing distribution coefficient value to each participant of the local energy community. The coefficients of all participating consumers must sum up to 1. In other words, if there is only one associated consumer, their coefficient will be 1 for each hour of the day. And if there are multiple consumers, the sum of all coefficients will also be 1. New improvements in the above-mentioned regulation expand the potential of LECs. These improvements introduce dynamic distribution coefficients for the energy produced by a LEC, while still maintaining the option of using fixed coefficients. Thus, the variable coefficients must be established for each hour of the year and may be changed every four months.

And how are these coefficients calculated? They can be determined based on several factors:

- The contracted power of each participating associated consumer.
- The financial contribution of each consumer to the photovoltaic installation.
- Other elements agreed upon by all participants in the shared self-consumption, such as balancing the economic savings among all participants.

This thesis does not address the optimization of the hourly value of distribution coefficients. However, the author recommends further investigation of this aspect in future research. Exploring the optimization of sharing coefficients (both dynamic and static) based on the desired objective could be engaging, whether it is minimizing the overall cost of the energy community or reducing the emissions associated with consumption, for instance.

To conclude this section, it is essential to remember and emphasize that regulations for Energy Communities are still evolving and can vary widely between countries and regions in a brief period.

4.3 Methodology

This section considers the existing regulatory framework in Spain and presents the methodology developed to optimize the operation of a Local Energy Communitiy. A scheme of the method followed is presented in Figure 4.2.

The methodology begins by setting up the case study and scenarios. The required input data includes the location of the renewable energy community, specifications of the BESS and PV systems, the number of electrical supply endpoints and their consumption profiles, PV system generation, hourly electricity price tariff for each end-user and grid GWP. The scenarios are established based on two variables linked to a specific analysis: sharing coefficient strategy and the REC optimization strategy. All possible combinations of these variables are established to generate a comprehensive set of scenarios for evaluation and discussion. The results will demonstrate the potential cost savings for the overall REC and the CO_2 emissions avoided through the optimization strategies designed in this chapter.

Regarding the LCA analysis, Table 4.2 displays the range of GWP indicators for the overall life cycle stages for each electricity generation source type, as reported in [8].

Table 4.2: Lyfe cycle emission factors for electricity generation sources [8, 10].

Generation source GS_i	**GWP range [kg** $\mathbf{CO_{2-eq}}$**/kWh]**	**Average GWP [kg** $\mathbf{CO_{2-eq}}$**/kWh]**
Hard coal	0.660-1.05	0.855
Lignite	0.800-1.30	1.05
Natural gas	0.38-1	0.690
Nuclear	0.003-0.035	0.019
Biomass	0.008-0.130	0.069
Hydro-power	0.002-0.02	0.011
Wind	0.003-0.041	0.022
Battery [10]	-	0.060

For this chapter, the average GWP value is used. The battery GWP value has been updated concerning the previous Chapter. Figure 4.3 illustrates the methodology used to calculate the coefficients associated with the hourly emissions of the energy mix using values from Table 4.2. This process begins by selecting the generation sources with the highest poundage in the energy mix. Ideally, the generation sources should rep-

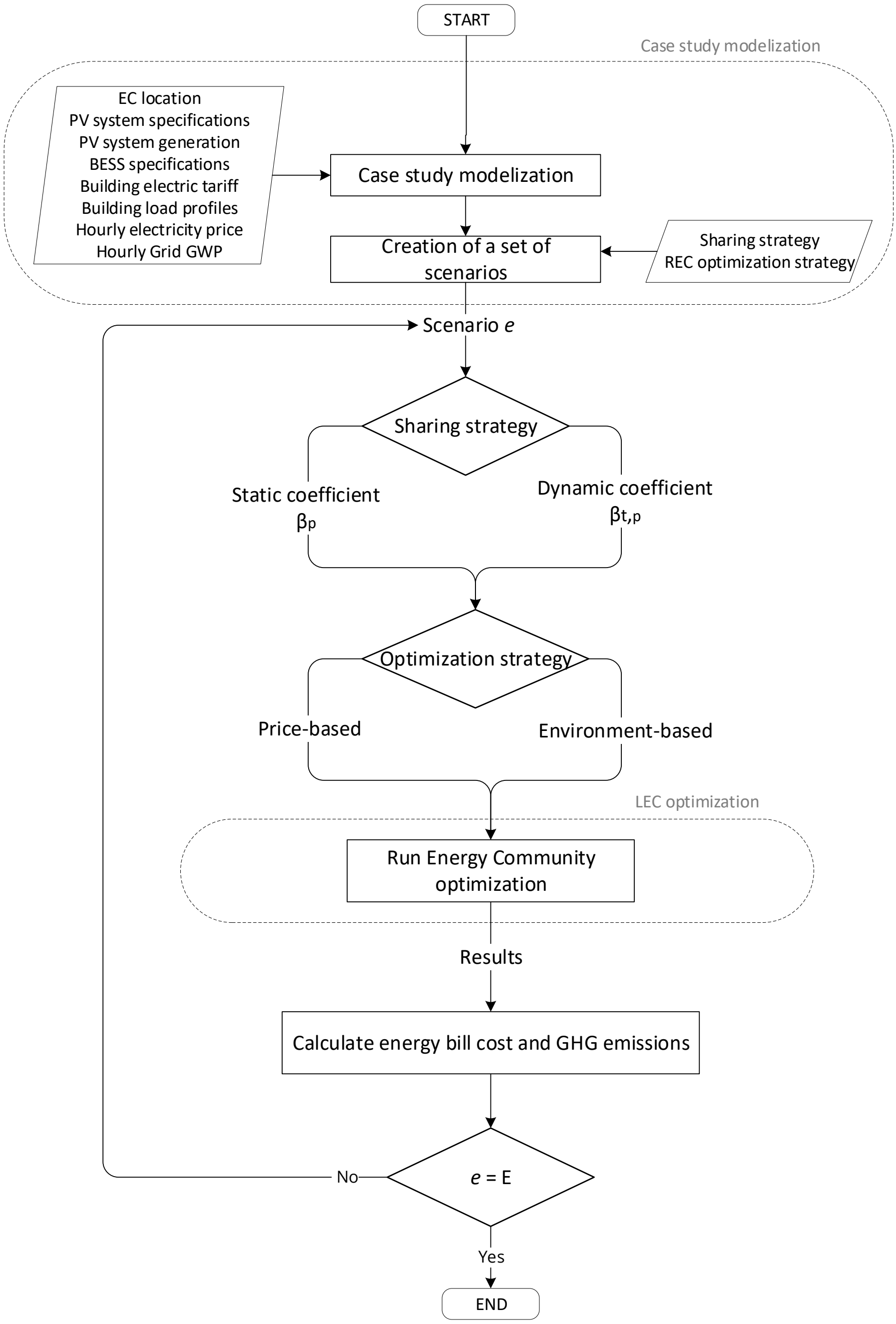

Figure 4.2: LEC methodology employed in this chapter.

resent almost all the generated power so that the grid emissions indicators are as precise as possible to reality. Once the generation sources (GS) have been specified, the data for the next day's scheduled generation, which are public and available the day before their provisioning, are acquired. For each hour, the amount of scheduled generation for that technology g is multiplied by its corresponding GWP emission factor. When calculated for the entire set of generation sources G, the total GWP of the network for the hourly time interval t is calculated. This process is repeated for each hour and day, successively.

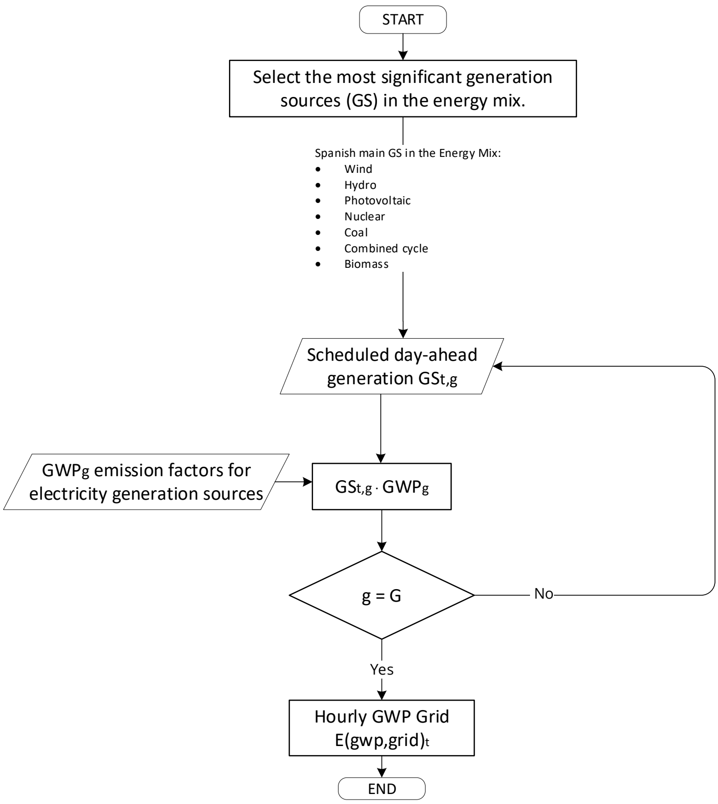

Figure 4.3: Methodology employed to calculate the hourly global warming potential of the grid.

4.4 Energy Community mathematical formulation

This section mathematically models the different assets that participate in the LEC, such as photovoltaic generation, BESS, and the group of consumers who will benefit from this clean energy generation. First, the objective functions and system constraints are defined, followed by the presentation of the mathematical models for each component that takes part in the LEC.

4.4.1 Objective functions

The objective functions in this chapter have a similar structure to those in Chapter 3. The main difference is that now there are multiple consumers p, and selling energy is considered in this case study. Therefore, as the primary purposes of each objective function have already been explained in detail in Chapter 3, Section 3.4.1, they will not be further elaborated here.

Price-based

This program focuses exclusively on the economic aspect. It minimizes the overall LEC electricity bill (4.1), considering the centralized battery degradation cost K_t^{cal} due to calendar aging, where P_t^{buy} and P_t^{sell} are the time-varying price for buying and selling electricity. The variables $\chi_{t,p}^{buy}$ and $\chi_{t,p}^{sell}$ refer to the virtual energy purchased and sold to the grid by each participant p, and P^{VAT} is the tax used. The objective function is expressed as

$$\min\ f_1 = \sum_{t=1}^{T}\sum_{p=1}^{P}(P_{t,p}^{buy}\chi_{t,p}^{buy} - P_{t,p}^{sell}\chi_{t,p}^{sell} + K_t^{cal}) \tag{4.1}$$

Environmental-based

This optimization strategy attempts to minimize the carbon footprint occasioned by the generation sources that provide electricity to each participant involved in the LEC. In this approach, only the emissions from the energy consumed are considered, not those sold to the grid. The objective function is formulated as follows.

$$\min f_2 = \sum_{t=1}^{T}\sum_{p=1}^{P}(E_t^{gwp,grid}\chi_{t,p}^{buy}) + E^{gwp,pv}W_t^{pv} + E^{gwp,bat}\sigma_t^{dis} \tag{4.2}$$

where $E_t^{gwp,grid}$ indicates the kg CO_{2-eq}/kWh of the grid on average per period t. It is calculated with the hourly energy production mix, taking the values of scheduled generation in the day-ahead market for each technology described in Table 4.2. The emission parameters associated with the battery $E^{gwp,bat}$ and photovoltaic $E^{gwp,pv}$ are also indicated in Table 4.2.

4.4.2 LEC constraints

Energy balance

The energy balance of the Local Energy Community allows for distinguishing the energy generated through the centralized PV and BESS and imported from the grid. In this case, a sharing coefficient $\beta_{t,p}$ is virtually associated with each participant. Therefore, if a participant p has a constant/static sharing coefficient of $\beta_{t,p} = 0.5$, it means that this consumer owns 50% of the electricity generated by the renewable generation of the community.

The energy balance expressed in 4.3 is designed to simulate that the participants have a battery and photovoltaic generation. To achieve a balanced energy system, the total electricity imported from the grid $\chi_{t,p}^{buy}$, must balance the production from generation units, consumption from load units, charging and discharging of the central BESS and energy sold for each period $t \in T$:

$$\beta_{t,p}(W_t^{pv} + \sigma_t^{dis} - \sigma_t^{ch}) + \chi_{t,p}^{buy} = W_{t,p}^{inflex,load} + \chi_{t,p}^{sell} \tag{4.3}$$

Not buy and sell at the same time

Binary variables $\delta_{t,p}^{buy}$ and $\delta_{t,p}^{sell}$ are now introduced in order to ensure that it is not possible to sell and buy in the same period for each participant. It is possible that when visualizing the total consumption of the LEC, there may be periods where consumption and selling occur simultaneously. However, this is due to the fact that one participant is selling while another is not. This happens because each client may have different distribution coefficients, resulting in one having surpluses while the other does not.

$$\delta_{t,p}^{buy} + \delta_{t,p}^{sell} \leq 1 \tag{4.4}$$

Prosumers capacity limits

Electricity bought and sold must be below power limits, according to the terms stipulated in the retail contract:

$$\chi_{t,p}^{buy} \leq \delta_{t,p}^{buy} \cdot X_p^{max,import} \tag{4.5}$$

$$\chi_{t,p}^{sell} \leq \delta_{t,p}^{sell} \cdot X_p^{max,export} \tag{4.6}$$

Net generation

The net-generation θ_t follows this equation

$$\theta_t^{lec} = W_t^{pv} + \sigma_t^{dis} - \sigma_t^{ch} \tag{4.7}$$

The individualized net hourly energy generated by those energy community participants p that carry out collective self-consumption, $\theta_{t,p}$, is

$$\theta_{t,p} = \beta_{t,p}\theta_t^{lec} \tag{4.8}$$

where θ_t^{lec} represents the total hourly net energy produced by the generator smart meter and $\beta_{t,p}$denotes the hourly distribution coefficient among consumers participating in the collective self-consumption of the energy generated, in period t. The exact value is repeated for all t when referring to the static sharing coefficient. Consumers are required to submit the coefficients of participants involved in self-consumption for all hours of the current year, which cannot be modified within the same year.

Sharing strategy

The sum of all sharing coefficients allocated to the LEC participants for each period t must equal 1.

$$\sum_{p=1}^{P} \beta_{t,p} = 1 \tag{4.9}$$

4.5 Case Study: Optimization of a Spanish Energy Community

This study is focused on an existing renewable energy community located in Spain, consisting of four office buildings and a collective-owned centralized photovoltaic system and a Li-ion BESS. These components are connected downstream of an inverter that feeds energy back into the grid, which is then compensated in the electricity bill of each participant based on their assigned generation percentage. In addition, this smart meter is bidirectional and capable of purchasing electricity from the grid when it is necessary to store energy in the battery. Each building has its own smart meter and purchases energy from the grid since they have no flexible assets behind the meter. Figure 4.4 illustrates the components of the renewable energy community: four office buildings and a centralized PV and BESS as a collective renewable generation unit. The Local Energy Community Management System is responsible for controlling and optimizing the LEC flexible asset and must send optimal operating set-points for the centralized BESS, the only flexible source as illustrated in Figure 4.4.

This case study is based on an actual energy community located in Spain. Data for PV generation and building consumption ranges from March 3 to December 31, 2022. According to the European and Spanish regulations, this study belongs to the category of a renewable energy community with collective self-consumption and surplus compensation [369]. In other words, the energy generated is used to supply multiple consumption points. This group of participants agrees to distribute the renewable energy generated and applies static sharing coefficients for each consumer. For each end-user, any surplus energy that is not

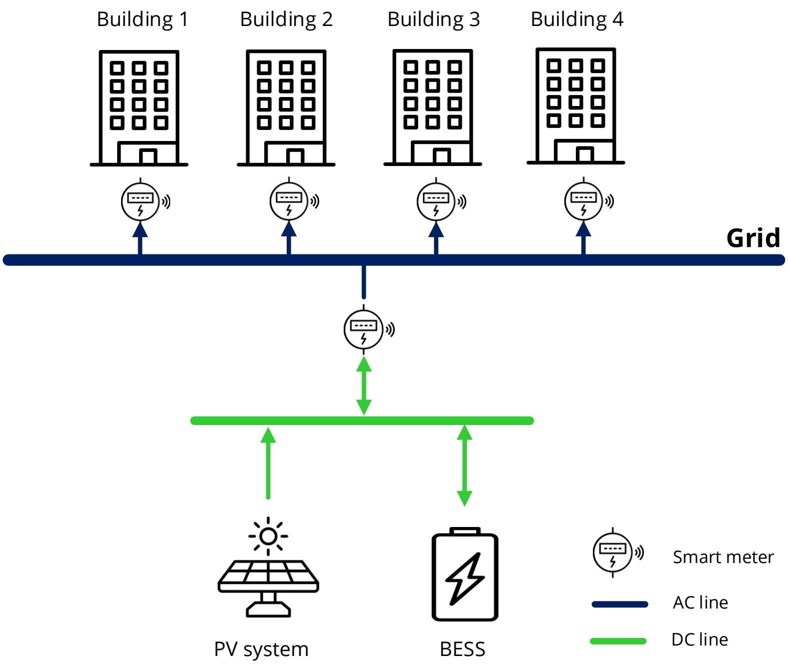

Figure 4.4: Case study local energy community simplified scheme.

consumed is fed back into the grid and economically compensated on its electricity bill.

4.5.1 Energy Community specifications

The BESS parameters and specifications used in this study are listed in Table 4.3. The optimization imposes that the state of charge of the battery is the same at the beginning $\sigma_{t=0}^{soc}$ and at the end of the optimization horizon $\sigma_{t=end}^{soc}$, which is 24 hours ahead in this study. This prevents the battery from being completely discharged at the end of the day.

Table 4.3: LEC BESS parameters.

Input Parameters	**Value**
Maximum charging power allowed	90 kW
Maximum discharging power allowed	90 kW
Maximum SOC	189.9 kWh
Minimum SOC	31.65 kWh
Efficiency charging	0.95
Efficiency discharging	0.95
$\sigma_{t=0}^{soc}$	150 kWh
$\sigma_{t=end}^{soc}$	150 kWh

Table 4.4 lists the static parameters related to each building that constitutes the energy community and Table 4.5 indicates the static sharing coefficient rate of generation corresponding to each building.

4.5.2 Grid GWP calculation

The proposed case study and scenarios used data from March 3rd - December 31st, 2022, the available date range from the energy community. The hourly share of each generation source (see Table 4.2) in the Spanish energy mix is illustrated in Figure 4.5. The selected generation types are primarily responsible for the total generation and cover

Table 4.4: LEC participants maximum contracted power.

Input Parameters	**B1**	**B2**	**B3**	**B4**
Spanish electricity tariff type	3.0	3.0	3.0	3.0
Maximum contracted power Period 1	70	43.65	20.785	75
Maximum contracted power Period 2	70	43.65	20.785	75
Maximum contracted power Period 3	70	43.65	20.785	75
Maximum contracted power Period 4	70	43.65	20.785	75
Maximum contracted power Period 5	70	43.65	20.785	75
Maximum contracted power Period 6	70	43.65	20.785	75

Table 4.5: LEC participants static sharing coefficient.

	B1	**B2**	**B3**	**B4**
Static sharing coefficient	0.35	0.15	0.02	0.04

96% of the total. The presence of coal is almost nonexistent, but during periods of high energy demand, the energy produced from it increases. Nuclear power serves as the base generation. In summer, photovoltaic and combined cycle generation significantly increase, mainly due to the rise in demand caused by high temperatures in most regions of the country. Hydroelectric generation increases its share in the spring months and late winter months.

Following the methodology presented in Chapter 3, the hourly grid GWP is calculated and shown in Figure 4.6. As expected, fossil-based generation sources such as coal, combined cycle and cogeneration contribute the most to the greenhouse gas emissions in the energy mix, despite not being the generation sources with higher production. For instance, although the energy produced from combined cycle cogeneration and coal in the mix is not the most significant, they are the sources that contribute the most CO_2 emissions to the grid.

Finally, the hourly grid GWP is calculated and displayed in Figure 4.7. Summer months display higher emissions associated with the grid energy mix, mainly due to the increased demand during these months

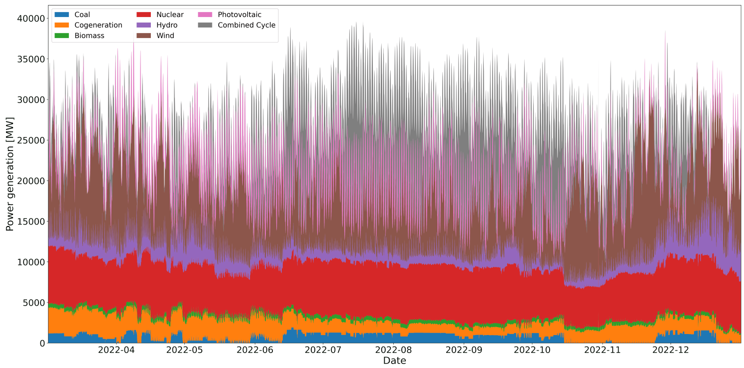

Figure 4.5: Generation sources energy mix sharing for the case study.

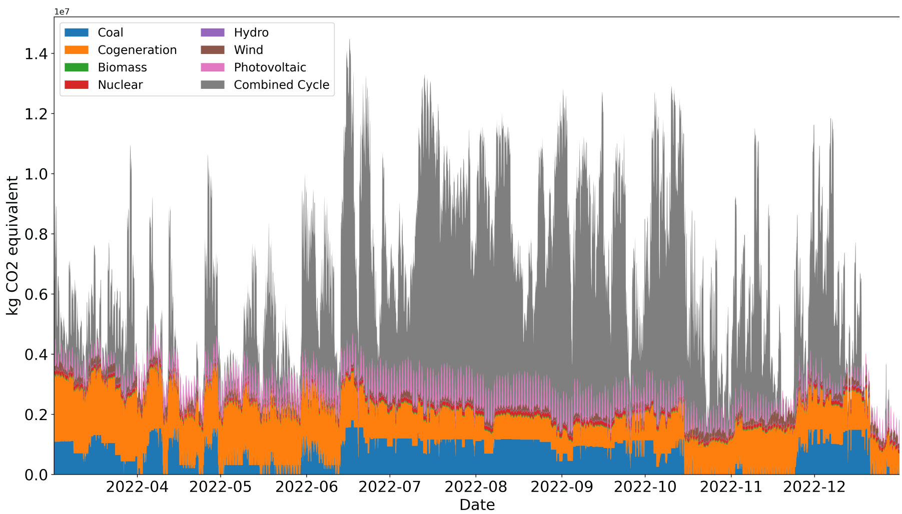

Figure 4.6: Spanish energy mix kg CO_2 equivalent in the energy mix.

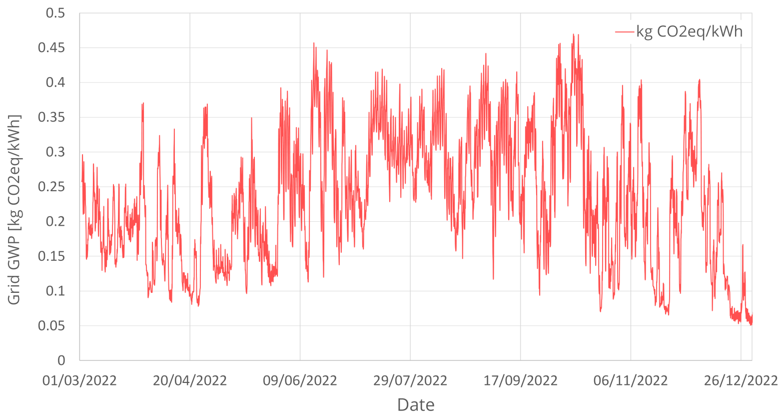

Figure 4.7: Spanish GHG emissions in the energy mix grid.

and lower production of specific renewable resources such as hydro and wind power. However, thanks to the increased presence of photovoltaic generation in the last years, emissions have been partially reduced during the summer months compared to other past years. Low emissions levels are usually associated with low energy demand in the system (i.e., mild temperatures) and high penetration of renewables, especially wind power.

4.6 Results

This section presents the results of the different scenarios analyzed in the proposed case study. To enhance the clarity and comprehensibility of the results, only two representative days are displayed to illustrate the behavior of the centralized battery and energy community for the price and environment optimization strategies. Additionally, this section includes an analysis of the energy community costs, as well as the GHG emissions associated with their consumption for the case study period. This information is valid to evaluate the effectiveness of the proposed optimization strategies.

Table 4.6: Baseline LEC cost and GHG emissions.

Building	**Total cost** (€)	**Total GHG emissions** (t CO_2)
B1	13 414.72	18.81
B2	6 004.96	7.04
B3	390.46	0.47
B4	30 326.32	35.63
LEC	**50 136.45**	**61.94**

4.6.1 LEC baseline consumption

Prior to presenting the results of the two optimization strategies for the LEC, the cost and GHG emissions associated with the consumption of its participants are shown, assuming they did not have access to collective generation and had to purchase all energy directly from the grid. The aim is to compare the effectiveness of the optimization developed based on these baseline data. Table 4.6 shows the cost and GHG emissions costs and GHG emissions associated with each building over the 10-month study period. Thus, the total LEC electricity cost is 50 136.45 €, and the emissions associated are 61.94 tones of CO_{2-eq}.

4.6.2 Price-based approach

The price-based scenario aims to minimize the overall energy bill cost for participants of the LEC by utilizing the flexibility of the central BESS and the inflexible PV generation.

To demonstrate the feasibility and functionality of the developed Local Energy Management System (LEMS), its performance over two consecutive days is presented in Figure 4.8. This figure consists of four sub-figures, which are described and explained in detail through this sub-section.

The upper graph (Figure 4.8a) displays the hourly price for buying and selling electricity on the grid. The purchase (Price buy) and self-consumption (Price sell) SPOT prices are used in this study. The buying price is always higher than the selling price because the taxes for purchasing are considered. Also, in this Figure, the start of the second day (Day *d+1*) is indicated by a dashed black vertical line on all graphs to aid the reader. Additionally, to facilitate comprehension, periods with low electricity prices from the grid are highlighted in blue, while periods with higher prices are emphasized in light red. These elements will help with the visual explanation.

The following image (Figure 4.8b) displays the behavior of the centralized battery of the LEC. It can be observed that during periods with high purchase prices from the grid, the battery discharges to reduce the costs associated with buying energy from the grid for the users of the community and/or selling excess energy to obtain economic remuneration in return.

Figure 4.8c shows the state of charge of the centralized battery, the baseline consumption of the energy community (i.e., the actual and inflexible consumption of all buildings that are part of the LEC), and the consumption associated with the LEC, considering centralized generation sources (battery and photovoltaic generation). During periods of low prices (blue time slots), the LEC takes advantage of buying energy from the grid to charge the battery and use this energy later during periods of high prices. Therefore, those consumption peaks are the battery charge, but they are associated with each user, as the cost is distributed among them, using the corresponding distribution coefficient. On the other hand, during periods of high prices (red time slots), the battery discharges to alleviate the cost for the participants of the LEC. Thus, the consumption of the LEC is below the baseline, thanks to photovoltaic and battery discharge.

Finally, Figure 4.8d shows the generation sources (battery discharge and PV) and the energy sold to the grid, for which participants will receive economic benefits for the supplied energy. To facilitate understanding of this graph, the author has assigned negative values to the generation sources or energy sold and positive values to consumption sources. According to the amount of energy sold to the grid (see periods 9 to 11 on the day *d*), the battery discharges more than needed to meet user demand and sells this excess energy at a higher selling price. This occurs again during periods 21 (day *d*) and 32-34 (day *d+1*). The

Table 4.7: Price-based LEC cost and GHG emissions.

Building	**PB Total cost** (€)	**Baseline**	**PB Total GHG** (t CO_2)	**Baseline**
B1	9 529.01	↓ 29.0%	21.65	↑ 15.1%
B2	4 301.87	↓ 28.4%	11.15	↑ 58.3%
B3	184.17	↓ 52.8%	5.64	↑ 1108.2%
B4	24 658.09	↓ 18.7%	36.14	↑ 1.4%
LEC	**38 673.14**	↓ 22.9%	**74.58**	↑ 20.4%

reason is that selling energy to the grid is very profitable, making it worthwhile.

Overall, the case study results demonstrate the effectiveness of the LEMS in reducing costs and generating economic benefits for the energy community participants. The price-based results for 10 months of time horizon, and assuming perfect predictions, are summarized in the following Table 4.8. The objective is to reduce the overall LEC cost above the individual benefit. The LEC reduces the energy bill a 22.9% (11 463.31 €) in exchange for raising the GHG-associated emissions by 20.4% (an increment of 12.64 tones of GHG emissions), mainly due to the BESS usage, since when it buys energy from the grid, the grid and BESS emissions are considered.

It should be mentioned that the installed photovoltaic generation covers, on average, 17% of the total LEC consumption; thus, the savings derived from photovoltaics are not very noticeable. However, significant savings can be achieved thanks to the battery, which purchases energy during low-cost periods and injects it into the grid during high-cost periods.

4.6.3 Environment-based approach

This optimization strategy aims to minimize the GHG emissions associated with the electricity consumption of the energy community participants. To quantify these emissions, the GHG of the grid is considered, which varies hourly depending on the energy mix, centralized PV generation, and BESS use.

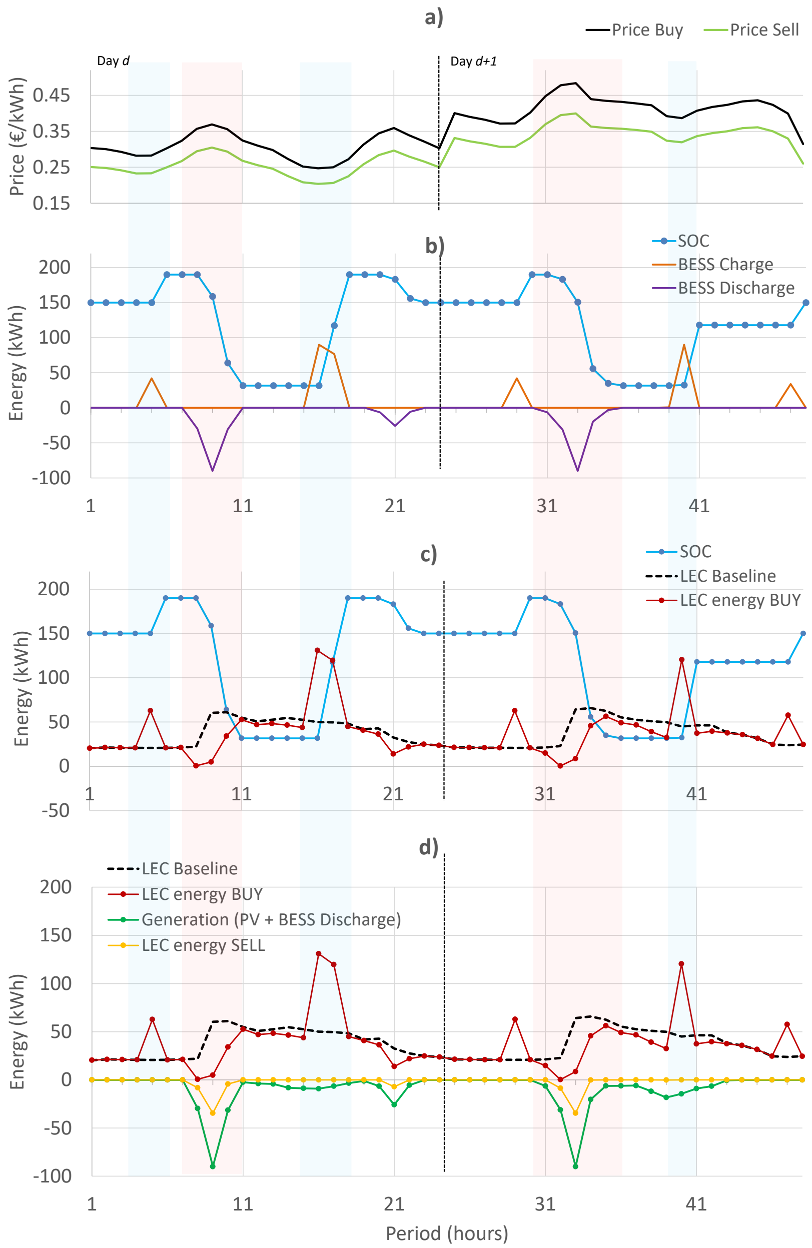

Figure 4.8: Price-based LEC optimization results. a) Electricity price for buying and selling to the grid. b) BESS performance c) LEC behavior after optimization d) LEC consumption, net generation produced and electricity sold.

To minimize greenhouse gas emissions associated with consumption, two different days are selected to explain the different behavior of the LEMS depending on the energy mix: the first with a high penetration of renewables in the grid, resulting in a low GWP index, and the second with a high percentage of fossil sources in the energy mix.

4.6.3.1 Low grid GWP

Following the same format as the previous approach, Figure 4.9 presents the results of the energy community for two days of optimization.

The figure above (Figure 4.9a) shows the emissions curve associated with the hourly energy mix of the grid. The higher the curve, the greater the emissions associated with consumption if energy is purchased from the grid.

Figure 4.9a displays the emissions curve associated with the hourly energy mix of the grid. The higher the curve, the greater the emissions associated with consumption if energy is purchased from the grid.

In the second image, Figure 4.9b, the battery is discharged during periods of the day with higher pollution indices (period 8), thus reducing emissions associated with consumption. As the battery capacity must be the same at the beginning and end of the day (150 kWh), surplus photovoltaic energy is used to charge the battery during period 16. On day *d+1*, the battery is slightly discharged during the period of maximum GHG emissions in the grid and charged with surplus solar energy. On day *d+1*, the battery is hardly used because, thanks to the high penetration of renewables, it is cleaner to purchase energy from the grid.

In Figure 4.9d, it can be observed that photovoltaic energy supplies part of the local energy community's consumption, and in specific periods, this photovoltaic energy is used to charge the battery.

Overall, the case study demonstrates the potential benefits of combining renewable energy sources with battery storage systems to reduce greenhouse gas emissions and increase energy efficiency.

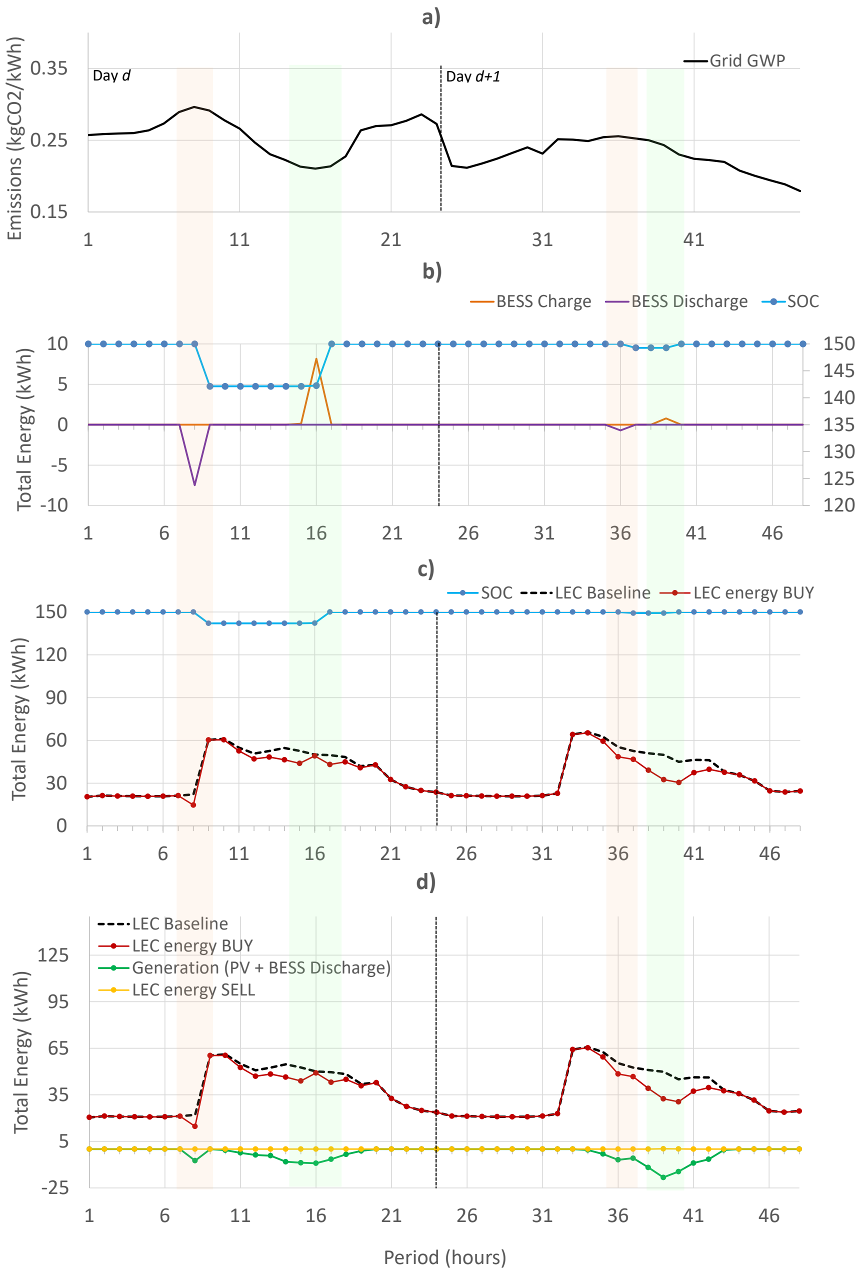

Figure 4.9: Environment-based LEC optimization results with high renewables penetration in the grid. a) Hourly grid emissions. b) BESS performance c) LEC behavior after optimization d) LEC consumption, net generation produced and electricity sold.

4.6.3.2 High grid GWP

For this subsection, two days with a high percentage of fossil sources in the energy mix have been selected. Figure 4.10a shows the GWP with high GHG indices and the GWP of the previous example to compare both magnitudes.

At first glance, the main difference is that the use of the battery increases as the GWP index in the national energy system becomes higher. For the first day (d), the battery discharges during the period with the highest pollution associated with the grid. In the following periods, the battery recovers its state of charge thanks to photovoltaic generation. From periods 21-24, the battery discharges, coinciding with the hours of highest emissions on the grid. For the next day (day d+1), the battery starts with the initial state of charge of 150 kWh, as the restriction imposes. The BESS discharges from 28 to 34, also coinciding with a period of high emissions. From there, the battery recovers its SOC thanks to the photovoltaic energy from the LEC and by charging it in period 42 at maximum power (90 kW) to meet the SOC restriction. This charge is made during the period when emissions are the lowest during that day.

In Figure 4.10c, it can be observed that the LEC only makes a greater electricity purchase than the baseline to charge the battery in period 40. As mentioned before, this period for day d+1 is the one with the lowest CO2.

In Figure 4.10d, it can be seen that no energy is sold to the grid and that the photovoltaic generation covers part of the participants' demand. During the night and in periods of high emissions, the battery discharges part of its capacity to reduce the emissions associated with consumption.

4.7 Conclusions

The LEC optimization results demonstrate the feasibility and satisfactory performance of the approaches proposed in this thesis. On the one hand, the price-based optimization achieves savings of approximately

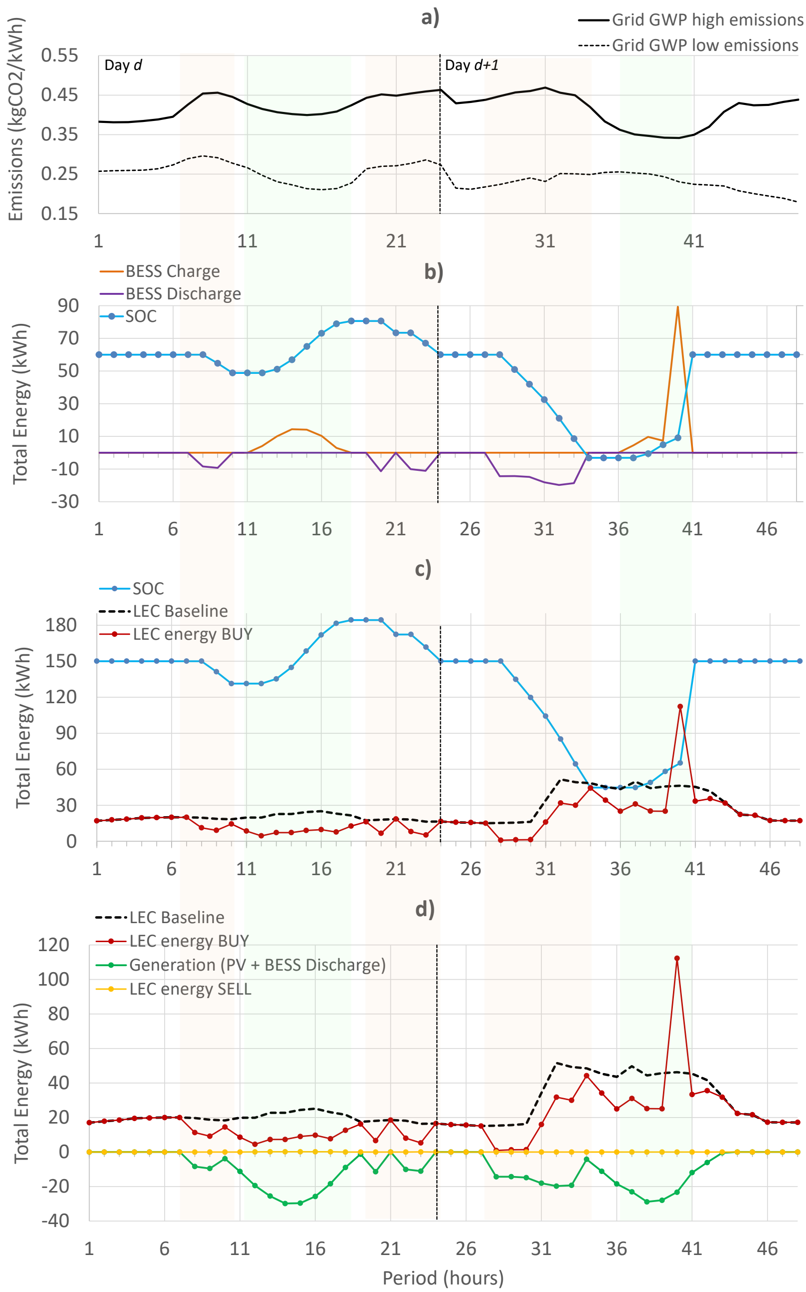

Figure 4.10: Environment-based LEC optimization results with low renewables penetration in the grid. a) Hourly grid emissions for low and high renewable penetrations in the grid. b) BESS performance c) LEC behavior after optimization d) LEC consumption, net generation produced and electricity sold.

Table 4.8: Environment-based LEC cost and GHG emissions.

Building	EB Total cost (€)	Baseline	EB Total GHG (t CO_2)	Baseline
B1	10 131.23	↓ 24.5%	16.82	↓ 10.6%
B2	4 590.18	↓ 23.6%	7.30	↑ 3.7%
B3	210.17	↓ 46.2%	2.28	↑ 3.87%
B4	25 757.76	↓ 15.1%	31.88	↓ 10.5%
LEC	**40 689.35**	**↓ 18.8%**	**58.27**	**↓ 5.9%**

23% (11 500 €) compared to the LEC baseline consumption. The contribution of photovoltaics compared to consumption is lower than this percentage, so the battery plays a crucial role in achieving these savings. However, the associated emissions increase by 20%. On the other hand, the strategy of minimizing emissions associated with consumption achieves a reduction of approximately 6% of these contaminants compared to the baseline consumption, avoiding the emissions of 3.7 tons of CO_{2-eq}. The difference in GHG release between the price and environmental-based strategies is 21%.

The positive aspect is that the environmental strategy, in addition to minimizing emissions, is capable of reducing costs by 18% (approximately 10 000 €) compared to the baseline, making it feasible to choose to enhance the environment without incurring high costs, which is a significant advantage.

However, there are still some challenges confronting the proliferation of LECs. A significant challenge is regulatory barriers due to the complex framework required to obtain the necessary permits and approvals for installation and operation. This can make it difficult for LECs to secure financing and access to the energy grid, as conventional investors may be uncertain about investing in short-time tested business models. Additionally, technical challenges associated with integrating renewable energy sources into the energy grid and storage systems are common. Also, at this early stage, LECs may lack the technical expertise to design, build, and operate renewable energy projects. Finally, due to a lack of awareness, many citizens may be unaware of the benefits and opportunities offered by energy communities, making it challenging to attract members and build critical support.

Despite these obstacles, LECs offer multiple benefits to neighborhoods and the environment, providing promising investment opportunities. One main advantage is the reduction of dependence on fossil fuels, resulting in a decrease in greenhouse gas emissions. By generating and consuming energy locally, LECs avoid losses due to the Joule effect and costly long-distance energy transmission and distribution lines. Furthermore, LECs can enhance energy security by reducing the vulnerability of areas to power outages and disruptions. They also promote community engagement and empowerment by enabling end-users and small groups to participate in the energy system, offering communities the possibility of generating income and having a more significant influence on the electricity market.

Chapter 5

Personalized Federated Learning for Energy Management Systems

This chapter proposes a personalized federated learning methodology for home energy management systems demand forecasting that incorporates a cost-oriented loss function while preserving customers' data privacy and security.

5.1 Introduction

The energy sector is undergoing a rapid transformation towards a distributed, digitized and decarbonized system, moving away from the traditional centralized, rigid, fossil fuel-based energy system [370]. The advent of digitization has given consumers the availability of fine-grained electricity consumption and self-generation data collected by smart meters. This has enabled the creation of data-driven energy services that cater the needs of end-users, such as Home Energy Management Systems (HEMS). In addition, recent advances in Artificial Intelligence (AI) technology boost the extraction of better meaningful insights and valuable information from smart meter data, leading to the development of more accurate prediction models and innovative energy services [263]. As a result, energy providers can offer personalized solutions to customers, ultimately creating a more efficient, sustainable, and customer-centric energy system.

The transition to a more sustainable energy system offers many advantages but also presents technical challenges that must be addressed to achieve digital maturity in the distribution network. This transformation requires continuous evolution of the energy legal framework [371]. For instance, the European Union's enforcement of the General Data Protection Regulation (GDPR) [372] requires the implementation of robust data privacy and security measures. To meet these challenges, Federated Learning (FL) technology has emerged as a promising solution [373–376].

Federated learning, as one of the collaborative learning techniques, is a machine learning technique that enables the training of a centralized model using multiple decentralized edge devices or servers holding local data samples without exchanging them [377]. This approach is particularly suitable for addressing customer data privacy and security concerns when using smart meter data in energy services and solutions [378]. Furthermore, FL mitigates the challenges of data silos in the energy sector by enabling collaboration between stakeholders, such as energy providers and customers, while preserving privacy, security, access rights and reducing or eliminating cloud-computing costs [379]. These advantages make FL a promising solution for power system applications.

In power systems, FL has demonstrated its applicability in various areas, including energy management systems, load forecasting, and anomaly detection, among others [380]. Based on this state of the art, literature has moved towards personalized FL models in recent years, where each edge server or device personalizes the federated forecasting model by re-training the global model with their historical data. Compared with traditional FL algorithms, the aforementioned method balances the global FL models meanwhile respecting the local data distribution. In [381], a novel personalized federated learning approach is proposed for individual consumer load predictions to improve forecasting accuracy with privacy protection.

From an economic perspective, the growth of distributed energy resources (DER) and the implementation of demand response programs for enhancing flexibility pose a significant challenge in maintaining a stable and balanced electricity system [382]. Consequently, this deviation from the scheduled energy could result in high operating costs or economic penalties for electricity market participants, such as BRPs.

State-of-the-art forecasting models typically use the Mean Square Error (MSE) loss function for training; however, MSE does not accurately reflect the actual costs associated with forecasting errors in power systems since cost functions are often neither symmetric nor linear. To address this issue, [383] published the first cost-oriented loss function method. This solution better captures the economic impact of forecasting errors, optimizing the trade-off between forecast accuracy and its associated cost. Recent literature proposes various approaches, such as a load forecast differentiable cost-oriented loss function that uses an optimal piece-wise linear approximation method [384] and a wind power generation model using a cost-oriented loss function [385]. Additionally, [386] presents two cost-oriented approaches to improve day-ahead load forecasts in terms of a more cost-effective real-time operation with respect to actual loads.

5.2 Federated Learning in Energy Systems

This section introduces the FL and reviews the existing FL applications within the energy system domain. Following this, the proposed personalized FL methodology is presented.

Within the energy system domain, massive data is being generated daily from multiple digital sources such as smart meters, sensors/sensing devices installed in transformers or transmission lines [387], and this amount of data is expected to increase over the years. Typically, this data is stored in a centralized system, which incurs high storage and communication infrastructure costs. However, some of these data may contain users' private information, which could conflict with GDPR security requirements [372]. To provide a solution to these issues, FL technology has emerged as one of the most promising solutions [388]. FL processes data locally at each source instead of central collection and processing. This reduces the amount of data transmitted over the network, enhances data privacy, and allows models to be trained on a larger dataset by combining data from multiple sources. Supervised learning energy services using FL can be classified into the following FL types, depending on their data partitioning category [192]:

- Horizontal Federated Learning (HFL) is introduced in scenarios where data sets share the same feature space but different samples (Figure 5.1 (a)).
- Vertical Federated Learning (VFL) applies to the cases where two data sets share the same sample ID space but differ in feature space (Figure 5.1 (b)).
- Federated Transfer Learning (FTL) applies to scenarios where two data sets differ not only in samples but also in feature space (Figure 5.1 (c)).

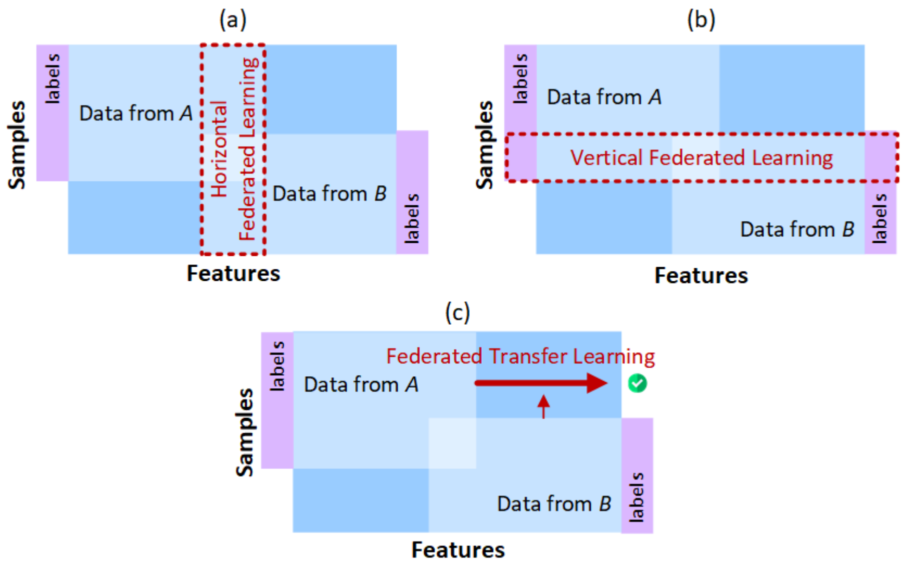

Figure 5.1: Categorization of federated learning regarding data partition for supervised learning: (a) Horizontal federated learning, (b) Vertical federated learning and (c) Federated transfer learning.

General FL concepts are reviewed in [192], which defines the architectures and applications for the federated learning framework and provides a comprehensive survey of existing works on this subject, while [194,389] focus more on the challenges and future directions. Specifically, [390] conducts a comprehensive review of FL applications in the energy domain. With regard to energy management systems, [391] reports one of the first use of federated learning to support the Microgrid Energy Management System.

Key applications of FL in power systems and their article-related are listed in Table 5.1.

Table 5.1: Main federated learning applications in the energy system domain.

FL Applications in Power Systems	References
Anomaly detection, predictive maintenance	[392,393]
Consumer characteristic identification	[373,394]
Electrical load forecasting	[376,395,396]
Energy management systems	[391,397]
Energy theft detection	[398]
EV demand forecasting and charging station recommendation	[341,399,400]
Load clustering	[376,396]
Solar generation disaggregation	[401]
Solar irradiation forecasting	[402]
Voltage forecasting	[403]

5.3 Personalized Federated Learning Methodology for HEMS

The challenge of creating a unique global ML model common to a group of customers is that the global model may not fit all the customers' patterns. A proposed solution to this problem is personalization to understand user behavior and adapt to it. It consists of retraining the centralized model using user-specific data to build a personalized model for each user. Personalized FL is essential in EMS-related solutions because it enables the creation of tailored forecasting models that are able to capture user behavior and adapt the model to their patterns, providing more accurate and reliable predictions for individual users. The literature proposes different ways of achieving this:

- Retrain the global FL for a small number of epochs locally using the client's data exclusively [395].
- Retrain the global FL adding to the loss function a regularization penalty [381].
- Retrain the global FL through a layer-wise parameter aggregation strategy. Only the shallow layer parameters are uploaded to the central server and aggregated together with those from other utilities, while the deep layer parameters are kept locally [401].

This study applies the personalization technique proposed in [395], where the centralized model is retrained with only five epochs. Figure 5.2 depicts the steps followed by the personalized FL approach here proposed for HEMS cost-oriented load forecasting:

- **Step 1a.** A clustering of clients is incorporated before initiating the FL procedure [395]. The selected clustering criteria are the similarity of households' average daily consumption profiles and geographical proximity. This is due to the fact that clients' load profiles require to have similar patterns. Otherwise, the global federated model may diverge [15] because of data heterogeneity among end-users. In this thesis, clusters have been computed in a centralized manner. Then, **Step 1b** calculates the cost-oriented loss function of the desirable households and for the cluster, if needed. From this step onwards, the procedure for a single cluster is explained since the strategy is identical for all clusters.
- **Step 2.** In the first training round, once the deep neural network structure is specified, the central cloud server randomly initializes the global model weights; if not, the server proceeds with the weights obtained from the previous training round (Step 7).
- **Step 3.** The server sends a copy of the global model to clients within the same cluster, which develops its own model.
- **Step 4.** Each client trains the initial model using its local consumption data. Households receive a copy of the global model and train it using only the local data. On the other hand, spe-

cifics of the local training depend on the type of loss function approach selected, as explained in Section II.

- **Step 5.** Model weights updates are returned to the server for aggregation.
- **Step 6.** Central server aggregates the individual model updates from the same cluster.
- **Step 7.** Aggregated models are distributed back to their participants' respective clusters. The process repeats from Step 2 until convergence.
- **Step 8.** Personalization. It consists of retraining the global centralized FL model using user-specific data to build a personalized FL model for each client. This is achieved through retraining the model for a small number of epochs locally using the user's data exclusively [395]. Thus, this new personalized model is kept locally, so it never leaves the house, preserving personal data.

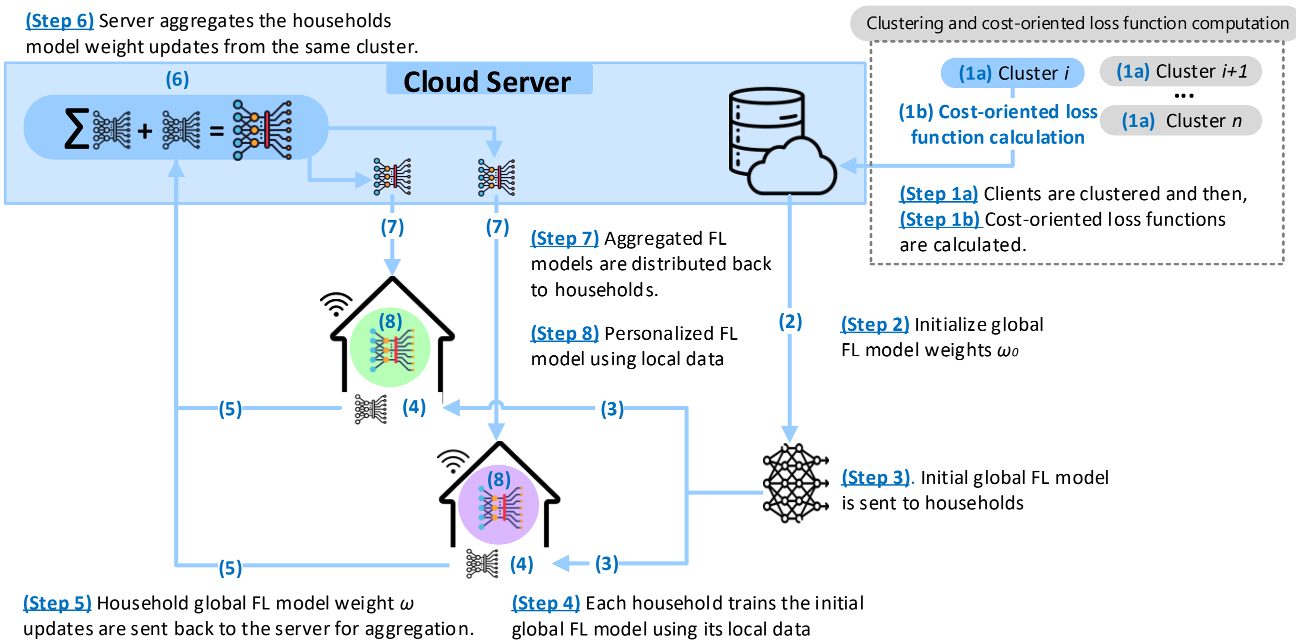

Figure 5.2: Personalized FL methodology for HEMS load forecasting. Cluster i example.

5.4 Imbalance cost-oriented load forecasting for Energy Management Systems

This section provides a detailed explanation of Steps 1a and 1b (clustering and cost-oriented loss function computation), shown in the proposed methodology in Figure 5.2. In Section 5.4.1, households are clustered based on their daily consumption patterns (Step 1a). The general methodology for calculating the cost functions is then described in Section 5.4.2. This methodology can be applied to individual houses as well as to an entire cluster (Step 1b).

5.4.1 Clustering households consumption profiles

In traditional machine learning algorithms, it is commonly assumed that all input data follows the principle of being Independent and Identically Distributed (i.i.d.), indicating that each data record in the dataset is independent and drawn from the same distribution. However, in the context of federated learning, where smart meter data is distributed across multiple smart meter devices, this assumption of i.i.d. data is inaccurate. Each household exhibits its unique consumption distribution due to evident variations in user behaviors and demographics. Consequently, models trained on one smart meter's data may not effectively generalize to another end-user due to differences in their data distributions.

For all the above-exposed reasons, an effective strategy to enhance the convergence and performance of aggregated models in federated learning is to cluster households with similar properties and consumption profiles. This approach, as suggested by [395], involves grouping together customers with similar i.i.d. data by clustering end-users according to their consumption profiles. Subsequently, further training is conducted using federated learning for each isolated cluster. Individual model updates provided by customers within the same cluster are then aggregated to create a specific global model tailored to that group of households.

Moreover, in the electricity sector smart meter data are owned by different retailers who may not be willing to share their data. In such situations, the consumer clustering approach cannot directly make use of the entire dataset. Previous research has focused on identifying social-demographic characteristics of electricity consumption [373, 394], and load clustering [376, 396]. However, this thesis presents a centralized approach to clustering clients using unsupervised learning techniques, as FL clustering is out of this thesis' scope.

5.4.2 Cost-oriented loss function calculation

Accurately predicting loads is an effective strategy for reducing imbalance costs associated with HEMS' performance. Both over and under-forecasts may result in additional costs due to imbalance penalizations. Currently, the majority of load forecast applications use the quadratic loss function $\mathcal{L}_q$. However, this quadratic approach is not ideal for cases where the costs of over-consumption and under-consumption have different economic impacts [404]. In literature, [384] presents a load forecast differentiable cost-oriented loss function by applying an optimal piecewise linear approximation method and the Huber norm embedding technique. The following study [385] develops a wind power generation model using a cost-oriented loss function. The quadratic loss function (5.1) is considered as follows:

$$\mathcal{L}_q(\epsilon) = (\hat{y} - y)^2 \tag{5.1}$$

where the forecasting error $= \hat{y} - y$, predictions are represented by $\hat{y}$ and the actual value is y.

Imbalance cost arises when forecasting error results in non-optimal decisions. Thus, the cost $C(\epsilon)$ is calculated by generating various HEMS scenarios of imperfect load forecast, quantifying the economic costs associated in (5.2).

$$C(\epsilon) = C(\hat{y}) - C(y) \tag{5.2}$$

The cost-oriented loss function $\mathcal{L}_{co}(\epsilon)$ has to satisfy the following three conditions [383]:

1) $\mathcal{L}_{co}(0) = 0$. There is no cost if the forecasting error ϵ is zero.

2) $min\ \mathcal{L}_{co}(\epsilon) = 0$. The cost must be greater or equal to zero.
3) $\mathcal{L}_{co}(\epsilon)$ is monotonic non-decreasing as ϵ moves away from zero, or, in other words, it is a convex function such as $f(x)'' \geq 0\ \forall x$.

Moreover, the imbalance cost function proposed for each household has the property that it is asymmetric, i.e., $\mathcal{L}_{co}(-\epsilon) \neq \mathcal{L}_{co}(\epsilon)$, since the associated economic impact of imperfect forecasts is asymmetric [404].

The cluster and individual household cost functions have been linearized based on the methodology presented in [384]. Thus, they are calculated as presented in (5.3):

$$\mathcal{L}_{co}(\epsilon) = \begin{cases} A\epsilon & \text{if } -\infty < \epsilon < 0 \\ B\epsilon & \text{if } 0 \leq \epsilon < \infty \end{cases} \tag{5.3}$$

When predicting individual customer load consumption in HEMS, it is important to think about the overall imbalance costs assumed by retailer companies due to deviations within their portfolio. To account for this associated cost resulting from errors in predictions, cost functions are constructed using end-users historical data. An analysis of the upward and downward imbalance of average prices is conducted using the average costs from the previous month. Cost-oriented loss functions are then used in Step 4 and Step 8 (refer to Figure 5.2). The HEMS costs of load forecasting errors are quantified by calculating the ideal cost with accurate forecasts and the actual cost with imbalance penalizations due to forecasting errors. This study assumes deterministic imbalance costs.

The methodology proposed for the cost-oriented FL load forecasting model, shown in Figure 5.3, is based on [384] and comprises the following phases:

- **Phase 1**: Cost-oriented loss function generation. Historical HEMS optimization results are used to associate load forecasting errors with economic imbalance costs.
- **Phase 2**: Cost-oriented loss function linear approximation. In this step, a linear cost function is formulated using the discrete imbalance cost and load forecast error data from Phase 1.

- **Phase 3**: Cost-oriented loss function integration in FL models. The cost-oriented loss function obtained in Phase 2 is integrated within the FL models. This loss function is used to train the global FL models and, most importantly, for personalizing the FL model using individual household cost functions in Step 8.

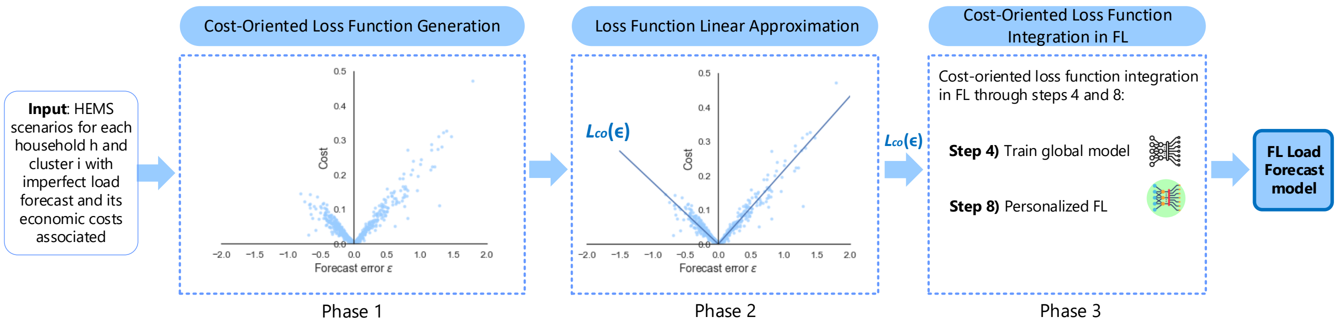

Figure 5.3: Cost-oriented loss function strategy for HEMS load forecasting.

The methodology presented in Figure 5.2 is utilized for training the HEMS load forecasting model in a distributed manner, with each household having its own personalized FL model. The resulting models are then utilized during the daily operation of each HEMS, as illustrated in Figure 5.4. The load forecast input data are fed into the FL model in order to predict the next day's load consumption. The day-ahead consumption prediction, along with other data, is then used for HEMS optimization. The HEMS output data is then used to create further cost functions based on these historical results.

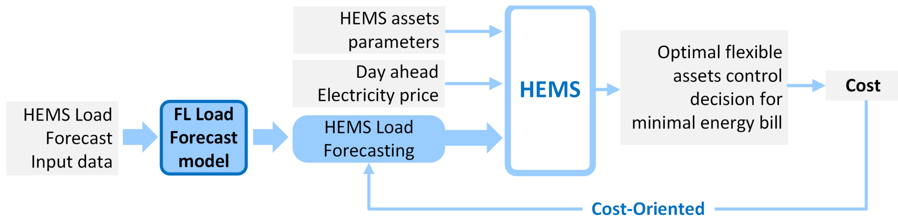

Figure 5.4: Cost-oriented loss function for HEMS load forecasting.

The details of the personalized FL procedure for HEMS load forecasting using a cost-oriented loss function are explained in Algorithm 1.

5.5 Case study and results

This section aims to verify the personalized FL effectiveness and economic benefit of the methodology proposed. For this purpose, we propose a study with actual household data using different cost functions in FL models in order to compare the results with traditional machine learning models trained locally.

5.5.1 Dataset Description and Experimental Settings

This study utilizes the publicly available "Smart Meter in London" dataset [405], consisting of a large amount of electricity consumption data gathered from 5,567 households in London between November 2011 and February 2014, with a 30-minute sampling interval. These

Algorithm 1 Personalized Federated Learning in Smart Homes using Cost-oriented loss functions.

Input: $\mathbf{D} = [D_1, D_2,...,D_N]$ historical data of N households within cluster C, individual cost-oriented loss function $L_{co}(\epsilon)_n$, federated iteration number I^{fl}, personalization iteration number I^p, central model weights Ω, local model weights ω.
Could server execution:
Initialize global model weights randomly Ω_0
for federated iteration $i = 1, 2..I^{fl}$ **do**
    $S_n \leftarrow$ random set of n households $\in C$
    Send Ω_0 weights to S_n households
    Client household execution:
    for each household $k \in S_n$ in parallel **do**
        Client household k updates local copy of $\omega_{i,k}$
        $\omega_{i+1,k} \leftarrow$ householdWeightsUpdate(k, ω_k)
        Client household k sends $\omega_{i+1,k}$ to **Server**
    end for
    Could server computes the global gradient:
    $\Omega_{i+1} \leftarrow \sum_{k \in K} \frac{n_k}{n} \omega_{i,k}$
    Cloud server updates the Ω_{i+1}
end for

Send global model Ω_{fl} to all households $\in C$ for personalization

$\omega_0 = \Omega_{ifl}$
Client household execution:
for personalization iteration i=1,2..I^p **do**
    Client household n personalizes ω_0 by minimizing the $L_{co}(\epsilon)_n$ on the D_n dataset for I^p epochs.
end for
Output: Cost-oriented personalized FL models ω_n

data are utilized for training load forecasting models for the next 24 hours, required as input in the HEMS optimization problem.

The objective function and constraints of the energy management system used in this study are based on previous work [406]. The objective function minimizes the electricity cost of purchasing power from the grid. All households are located in the same region (London) and it is assumed that they have inflexible PV generation and identical storage capacity. The evaluation period for this case study is one month (November 2013), using historical data backdated to this date to train the different forecasting models. The cost-oriented loss functions, both at household individual and cluster levels, are determined using the HEMS optimization cost results from the previous month. The features used for the load forecast include past consumption data from 1, 2, and 7 days before. A HFL structure is employed for this purpose.

The implementation of the deep neural network model is carried out using PyTorch [407], while federated learning is enabled by the PySyft library [408]. Table 5.2 presents the configuration of DNN models integrated with various loss functions.

Table 5.2: Hyperparameters of the deep learning model

Hyperparameter	**Value**
Number of hidden layers	2
Hidden layers neurons	(8,4)
Activation function	relu
Learning rate	0.01
Dropout	0.2
Momentum	0.95
Batch size	24

It should be noted that the deep learning feed-forward neural network hyperparameters analysis is not exhaustive, as it lies beyond the scope of this research. However, a brief justification for the chosen hyperparameter values is provided below.

The decision to set the number of hidden layers to 2 is based on the understanding that increasing the number of hidden layers enhances the model's complexity and its capability to capture hidden patterns within the data. This choice permits a balance between preserving simplicity and enabling the network to learn sophisticated patterns. The ReLU activation function is well-known for its effectiveness in deep neural networks. The learning rate determines the step size taken during each parameter update, affecting the convergence speed. A typical learning rate is 0.01, which enables a balance between faster convergence and stability. Dropout is a regularization technique that helps prevent overfitting by randomly disabling a fraction of neurons during training. A standard dropout rate is 0.2, which aids in preventing dependency on specific neurons and encourages the deep neural network to generalize better. Momentum is a parameter that affects the optimization process by adding a fraction of the previous update to the current update step. In deep learning, most practitioners set the momentum value between 0.9 - 0.99 without attempting to further tune this hyperparameter. Ultimately, the batch size determines the number of training examples processed in each iteration before updating the model parameters. In this case, 24 records are processed in each iteration.

5.5.2 Learning frameworks

This study introduces three learning frameworks to evaluate and compare day-ahead consumption forecast performance for households within the same cluster. The proposed methods are:

1. **Centralized Federated Learning (CFL)**: In this approach, a central server coordinates all participating households during the learning process. It selects random households within the same cluster at the beginning of the training process and aggregates the received model updates.
2. **Local Learning (LL)**: This approach employs state-of-the-art ML methods as a baseline for comparison. Each household has its local prediction model trained exclusively on its historical consumption data. Although this approach makes the model easily adaptable to the consumption pattern, a large amount of historical data is necessary to make the model reliable and robust. RF, MLP and MLR models are selected.

3. **Personalized Federated Learning (PFL)**: This approach is a combination of the previous methods. A centralized global model, common for each cluster, is trained first using a random set of households. Then, this model is sent to each user of that cluster to retrain the global model with their personal data, slightly modifying the weights of the CFL model to adapt it to each user and obtain better results. The PFL model is kept inside the end-user premises and should be periodically retrained, along with the CFL model.

The case study evaluates the federated learning and LL model approaches as illustrated in Figure 5.5. In Stage 1, a random set of households is selected to train and create the CFL model. In Stage 2, the trained CFL models are utilized to build personalized federated learning models in a different set of randomly selected households within the same cluster. A total of 13 models are evaluated, out of which 10 incorporate federated learning technology. To assess the effectiveness of the federated learning proposal, various combinations of loss functions are evaluated to determine the optimal strategy in terms of error and cost, as illustrated in Figure 5.5. To this end, different sets of households are used for training the CFL and for testing all FL and LL models. This results in the creation of thirteen scenarios, each representing a proposed model.

5.5.3 Clustering profiles and cost-oriented function generation

To cluster households based on the similarity of their hourly consumption profiles, the unsupervised learning algorithm K-Means [409] is employed. In this study, twenty different groups of end-users are created. For the sake of clarity, Figure 5.6a) shows only three clusters out of twenty to facilitate interpretation and reduce the number of lines in the chart. Each thin solid line represents the average hourly consumption of a household, with a different color depending on the group to which it belongs. The thick dashed line represents the average daily consumption of an entire cluster, which exhibits specific patterns in terms of shape and magnitude. In this study, cluster 4 is chosen to

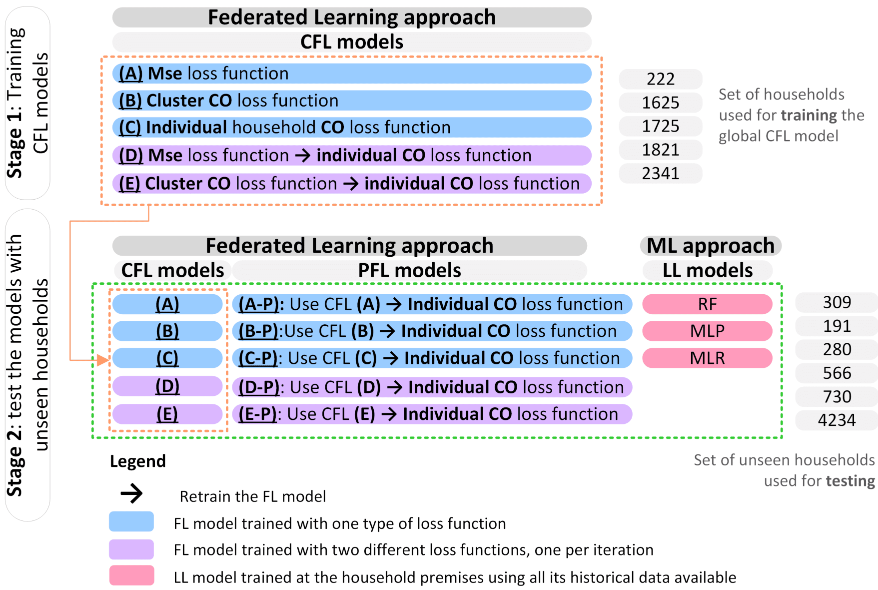

Figure 5.5: Case study: consumption prediction approaches (CFL, PFL, LL) with their respective loss function.

evaluate the proposed methodology.

In Figure 5.6b), the blue lines depict the average consumption of the houses used to train the CFL models, while the black lines correspond to the houses used to test the CFL, PFL, and LL models.

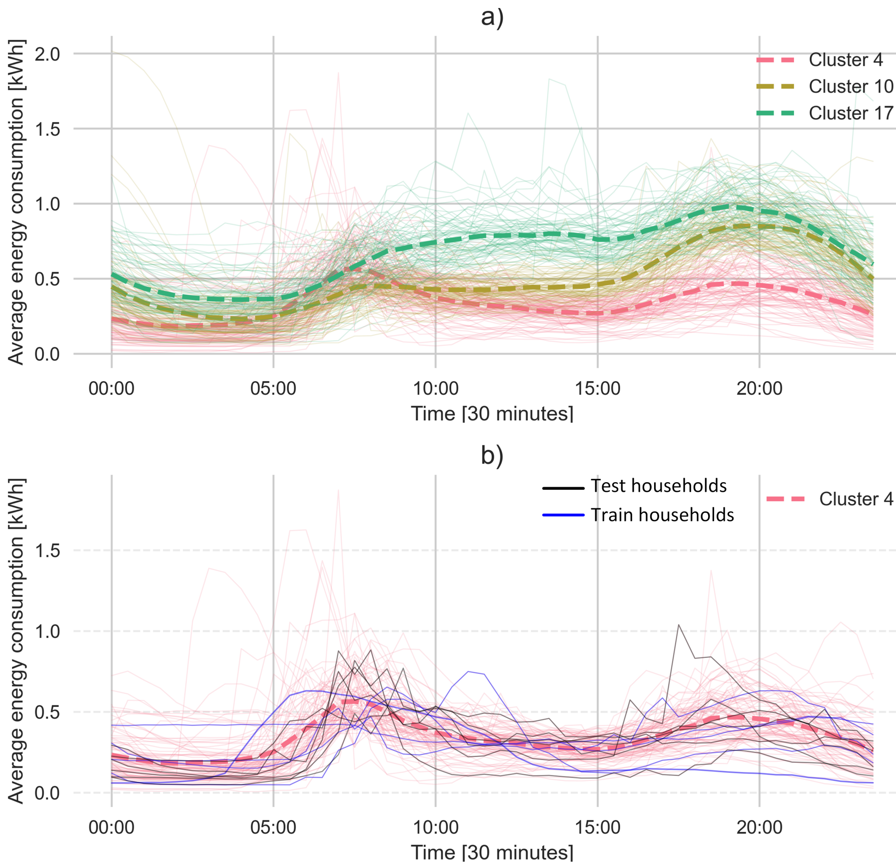

Figure 5.6: Clustering end-users' daily consumption patterns: a) displays three clusters and their associated consumption patterns and b) shows average daily consumption of the households selected for training and testing within cluster 4.

As shown in Step 1a) of Figure 5.2, once the clusters have been established, the next step is to calculate the cost-oriented functions using the cost-oriented methodology represented in Figure 5.3. Table

Table 5.3: Cost function coefficients A, B for training and testing households.

	ID used for training global FL models					ID used for testing FL models					
	222	1625	1725	1821	2341	191	280	309	566	730	4234
A	-0.166	-0.188	-0.185	-0.206	-0.199	-0.190	-0.181	-0.194	-0.200	-0.185	-0.185
B	0.215	0.208	0.218	0.231	0.208	0.208	0.216	0.233	0.237	0.197	0.212

5.3 displays the coefficients A and B resulting from Equation 5.3 for individual households. For the entire cluster 4, the coefficients $A = -0.1854$ and $B = 0.2034$ have been obtained, which are required for training the CFL (B) model using the general cost function.

5.5.4 HEMS load forecasting results

This subsection presents the results of the consumption estimations for November 2013. These predictions were used as input to run the HEMS for each individual household to calculate the error of the learning frameworks and their associated costs after optimization.

Table 5.4 displays the average performance in terms of error for each learning framework and household. As expected, since the measurement error metric is the MSE, models using loss functions with a quadratic format rather than cost functions show less error. The *diff* column represents how much more error in percentage each model has with respect to LL-MLP, which has the lowest error rate on average. On average, the LL obtains better error outcomes for two primary reasons. Firstly, these models are trained for each house with their historical data, which results in a better fit of the consumption pattern. Additionally, LL models use the quadratic function (5.1) as their loss function; therefore, the algorithms are trained to provide less error for this metric. In the case of CFL, the error rises significantly, specifically in those models where the quadratic loss function is not used but cost-oriented instead (B and E). Last but not least, PFL models have, on average, around 15% more MSE error than LL-MLP. This is because personalization is carried out by retraining the CFL models with individual end-users cost functions, thus prioritizing cost minimization

Table 5.4: Forecasting Mean Squared Error for each household tested.

		Mean Error (MSE)							
		191	**280**	**309**	**566**	**730**	**4234**	**Total**	% diff
CFL	**(A)**	0.566	0.110	1.057	0.209	0.617	0.151	2.713	9.38%
	(B)	0.595	0.143	1.111	0.225	0.656	0.144	2.877	15.98%
	(C)	0.558	0.112	1.062	0.205	0.614	0.149	2.703	8.97%
	(D)	0.559	0.110	1.055	0.209	0.606	0.151	2.692	8.55%
	(E)	0.618	0.146	1.142	0.226	0.657	0.139	2.929	18.11%
PFL	**(A-P)**	0.589	0.113	1.119	0.213	0.637	0.151	2.825	13.91%
	(B-P)	0.578	0.117	1.159	0.217	0.647	0.154	2.874	15.89%
	(C-P)	0.591	0.114	1.131	0.212	0.642	0.164	2.857	15.17%
	(D-P)	0.581	0.114	1.136	0.224	0.631	0.151	2.841	14.53%
	(E-P)	0.577	0.116	1.164	0.222	0.635	0.152	2.867	15.60%
LL	**RF**	0.532	0.112	0.981	0.210	0.539	0.130	2.507	1.06%
	MLP	0.503	0.102	1.012	0.202	0.531	0.129	2.480	0.00%
	MLR	0.530	0.109	1.032	0.209	0.598	0.137	2.617	5.52%

rather than error.

Figure 5.7 depicts a 48-hour extract of the load forecasting performance for households 566 and 4234, displaying the profiles of the models with the least error for each method. It is worth mentioning that the performance of the CFL model may vary depending on the consumption pattern of the test household, as it may not fit the generalization of the cluster model satisfactorily, as is the case in Figure 5.7 b). In such cases, personalization helps the global CFL model to obtain better results and adjust the predictions to the consumption pattern of the household.

The costs associated with the optimization of the HEMS using the learning frameworks proposed in subsection 5.5.4 for load estimation are summarized in Table 5.5. The results are presented in the same way as the previous table (Table 5.4), with the exception that this table displays the total costs for the houses used in the test case study. The *Total* column represents the sum of all costs, and *% diff* denotes the increase in cost relative to the most affordable model, which is the PFL-(E-P). The outcomes confirm that all PFL models result in lower costs than any other method, despite having the highest MSE error.

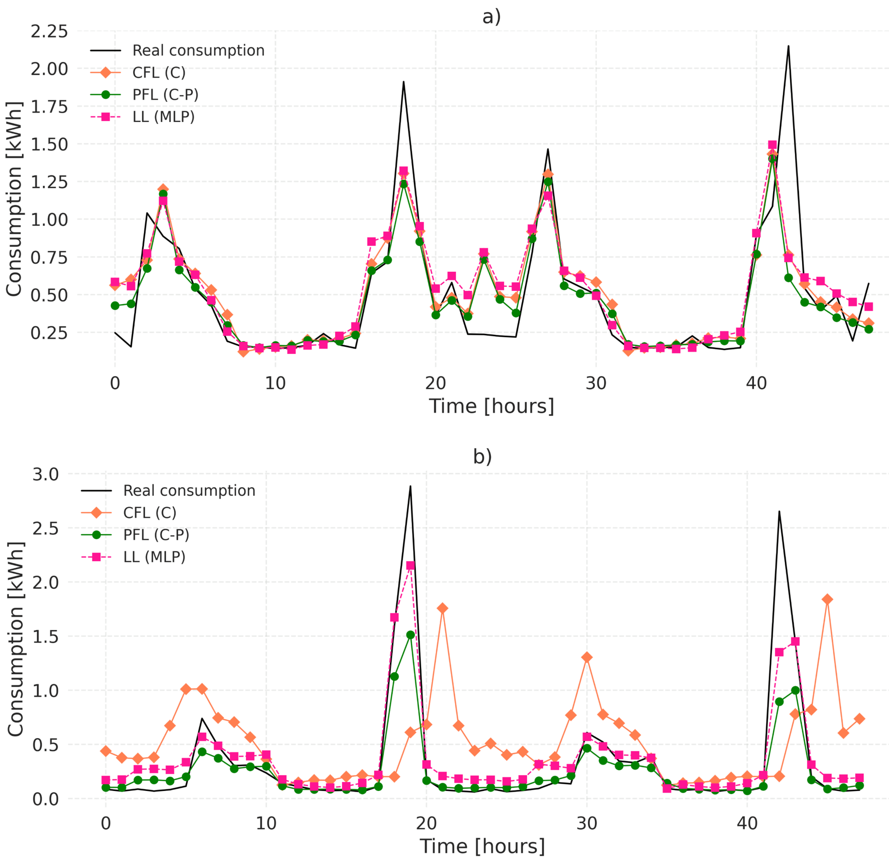

Figure 5.7: CFL, PFL and LL predictions for a) client 566 and b) client 4234, who did not participate in the CFL training model

Table 5.5: Evaluation cost of HEMS using different approaches.

		Cost (EUR)							
		191	**280**	**309**	**566**	**730**	**4234**	**Total**	**% diff**
CFL	**(A)**	179.42	137.15	232.93	128.54	166.87	89.98	934.90	3.40%
	(B)	176.36	142.59	231.88	125.48	164.84	85.12	926.27	2.45%
	(C)	176.35	137.60	231.25	125.75	164.32	86.10	921.37	1.90%
	(D)	177.64	137.87	232.34	127.81	164.93	90.56	931.15	2.99%
	(E)	177.69	144.71	232.73	125.88	164.84	81.65	927.51	2.58%
PFL	**(A-P)**	173.61	136.92	229.13	124.65	161.06	81.03	906.40	0.25%
	(B-P)	172.84	137.53	230.17	124.32	161.08	80.26	906.20	0.23%
	(C-P)	172.92	137.04	229.65	124.28	160.78	81.40	906.08	0.21%
	(D-P)	172.98	136.67	228.96	125.61	159.78	80.71	904.72	0.06%
	(E-P)	171.87	137.22	229.69	124.93	160.12	80.33	904.16	0
LL	**RF**	176.49	137.34	226.91	126.32	160.58	81.11	908.74	0.51%
	MLP	173.47	136.06	231.27	125.84	161.14	82.46	910.24	0.67%
	MLR	177.09	137.31	232.16	126.47	166.83	81.69	921.55	1.92%

Figure 5.8 provides a visual summary of the results presented in previous tables (Table 5.4 and Table 5.5). The graph shows the total cost of the tested houses in blue on the left axis and the sum of the obtained errors in gray on the right axis. The X-axis indicates the three different learning frameworks with their respective models.

It is worth noting that the accuracy of FL models, even after personalization, still varies depending on the user. This could be due to the fact that the consumption data of the households selected for training may have somewhat different consumption profiles, as seen in Figure 5.6. Ideally, this could be improved by using a larger number of clusters in the data set, resulting in a smaller number of users in each cluster with much more similar profiles. However, this increases complexity, so a middle ground must be found.

As expected, the PFL framework has the lowest total cost, but it also has the highest error compared to the other methods.

5.5.5 Sensitivity analysis

Suppose that a new household, whose consumption pattern aligns with cluster 4, wants to predict its energy consumption but has very few

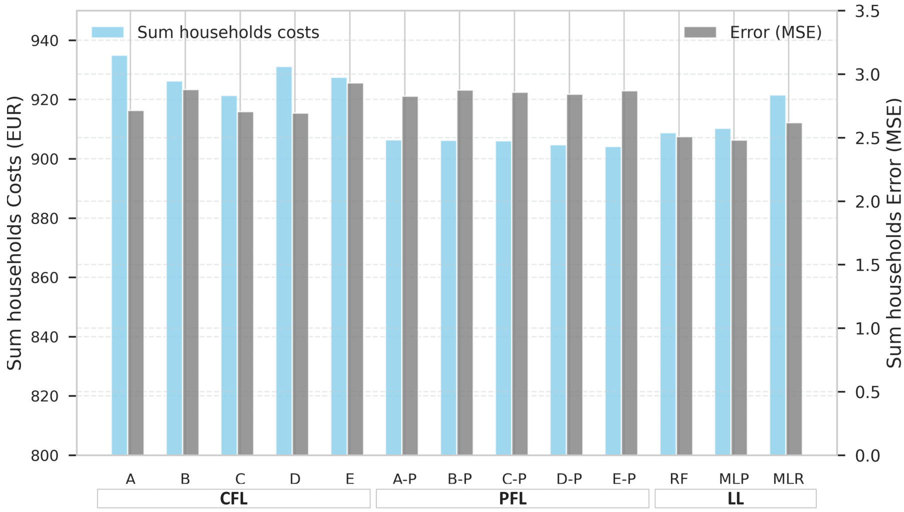

Figure 5.8: Case study results: Sum of households costs and error.

days of past consumption data available. In such a case, how would the three types of models analyzed in the preceding section perform in terms of error? To answer this question, we conducted a sensitivity analysis to examine how the MSE changes based on the number of past days that can be utilized for personalizing the FL model, and how the ML benchmark models perform using only these limited past days for training. It is noteworthy that the CFL model would not be affected by low data availability since it does not use historical data from the local test houses to predict energy consumption for unseen households.

To investigate how the amount of available historical data in each test household affects our framework and learning models, we conducted a sensitivity analysis by simulating the case study proposed in Figure 5.5 with varying numbers of historical days (d = 1, 3, 7, 14, 21, 30, 90, 365) to retrain PFLs and train LLs. Table 5.5 presents the average test households error of this analysis.

Figure 5.9 displays the performance of the CFL, LL, and PFL models under varying amounts of available historical data, with a logarithmic x-axis scale. For LL, there is a clear trend of improving error performance as the amount of available historical data increases, which aligns with

Table 5.6: Sensitivity analysis of the availability of historical data.

	Historical data used for personalization and LL training							
	1	3	7	14	21	30	90	365
Average CFL	0.349	0.349	0.348	0.347	0.349	0.348	0.348	0.3375
Average LL	0.453	0.405	0.398	0.382	0.366	0.366	0.355	0.3311
Average PFL	0.356	0.361	0.361	0.361	0.368	0.368	0.358	0.3546

the common intuition that more historical data yields better model learning. However, for federated learning models, the error remains relatively constant, indicating that CFL is capable of generalizing well. As expected, the MSE error for PFL is slightly higher than that for CFL, since PFL focuses on minimizing cost rather than quadratic error. With one month of historical data available, the local models exhibit an error rate similar to the FL approaches. After one year, the LL framework offers the most minor error rate, but this does not imply the least cost, as demonstrated in the previous Section 5.5.4.

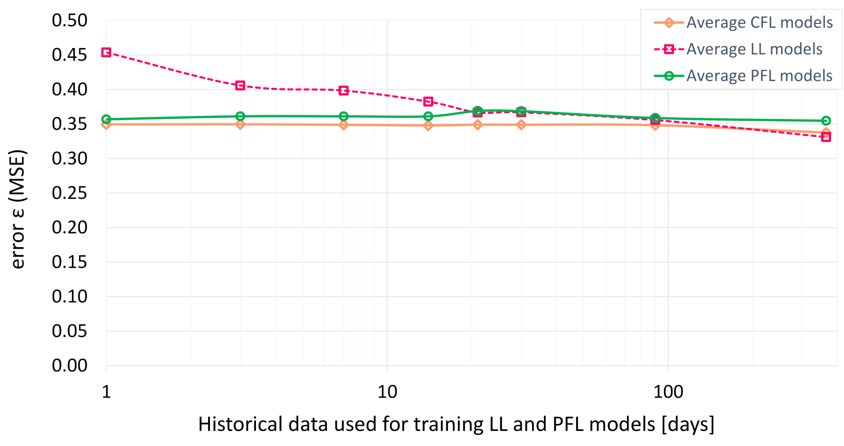

Figure 5.9: Sensitivity analysis for different historical data training.

In summary, this sensitivity analysis demonstrates that even with limited available data, reliable consumption predictions can be made for households within the same cluster.

5.6 Conclusions

This chapter presents a novel personalized federated learning methodology for home energy management systems, designed explicitly for demand forecasting. The PFL method incorporates cost-oriented loss functions, which minimize the risk of imbalance cost penalties for end-users, indirectly contributing to maintaining a balanced electricity grid. The performance of the proposed federated learning framework was evaluated and compared with local learning ML models. The results show that the PFL approach provides the lowest cost in all its models, compared to LL and CFL, saving on average more than 1% and 3%, respectively. Interestingly, PFL exhibits the largest average quadratic error rate, around 14% higher than the LL approach. Furthermore, the sensitivity study revealed that FL frameworks are essential for making reliable consumption predictions for houses with unknown historical data availability, as the FL frameworks maintain error performance regardless of the data available, unlike LL models. Additionally, LL models outperform the FL approach in terms of error after three months of past data available.

Despite the promising results of the proposed federated learning framework for home energy management systems, its implementation in energy systems faces several technical challenges related to data heterogeneity, improving privacy-preserving techniques, overcoming communication constraints, and exploring federated learning for real-time data processing. Data format and granularity collected from different households could be different, leading to data heterogeneity. To improve FL security and prevent attacks, privacy-preserving and cybersecurity strategies must be enhanced. Communication between edge devices and central servers must be optimized to minimize latency and bandwidth requirements. Implementing federated learning for real-time data processing is another crucial challenge for demand-response and grid management applications. Thus, the feasibility of the proposed method depends on the edge devices' capabilities to perform distributed local training, which may pose additional challenges in terms of computational power and storage capacity. Fortunately, recent advances in Information technology communications will unlock some of these current barriers, such as real-time communication.

In the context of deploying federated learning in a real environment, future studies in this area could explore hardware-in-the-loop investigations. This experimental approach involves integrating and testing real hardware within a simulation environment to evaluate its performance and behavior prior to real-world implementation. To conduct such an experiment, key steps must be followed. First, integrating hardware components is necessary to establish the federated learning setup. This involves ensuring compatibility, establishing communication protocols, and synchronizing data collection and transmission. Furthermore, it is crucial to implement a robust data security framework to protect sensitive consumer data during the hardware-in-the-loop experiment. This framework should incorporate encryption techniques and other privacy measures to ensure data privacy and security. Scalability is another essential concern to consider in terms of efficient computational resources. When transitioning to a real-life environment, scalability constraints become even more significant as the experiment is conducted with a larger number of participants or households and, therefore, devices.

In summary, conducting a hardware-in-the-loop experiment would provide researchers with valuable insights into the performance, behavior, and scalability of FL. This investigation plays a crucial role in ensuring the proper functioning of FL approach prior to actual deployment.

Future work could explore incorporating a cybersecurity layer into this approach to enhance its safety against potential attacks. Different deep neural network architectures, such as LTSM -designed explicitly for time-series estimations- can be analyzed to improve model performance further and reduce costs. Additionally, to capture the time-dependency of cost functions, time-varying cost-oriented loss functions could be investigated.

Chapter 6

Conclusions

This chapter presents the main findings of this thesis in response to the research questions posed in Chapter 1, Section 1.2. It also provides a summary of the contributions and proposes future lines of research.

6.1 Answers to Research Questions

This thesis addresses five research questions, which have been answered throughout the studies presented in the chapters of the thesis.

> **RQ1)** What novel data-driven energy services are likely to emerge or could benefit from the vast daily amount of operational and non-operational data related to distribution networks?

To identify the emerging or enhancing data-driven energy services resulting from the digitization of the distribution network, an exhaustive state-of-the-art study was conducted in Chapter 2. The primary energy-related services found were categorized into three major groups:

- **Distribution grid operation**. Responsible for ensuring the correct functioning of the distribution network. This category includes real-time operation and monitoring services, allowing real-time network state observability. Fraud detection and predictive maintenance of assets involved in the low-voltage network are also improved thanks to continuous sensorization and measurement of their condition.

- **Flexibility management**. Responsible for the flexibility market. This category includes prediction models for both generation and consumption, as well as available flexibility. Energy management systems and aggregated flexibility services for DSOs (to avoid network congestion) or BRPs (to avoid deviations in their portfolio) are also benefiting from digitization and high data availability. P2P trading, which aims to find a reliable way to exchange energy between different customers, local energy communities, and operators, is enabled thanks to Distributed Ledger Technology.
- **Planning**. Responsible for optimal investment strategies that contribute to long-term planning in the distribution grid. This category examines grid status and expansion criteria, selects the most appropriate technologies and optimal geographical locations, and contributes to grid support during a settled planning horizon. Mid- and long-term forecasting plays a crucial role in this category, estimating the associated costs for achieving specific planning goals or criteria while meeting the forecasted demand.

These innovative data-driven energy services are offered to a broad range of stakeholders involved in the energy domain, including DSOs, BRPs, prosumers, and aggregators. The purpose is to improve their performance and encourage the creation of novel business models in the energy sector to take advantage of the massive data being generated and underused. The scope of this thesis focuses on the category of flexibility management, specifically on intelligent energy management systems.

RQ2) Which Artificial Intelligence techniques are utilized in the development of these data-driven energy services in the distribution network, and how do they contribute to enhancing the sustainability, efficiency, and reliability of the system?

Once the innovative services have been identified, an exhaustive review of the most recent studies implementing AI techniques in each is carried out in Chapter 2. Key findings state that ensemble models present better results than single ML models by combining different data-driven algorithms. In recent years, deep learning algorithms have gained importance for time series prediction tasks and outperform most benchmark ML and statistical algorithms. Concerning classification tasks, traditional ML algorithms such as SVM or RF yet provide excellent results. For instance, the RF classifier outperforms when it comes to supervised classification tasks, while LSTM recurrent network is the predominant algorithm for time series forecasting. Unsupervised learning methods are mainly responsible for customer segmentation, building efficiency clustering and consumption profile grouping for non-technical losses detection. Finally, RL is widely applied in the literature to schedule flexible household assets optimally. However, the scarcity of physical experimentation in a realistic environment prevents its application in real-world buildings and households.

The main benefits of implementing AI methods within the distribution grid domain are addressed:

- Allow real-time distribution grid status estimation and gain observability, enhancing the monitoring and locating possible events in the network to provide a tool that enables the operator to react more rapidly when a fault or event occurs.
- Performing the predictive maintenance service increases the distribution network security and availability while diminishing the DSO costs.
- Use the aggregated flexibility available to avoid grid congestion.
- Optimal and reliable electricity trading among customers.
- Optimal medium and long term distribution grid planning

To conclude, equipping the distribution network with sensors to collect data is crucial for developing innovative energy services. Implementing data-driven techniques in energy service development is essential for creating a reliable, secure, and efficient Smart Grid. These methods have been shown to provide better performance than statistical benchmark procedures. However, challenges still need to be addressed to extend and improve AI applications in power systems. These challenges mainly include ICT infrastructure, data collection and governance, data integration and sharing, data processing and analysis, security and privacy, and the need for professionals in Big Data analytics.

> **RQ3)** How can electricity consumers and local energy communities contribute to enhancing sustainability through emissions reduction, and how can they be engaged and incentivized?

In a recent survey conducted by the EU, 90% of European citizens agreed that CO_2 emissions must be reduced to achieve the 2050 emissions neutrality target [14]. This motivation led to the development in Chapter 3 of a novel environmental-based HEMS strategy focused on minimizing **household**-related emissions. Then, this strategy evolved into a hybrid-based approach, combining GHG and electricity cost minimization.

- Environmental-based strategy. The key idea behind the novel environmental-based HEMS strategy is to empower end-users to actively participate in reducing GHG emissions by leveraging flexible assets such as batteries and controllable loads. The HEMS activates flexibility and adjusts end-users consumption based on the GHG emissions associated with the generation source used to meet their demand. This source can be self-consumption, such as batteries or photovoltaic panels installed behind the meter or from the electricity grid network. However, GHG emissions alone cannot represent environmental performance since they only account for emissions during the operational life. Instead, the Life Cycle Assessment (LCA) framework is employed in this thesis, which considers emissions from all processes associated with each generation technology, including raw material extraction, manufacturing, operation, and end-of-life. To this end, the Global Warming Potential (GWP) indicator is used as a reference in this work, which measures the CO_2 equivalent per kWh generated by each generation source.

However, the environmental-based strategy alone does not address the complex trade-offs between cost and emissions. Therefore, a hybrid-based approach was developed.

- Hybrid-based HEMS. It is a multi-objective strategy that combines emissions with cost reduction. This approach seeks to achieve the best trade-off between electricity cost and GHG emissions minimization, considering the life cycle analysis of the generation sources used to meet household demand.

The hybrid-based approach enables the end-user to modify the priority of their objective function in a fast and flexible manner by assigning weights to the multi-objective function. The end-user can then decide whether to prioritize minimizing the cost, reducing the environmental impact of their electricity consumption or a trade-off between these two objectives. This makes the HB HEMS an excellent option to encourage end-users to be engaged and participate in the fight against climate change without incurring high economic expenses.

After developing the environmental strategy at the consumer level, the concept is expanded to energy communities. For this purpose, a local energy management system has been created, consisting of a battery and centralized photovoltaic generation, the generation of which is distributed in an agreed manner among the members of the community. Together, they can agree to prioritize the reduction of emissions and minimize the total cost of the electricity bill.

The LEC optimization results demonstrate the feasibility and satisfactory performance of the proposed approaches. The price-based optimization achieves approximately 23% savings compared to the LEC baseline consumption. Similarly, the environmental strategy reduces about 6% of GHG emissions associated with consumption compared to the baseline. Furthermore, the environmental strategy also reduces costs by 18%, making it a sustainable and affordable LEC optimization choice. One significant finding is that buying energy directly from the grid could be more sustainable than using the battery with the surplus energy from the PV generator in case of low penetration of CO_2 generation in the energy mix.

> **RQ4)** Given the need to protect user data privacy, which machine learning distributed method can be applied to train energy management system prediction models without compromising personal data?

In recent years, federated learning, a machine learning method also known as collaborative learning, has gained notoriety for training prediction models involved in EMS methodologies in a distributed way without compromising personal data. This ensures that end-users information remains on their devices, ensuring the privacy and security of their data. The central server only receives the prediction model weights aggregated from several households, which is insufficient to infer personal data about individual users.

However, the challenge of creating a unique global ML model common to a group of customers is that the global model may not fit all the customers' patterns. Chapter 5 proposes personalization in federated learning as a solution to this problem in order to understand user behavior and adapt to it. It consists of retraining the centralized model using user-specific data to build a personalized model for each user. Personalized FL is essential in EMS-related solutions because it enables the creation of tailored forecasting models that are able to capture user behavior and adapt the model to their patterns, providing more accurate and reliable predictions for individual users.

RQ5) How can cost-oriented loss functions in EMS prediction models help to minimize imbalance penalizations due to energy consumption estimation errors?

Currently, the majority of load forecast applications use the quadratic loss function. However, it is not a suitable approach when over-consumption and under-consumption have different economic impacts, like deviation penalizations. In such cases, the prediction model should be trained to minimize the expected cost of imbalances rather than just the mean squared error (MSE).

To address this issue, this thesis proposes in Chapter 5 an innovative federated learning methodology for Home Energy Management Systems (HEMS). The presented approach incorporates cost-oriented loss functions when training demand forecasting models. This thesis tests different combinations and types of loss functions for both global and personalized federated learning models. By minimizing the expected cost of imbalances, the HEMS can better manage the risks associated with inaccurate energy consumption predictions, which may lead to costly penalties from energy suppliers. Additionally, cost-oriented loss functions can optimize the trade-off between cost and accuracy, allowing the HEMS to make more informed decisions about how to balance energy supply and demand. This methodology, tested using actual household data, has demonstrated that using cost-oriented loss functions achieves the lowest costs associated with imbalances. Additionally, the study has proven that global federated learning models enable households with similar consumption patterns but without extensive historical consumption data to achieve accurate load prediction outcomes through collaborative learning models.

6.2 Contributions

The major contributions for each chapter are summarized below:

- In **Chapter 2**, the AI methods with potential applications in distribution grid energy services are analyzed. These applications include operation and monitoring, predictive maintenance, non-technical losses detection, forecasting, flexibility management and planning of distribution grids. In addition, the relationships between the distribution grid applications are presented and mapped, proving that multiple services can be offered as a single clustered service to different stakeholders. Furthermore, future opportunities and challenges for the application of AI in distribution grids are identified.
- **Chapter 3** introduces an energy management system for a household and proposes a novel strategy to optimize the system operation by minimizing the greenhouse gas emissions associated with user consumption. To achieve this goal, the emissions associated with the generation sources used are quantified hourly, including those from self-consumption sources such as batteries and photovoltaics, as well as emissions associated with the hourly energy mix of the national electrical system, which varies depending on the availability of renewable sources. Furthermore, a degradation battery model is implemented in the HEMS, which takes into account calendar and cycling aging constraints. Then,

a hybrid-based strategy that combines environmental and cost-minimization strategies is proposed. The users give the weight that they consider appropriate to the approach that suits them best, and they can change it whenever they want. The proposed hybrid-based strategy enables the end-users to have a more influential role in the climate change solution by giving more weight to the multi-objective function's environmental component.

In line with the latter challenge, federated learning strategies that could enable the implementation of some of the identified data-driven services without compromising data privacy and security can be identified and compared to centralized learning approaches. On the other hand, potential data-driven services of interest for transmission grids could be investigated.

- **Chapter 4** extends the previous methodology to the scenario of a local energy community, which includes a centralized PV system and a battery. Thus, energy communities can offer not only energy and economic savings but also an opportunity to encourage and involve users in the fight against climate change. Price and environmental-based LEC optimization strategies are explored, resulting in promising outcomes for reducing electricity expenses and emissions.

- In **Chapter 5**, for the first time, a cost-oriented loss function is integrated with personalized federated learning (FL) models for load forecasting in HEMS. This approach ensures the protection of consumers' data privacy by avoiding the sharing of their smart meter information. The methodology demonstrates and validates the potential cost savings associated with imbalances through the use of cost-oriented personalized FL models in HEMS for end-users with diverse availability of historical data. Furthermore, this thesis estimates accurate day-ahead consumption profiles and imbalance cost savings for households without historical data using federated models.

6.3 Future work

Potential future research directions based on the contributions of this thesis are described by chapter:

- **Chapter 2**. The energy sector is increasingly interested in utilizing data science and AI capabilities to overcome their daily challenges However, despite the potential benefits, the application of big data methods in the energy sector is not yet fully matured. This brings new opportunities and future research directions for this emerging and promising research area. Furthermore, with AI technology constantly evolving at an ever-increasing speed, there is a continuous need for researchers to adapt the current state of the art to these recent technologies. In addition, the expected increase in data availability in the future will result in even greater accuracy and possibilities for energy services. However, to fully boost the potential of the distribution grid data, it is crucial to address some challenges, including IT infrastructure, data collection and governance, data integration and sharing, data processing and analysis, security and privacy, and the need for professionals with expertise in big data and AI analytics.
- **Chapter 3**. The price, environment and hybrid-based strategies proposed in Chapter 3 for HEMS demonstrated promising results in reducing costs and GHG emissions without incurring high expenses. However, further investigation is necessary to understand how this strategy can be applied on a larger scale. Future research could explore the implementation of hybrid-based energy management systems in larger communities or industrial settings, utilizing more flexible assets and incorporating external flexibility requests. While the thesis primarily focused on energy management systems with PV generation and storage units, other flexible assets could be integrated into the proposed energy systems to further optimize energy use. For instance, integrating electric vehicle charging stations and flexible loads.
- **Chapter 4**. Similar to the findings in Chapter 3, the price-based and environment-based strategies proposed in Chapter 4 for local energy communities have demonstrated the potential for

reducing overall LEC energy costs and GHG emissions without incurring high expenses. To further improve these strategies, future research could incorporate dynamic sharing coefficients that vary hourly and have different values on each day. Moreover, adding flexible elements within each building belonging to the energy community could optimize their behavior alongside the centralized battery. Additionally, distributed federated learning techniques could be analyzed for LEC forecasting tasks, such as the one presented in Chapter 5. Furthermore, exploring the participation of a LEC in a Flexibility Market, where it can offer its flexibility in exchange for economic remuneration, could be another future research path.

- **Chapter 5**. The personalized federated learning models proposed in Chapter 5 achieved accurate load forecasts while preserving consumers' privacy. However, privacy-preserving federated learning methods still require further development. Future research could investigate the incorporation of a cybersecurity layer to enhance the safety of the approach against potential attacks. Different deep neural network architectures, such as LTSM, which are explicitly designed for time-series estimations, could be explored to further improve the performance of federated learning models while maintaining strong privacy data guarantees. Furthermore, to capture the time-dependency of cost functions, researchers could investigate the use of time-varying cost-oriented loss functions.

Bibliography

[1] E. Commission and D.-G. for Climate Action, *Going climate-neutral by 2050 : a strategic long-term vision for a prosperous, modern, competitive and climate-neutral EU economy.* Publications Office, 2019. xv, 4, 5

[2] H. Ritchie, M. Roser, and P. Rosado, "Co and greenhouse gas emissions," *Our World in Data*, 2020, https://ourworldindata.org/co2-and-greenhouse-gas-emissions. xv, 5, 6

[3] E. Commission, "Driving forward the green transition and promoting economic recovery through integrated energy and climate planning," https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0564&from=EN, 2020, (Accessed on 04/08/2023). xv, 5, 7

[4] G. Shobha and S. Rangaswamy, "Computational analysis and understanding of natural languages: Principles, methods and applications Chapter 8 - Machine learning," *Handbook of Statistics*, vol. 38, pp. 197–228, 2018. xv, 21, 22

[5] P. Systems and M. Koivisto, *Finding Value in Big Data : Statistical Analysis inding Value in Big Data*, no. January, 2016. xv, 22

[6] "INVADE H2020: Deliverable D5.4, distributed and centralised energy storage in the distribution grid," Tech. Rep. 731148, 2020. xvi, 86

[7] REScoop EU., "Annual report 2021," Tech. Rep., 2022. xvi, 106

[8] R. Turconi, A. Boldrin, and T. Astrup, "Life cycle assessment (LCA) of electricity generation technologies: Overview, comparability and limitations," *Renewable and Sustainable Energy Reviews*, vol. 28, DOI 10.1016/j.rser.2013.08.013, pp. 555–565, 2013. xix, xx, 75, 76, 78, 109

[9] CEER, "Regulatory Aspects of Self- Consumption and Energy Communities CEER Report," no. June, p. 53, 2019. xx, 104

[10] L. V. Thomas, O. Schmidt, A. Gambhir, S. Few, and I. Staffell, "Comparative life cycle assessment of lithium-ion battery chemistries for residential storage," *J. Energy Storage*, vol. 28, DOI 10.1016/j.est.2020.101230, no. January, p. 101230, 2020. xx, 109

[11] U. Nations, "Paris agreement," Tech. Rep., 2015. 3

[12] The European Commission, "The European Green Deal," *Solar Power*, DOI 10.2307/j.ctvd1c6zh.7, p. 24, 2019. 4, 101

[13] IRENA, *Power system flexibility for the energy transition*, no. December, 2018. [Online]. Available: https://www.irena.org/publications/2018/Nov/Power-system-flexibility-for-the-energy-transition 4

[14] European Commission, *Special Eurobarometer 513. Climate Change*, no. April, 2021. [Online]. Available: https://europa.eu/eurobarometer/surveys/detail/2273 8, 162

[15] N. Good, K. A. Ellis, and P. Mancarella, "Review and classification of barriers and enablers of demand response in the smart grid," *Renewable and Sustainable Energy Reviews*, vol. 72, DOI 10.1016/j.rser.2017.01.043, no. January 2016, pp. 57–72, 2017. 19

[16] A. A. Khan, S. Razzaq, A. Khan, K. Fatima, and Owais, "HEMSs and enabled demand response in electricity market: An overview," *Renewable and Sustainable Energy Reviews*, vol. 42, DOI 10.1016/j.rser.2014.10.045, pp. 773–785, 2015. 19

[17] M. H. Rehmani, M. Reisslein, A. Rachedi, M. Erol-Kantarci, and M. Radenkovic, "Integrating Renewable Energy Resources into the Smart Grid: Recent Developments in Information and Communication Technologies," *IEEE Transactions on Industrial Informatics*, vol. 14, DOI 10.1109/TII.2018.2819169, no. 7, pp. 2814–2825, 2018. 19

[18] B. P. Bhattarai, S. Paudyal, Y. Luo, M. Mohanpurkar, K. Cheung, R. Tonkoski, R. Hovsapian, K. S. Myers, R. Zhang, P. Zhao, M. Manic, S. Zhang, and X. Zhang, "Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions," *IET Smart Grid*, vol. 2, DOI 10.1049/iet-stg.2018.0261, no. 2, pp. 141–154, 2019. 19

[19] A. Ghosal and M. Conti, "Key Management Systems for Smart Grid Advanced Metering Infrastructure: A Survey," *IEEE Communications Surveys and Tutorials*, vol. 21, DOI 10.1109/COMST.2019.2907650, no. 3, pp. 2831–2848, 2019. 19

[20] T. Alquthami, A. M. Alsubaie, and M. Anwer, "Importance of Smart Meters Data Processing – Case of Saudi Arabia," DOI 10.1109/icecta48151.2019.8959739, pp. 1–5, 2020. 19

[21] T. Alquthami, A. Alamoudi, A. M. Alsubaie, A. Bin, J. Nassir, M. Anwar, and S. Al, "Analytics framework for optimal smart meters data processing," *Electrical Engineering*, DOI 10.1007/s00202-020-00949-0, 2020. 19

[22] "Front Matter," *Big Data Application in Power Systems*, DOI 10.1016/b978-0-12-811968-6.09985-x, pp. i–ii, 2018. 20, 23, 28, 29

[23] K. Zhou, C. Fu, and S. Yang, "Big data driven smart energy management: From big data to big insights," *Renewable and Sustainable Energy Reviews*, vol. 56, DOI 10.1016/j.rser.2015.11.050, no. 2016, pp. 215–225, 2016. 20, 66, 67

[24] T. International Renewable Energy Agency, "Artificial Intelligence and Big Data Innovation Landscape Brief," 2019. 20, 21, 67, 69

[25] R. Gupta, S. Tanwar, S. Tyagi, and N. Kumar, "Machine Learning Models for Secure Data Analytics: A taxonomy and threat model," *Computer Communications*, vol. 153, DOI 10.1016/j.comcom.2020.02.008, no. November 2019, pp. 406–440, 2020. 21

[26] Z. M. Fadlullah, F. Tang, B. Mao, N. Kato, O. Akashi, T. Inoue, and K. Mizutani, "State-of-the-Art Deep Learning: Evolving Machine Intelligence Toward Tomorrow's Intelligent Network Traffic Control Systems," *IEEE Communications Surveys and Tutorials*, vol. 19, DOI 10.1109/COMST.2017.2707140, no. 4, pp. 2432–2455, 2017. 21

[27] D. S. Boning, I. A. M. Elfadel, and X. Li, "A Preliminary Taxonomy for Machine Learning in VLSI CAD," in *Machine Learning in VLSI Computer-Aided Design*, I. A. M. Elfadel, D. S. Boning, and X. Li, Eds., pp. 1–16. Cham: Springer International Publishing, 2019. [Online]. Available: https://doi.org/10.1007/978-3-030-04666-8{_}1 21

[28] M. S. Ibrahim, W. Dong, and Q. Yang, "Machine learning driven smart electric power systems: Current trends and new perspectives," *Applied Energy*, vol. 272, DOI 10.1016/j.apenergy.2020.115237, no. May, p. 115237, 2020. 21

[29] R. Sutton and A. Barto, "Reinforcement Learning: An introduction," *The MIT Press*, DOI 10.4018/978-1-60960-165-2.ch004, pp. 63–80, 2015. 21

[30] M. Glavic, R. Fonteneau, and D. Ernst, "Reinforcement Learning for Electric Power System Decision and Control: Past Considerations and Perspectives," *IFAC-PapersOnLine*, vol. 50, DOI 10.1016/j.ifacol.2017.08.1217, no. 1, pp. 6918–6927, 2017. 21

[31] A. T. Perera and P. Kamalaruban, "Applications of reinforcement learning in energy systems," *Renewable and Sustainable Energy Reviews*, vol. 137, DOI 10.1016/j.rser.2020.110618, no. December 2020, p. 110618, 2021. 21

[32] L. Duan, J. Hu, G. Zhao, K. Chen, S. X. Wang, and J. He, "Method of inter-turn fault detection for next-generation smart transformers based on deep learning algorithm," *High Voltage*, vol. 4, no. 4, pp. 282–291, 2019. 22

[33] Y. Li, F. Yu, Q. Cai, K. Yuan, R. Wan, X. Li, M. Qian, P. Liu, J. Guo, J. Yu, T. Zheng, H. Yan, P. Hou, Y. Feng, S. Wang, and L. Ding, "Image fusion of fault detection in power system based on deep learning," *Cluster Computing*, vol. 22, DOI 10.1007/s10586-018-2264-2, 2019. 22

[34] A. Moradzadeh and K. Pourhossein, "Short Circuit Location in Transformer Winding Using Deep Learning of Its Frequency Responses," in *2019 International Aegean Conference on Electrical Machines and Power Electronics (ACEMP) 2019 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM)*, pp. 268–273, 2019. 22

[35] Y.-Y. Hong and C. L. P. P. Rioflorido, "A hybrid deep learning-based neural network for 24-h ahead wind power forecasting," *Applied Energy*, vol. 250, DOI https://doi.org/10.1016/j.apenergy.2019.05.044, pp. 530–539, 2019. 22

[36] M. Abdel-Nasser and K. Mahmoud, "Accurate photovoltaic power forecasting models using deep LSTM-RNN," *Neural Computing and Applications*, DOI 10.1007/s00521-017-3225-z, 2017. 22

[37] A. Brusaferri, M. Matteucci, P. Portolani, and A. Vitali, "Bayesian deep learning based method for probabilistic forecast of day-ahead electricity prices," *Applied Energy*, vol. 250, DOI https://doi.org/10.1016/j.apenergy.2019.05.068, pp. 1158–1175, 2019. 22

[38] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2014. 22

[39] D. Bzdok, N. Altman, and M. Krzywinski, "Statistics versus Machine Learning," *Nature Methods*, vol. 14, pp. 232–234, 2018. 22, 23

[40] P. Bruce and A. Bruce, *Practical statistics*, vol. 3, 2017. 22

[41] Baron-Cohen Simon, *Practical statistics*, vol. 2, no. 3, 1989. 22

[42] Y. Zhang, T. Huang, and E. F. Bompard, "Big data analytics in smart grids: a review," *Energy Informatics*, vol. 1, DOI 10.1186/s42162-018-0007-5, no. 1, pp. 1–24, 2018. 23

[43] A. E. Saldaña-gonzález, A. Sumper, and M. Aragüés-peñalba, "Advanced Distribution Measurement Technologiesand Data Applications for Smart Grids : A Review," *Energies*, DOI 10.3390/en13143730, pp. 1–34, 2020. 23

[44] A. Dehghan-Banadaki, T. Taufik, and A. Feliachi, "Big data analytics in a day-ahead electricity price forecasting using tensorflow in restructured power systems," *Proceedings - 2018 International Conference on Computational Science and Computational Intelligence, CSCI 2018*, DOI 10.1109/CSCI46756.2018.00207, pp. 1065–1069, 2018. 23, 50

[45] U.S. Department of Energy, "Advanced Metering Infraestructure and Customer Systems - Results from the smart grid investment grant program," *Office of Electricity Delivery and Energy Reliability*, pp. 1–98, 2016. 24

[46] A. Shahsavari, M. Farajollahi, E. M. Stewart, E. Cortez, and H. Mohsenian-Rad, "Situational Awareness in Distribution Grid Using Micro-PMU Data: A Machine Learning Approach," *IEEE Transactions on Smart Grid*, vol. 10, DOI 10.1109/TSG.2019.2898676, no. 6, pp. 6167–6177, 2019. 24

[47] J. Ma, J. C. Cheng, F. Jiang, W. Chen, M. Wang, and C. Zhai, "A bi-directional missing data imputation scheme based on LSTM and transfer learning for building energy data," *Energy and Buildings*, vol. 216, DOI 10.1016/j.enbuild.2020.109941, 2020. 27

[48] S. Jung, J. Moon, S. Park, S. Rho, S. W. Baik, and E. Hwang, "Bagging Ensemble of Multilayer Perceptrons for Missing Electricity Consumption Data Imputation," *Sensors*, vol. 20, no. 1772, pp. 1–16, 2020. 27

[49] I. Spinelli, S. Scardapane, and A. Uncini, "Missing data imputation with adversarially-trained graph convolutional networks," *Neural Networks*, vol. 129, DOI 10.1016/j.neunet.2020.06.005, pp. 249–260, 2020. 27

[50] J. Qiao, B. Zhu, X. Wang, and K. Luo, "Application research of artificial intelligence technology in error diagnosis of electric energy meter," *2019 IEEE 4th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2019*, DOI 10.1109/ICCCBDA.2019.8725622, pp. 155–159, 2019. 28

[51] R. Arghandeh, M. Gahr, A. Von Meier, G. Cavraro, M. Ruh, and G. Andersson, "Topology detection in microgrids with micro-synchrophasors," *IEEE Power and Energy Society General Meeting*, vol. 2015-Septe, DOI 10.1109/PESGM.2015.7286053, 2015. 29, 31

[52] C. Vitiello, S, Vasiljevksa, J, Filiou, "Cost-benefit analyses & state of play of smart metering deployment in the EU-27," Tech. Rep., 2014. 29

[53] Y. Zhao, J. Chen, and H. V. Poor, "Efficient neural network architecture for topology identification in smart grid," *2016 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2016 - Proceedings*, DOI 10.1109/GlobalSIP.2016.7905955, pp. 811–815, 2017. 29, 30, 31

[54] R. Sevlian, Y. Zhao, A. Goldsmith, R. Rajagopal, and H. V. Poor, "Distribution System Outage Detection using Consumer Load and Line Flow Measurements," no. May, 2015. 29

[55] J. Lin, C. Wan, Y. Song, R. Huang, X. Chen, W. Guo, Y. Zong, and Y. Shi, "Situation awareness of active distribution network: Roadmap, technologies, and bottlenecks," *CSEE Journal of Power and Energy Systems*, vol. 2, DOI 10.17775/cseejpes.2016.00033, no. 3, pp. 35–42, 2016. 29, 31

[56] S. Bolognani, N. Bof, D. Michelotti, R. Muraro, and L. Schenato, "Identification of power distribution network topology via voltage correlation analysis," *Proceedings of the IEEE Conference on Decision and Control*, DOI 10.1109/CDC.2013.6760120, pp. 1659–1664, 2013. 29, 30, 31

[57] Y. Zhao, J. Chen, and H. V. Poor, "Learning to infer: A new variational inference approach for power grid topology identification," *IEEE Workshop on Statistical Signal Processing Proceedings*, vol. 2016-Augus, DOI 10.1109/SSP.2016.7551827, pp. 1–5, 2016. 29, 30, 31

[58] G. Cavraro, R. Arghandeh, K. Poolla, and A. Von Meier, "Data-driven approach for distribution network topology detection," *IEEE Power and Energy Society General Meeting*, vol. 2015-Septe, DOI 10.1109/PESGM.2015.7286490, pp. 1–5, 2015. 29, 31, 32

[59] K. M. Rogers, R. D. Spadoni, and T. J. Overbye, "Identification of power system topology from synchrophasor data," *2011 IEEE/PES Power Systems Conference and Exposition, PSCE 2011*, vol. 61801, DOI 10.1109/PSCE.2011.5772462, pp. 1–8, 2011. 29, 31

[60] V. Kekatos, G. B. Giannakis, and R. Baldick, "Grid topology identification using electricity prices," *IEEE Power and Energy Society General Meeting*, vol. 2014-Octob, DOI 10.1109/PESGM.2014.6939474, no. October, pp. 1–5, 2014. 29, 31

[61] V. Kekatos, G. B. Giannakis, and R. Baldick, "Online Energy Price Matrix Factorization for Power Grid Topology Tracking," *IEEE Transactions on Smart Grid*, vol. 7, DOI 10.1109/TSG.2015.2469098, no. 3, pp. 1239–1248, 2016. 29, 31

[62] A. Anwar, A. Mahmood, and M. Pickering, "Estimation of smart grid topology using SCADA measurements," *2016 IEEE International Conference on*

Smart Grid Communications, SmartGridComm 2016, DOI 10.1109/SmartGridComm.2016.7778817, no. 1, pp. 539–544, 2016. 29, 31

[63] G. Cavraro and R. Arghandeh, "Power Distribution Network Topology Detection with Time-Series Signature Verification Method," *IEEE Transactions on Power Systems*, vol. 33, DOI 10.1109/TPWRS.2017.2779129, no. 4, pp. 3500–3509, 2018. 29, 31

[64] J. D. Watson, J. Welch, and N. R. Watson, "Use of smart-meter data to determine distribution system topology," *The Journal of Engineering*, vol. 2016, DOI 10.1049/joe.2016.0033, no. 5, pp. 94–101, 2016. 30, 31

[65] S. Zhang, Q. Ding, G. Hu, and C. Nan, "Power Distribution Network Topology Reconstruction Based on Adaptive Lasso," *2nd IEEE Conference on Energy Internet and Energy System Integration, EI2 2018 - Proceedings*, DOI 10.1109/EI2.2018.8582616, pp. 1–6, 2018. 30, 31

[66] M. Zhang, W. Luan, S. Guo, and P. Wang, "Topology identification method of distribution network based on smart meter measurements," *China International Conference on Electricity Distribution, CICED*, DOI 10.1109/CICED.2018.8592228, no. 201808080000037, pp. 372–376, 2018. 30, 31

[67] G. Cavraro, V. Kekatos, and S. Veeramachaneni, "Voltage Analytics for Power Distribution Network Topology Verification," *IEEE Transactions on Smart Grid*, vol. 10, DOI 10.1109/TSG.2017.2758600, no. 1, pp. 1058–1067, 2019. 30, 31

[68] Y. Liao, S. Member, Y. Weng, G. Liu, and S. Member, "Urban MV and LV Distribution Grid Topology Estimation via Group Lasso," vol. 34, no. 1, pp. 12–27, 2019. 30, 31

[69] L. Richaud, R. Pellerej, C. Benoit, and E. Ramos, "Analysis of Voltage Patterns for Topology Identification and Gis," in *25th International Conference on Electricity Distribution*, no. June, pp. 3–6, 2019. 30

[70] R. Jercic, I. Pavic, and I. Damjanovic, "New algorithm for identifying network topology based on artificial neural networks," *2019 2nd International Colloquium on Smart Grid Metrology, SMAGRIMET 2019 - Proceedings*, DOI 10.23919/SMAGRIMET.2019.8720364, pp. 1–5, 2019. 30, 31

[71] J. Yu, Y. Weng, and R. Rajagopal, "PaToPa: A Data-Driven Parameter and Topology Joint Estimation Framework in Distribution Grids," *IEEE Transactions on Power Systems*, vol. 33, DOI 10.1109/TPWRS.2017.2778194, no. 4, pp. 4335–4347, 2018. 30, 31

[72] M. Lisowski, R. Masnicki, and J. Mindykowski, "PLC-Enabled Low Voltage Distribution Network Topology Monitoring," *IEEE Transactions on Smart Grid*, vol. 10, DOI 10.1109/TSG.2019.2904681, no. 6, pp. 6436–6448, 2019. 30, 31

[73] D. Deka, S. Backhaus, and M. Chertkov, "Structure learning in power distribution networks," *IEEE Transactions on Control of Network Systems*, vol. 5, DOI 10.1109/TCNS.2017.2673546, no. 3, pp. 1061–1074, 2018. 31

[74] Y. Guo, Z. Yang, S. Feng, and J. Hu, "Complex Power System Status Monitoring and Evaluation Using Big Data Platform and Machine Learning Algorithms: A Review and a Case Study," *Complexity*, vol. 2018, DOI 10.1155/2018/8496187, 2018. 31, 32

[75] M. Baran and T. E. McDermott, "Distribution system state estimation using AMI data," *2009 IEEE/PES Power Systems Conference and Exposition, PSCE 2009*, vol. 15236, DOI 10.1109/PSCE.2009.4840257, pp. 1–3, 2009. 32

[76] B. Matthiss, J. Erb, and J. Binder, "Using Smart Meters for Distribution Grid State Estimation," *SEST 2019 - 2nd International Conference on Smart Energy Systems and Technologies*, DOI 10.1109/SEST.2019.8849096, pp. 1–5, 2019. 32

[77] M. Pignati, M. Popovic, S. Barreto, R. Cherkaoui, G. Dario Flores, J. Y. Le Boudec, M. Mohiuddin, M. Paolone, P. Romano, S. Sarri, T. Tesfay, D. C. Tomozei, and L. Zanni, "Real-time state estimation of the EPFL-campus medium-voltage grid by using PMUs," *2015 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2015*, DOI 10.1109/ISGT.2015.7131877, 2015. 32

[78] M. Huang, Z. Wei, G. Sun, and H. Zang, "Hybrid State Estimation for Distribution Systems With AMI and SCADA Measurements," *IEEE Access*, vol. 7, DOI 10.1109/access.2019.2937096, pp. 120 350–120 359, 2019. 32

[79] A. M. Prostejovsky, "Increased Observability in Electric Distribution Grids: Emerging Applications Promoting the Active Role of Distribution System Operators," Ph.D. dissertation, 2017. 32

[80] K. R. Mestav, J. Luengo-Rozas, and L. Tong, "Bayesian state estimation for unobservable distribution systems via deep learning," *arXiv*, vol. 34, no. 6, pp. 4910–4920, 2018. 32

[81] M. Nithya, S. Nagarajan, and P. Navaseelan, "Fault detection of wind turbine system using neural networks," *Proceedings - 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development, TIAR 2017*, vol. 2018-Janua, DOI 10.1109/TIAR.2017.8273694, pp. 103–108, 2018. 32

[82] Grid Solutions, "Distribution Feeder Principles," Tech. Rep., 2015. 33

[83] G. Kaikai, C. Kai, G. Ran, W. Ruoyue, P. Zhonghan, and S. You, "Typical Ice-Wind Fault Analysis Based on Correlation Analysis and Naive Bayesian Method," *2020 IEEE 3rd International Conference on Electronics Technology, ICET 2020*, DOI 10.1109/ICET49382.2020.9119709, pp. 372–377, 2020. 33

[84] R. Tervo, J. Karjalainen, and A. Jung, "PREDICTING ELECTRICITY OUTAGES CAUSED BY CONVECTIVE STORMS Roope Tervo , Joonas Karjalainen Observing and information systems centre Finnish Meteorological Institute Alexander Jung Dept of Computer Science Aalto University," *2018 IEEE Data Science Workshop (DSW)*, DOI 10.1109/DSW.2018.8439906, pp. 145–149, 2018. 33, 34

[85] N. Sapountzoglou, J. Lago, and B. Raison, "Fault diagnosis in low voltage smart distribution grids using gradient boosting trees," *Electric Power*

Systems Research, vol. 182, DOI 10.1016/j.epsr.2020.106254, no. September 2019, p. 106254, 2020. 33, 34

[86] D. Chakraborty, U. Sur, and P. K. Banerjee, "Random forest based fault classification technique for active power system networks," *2019 5th IEEE International WIE Conference on Electrical and Computer Engineering, WIECON-ECE 2019 - Proceedings*, DOI 10.1109/WIECON-ECE48653.2019.9019922, pp. 2–5, 2019. 33, 34, 63

[87] M. Shafiullah, M. A. Abido, and Z. Al-Hamouz, "Wavelet-based extreme learning machine for distribution grid fault location," *IET Generation, Transmission and Distribution*, vol. 11, DOI 10.1049/iet-gtd.2017.0656, no. 17, pp. 4256–4263, 2017. 33, 34

[88] W. Fei and P. Moses, "Fault current tracing and identification via machine learning considering distributed energy resources in distribution networks," *Energies*, vol. 12, DOI 10.3390/en12224333, no. 22, pp. 1–12, 2019. 33, 34

[89] I. Ullah, R. U. Khan, F. Yang, and L. Wuttisittikulkij, "Deep learning image-based defect detection in high voltage electrical equipment," *Energies*, vol. 13, DOI 10.3390/en13020392, no. 2, 2020. 33

[90] Y. D. Mamuya, Y. D. Lee, J. W. Shen, M. Shafiullah, and C. C. Kuo, "Application of machine learning for fault classification and location in a radial distribution grid," *Applied Sciences (Switzerland)*, vol. 10, DOI 10.3390/app10144965, no. 14, 2020. 33, 34

[91] R. Moghaddass and J. Wang, "A hierarchical framework for smart grid anomaly detection using large-scale smart meter data," *IEEE Transactions on Smart Grid*, vol. 9, DOI 10.1109/TSG.2017.2697440, no. 6, pp. 5820–5830, 2018. 33

[92] W. Li, D. Deka, M. Chertkov, and M. Wang, "Real-Time Faulted Line Localization and PMU Placement in Power Systems through Convolutional Neural Networks," *IEEE Transactions on Power Systems*, vol. 34, DOI 10.1109/TPWRS.2019.2917794, no. 6, pp. 4640–4651, 2019. 33, 34

[93] Y. Li, X. Song, S. Zhao, and F. Gao, "A Line-Fault Cause Analysis Method for Distribution Network Based on Decision-Making Tree and Machine Learning," *Proceedings - 2020 5th Asia Conference on Power and Electrical Engineering, ACPEE 2020*, DOI 10.1109/ACPEE48638.2020.9136293, pp. 1–5, 2020. 33

[94] H. Jiang, Y. Li, Y. Zhang, J. J. Zhang, D. W. Gao, E. Muljadi, and Y. Gu, "Big Data-Based Approach to Detect, Locate, and Enhance the Stability of an Unplanned Microgrid Islanding," *Journal of Energy Engineering*, vol. 143, DOI 10.1061/(ASCE)EY.1943-7897.0000473, no. 5, pp. 1–11, 2017. 33

[95] C. Cepeda, C. Orozco-Henao, W. Percybrooks, J. D. Pulgarín-Rivera, O. D. Montoya, W. Gil-González, and J. C. Vélez, "Intelligent fault detection system for microgrids," *Energies*, vol. 13, DOI 10.3390/en13051223, no. 5, 2020. 33, 34

[96] T. Gush, S. B. A. Bukhari, K. K. Mehmood, S. Admasie, J. S. Kim, and C. H. Kim, "Intelligent fault classification and location identification

method for microgrids using discrete orthonormal stockwell transform-based optimized multi-kernel extreme learning machine," *Energies*, vol. 12, DOI 10.3390/en12234504, no. 23, 2019. 33, 34

[97] I. Almutairy and M. Alluhaidan, "Fault Diagnosis Based Approach to Protecting DC Microgrid Using Machine Learning Technique," *Procedia Computer Science*, vol. 114, DOI 10.1016/j.procs.2017.09.019, pp. 449–456, 2017. 33, 34

[98] M. Al Karim, J. Currie, and T. T. Lie, "Dynamic Event Detection Using a Distributed Feature Selection Based Machine Learning Approach in a Self-Healing Microgrid," *IEEE Transactions on Power Systems*, vol. 33, DOI 10.1109/TPWRS.2018.2812768, no. 5, pp. 4706–4718, 2018. 33, 34

[99] J. Y. Hsu, Y. F. Wang, K. C. Lin, M. Y. Chen, and J. H. Y. Hsu, "Wind turbine fault diagnosis and predictive maintenance through statistical process control and machine learning," *IEEE Access*, vol. 8, DOI 10.1109/ACCESS.2020.2968615, pp. 23 427–23 439, 2020. 33

[100] S. Lu, T. Sirojan, B. T. Phung, D. Zhang, and E. Ambikairajah, "DA-DCGAN: An Effective Methodology for DC Series Arc Fault Diagnosis in Photovoltaic Systems," *IEEE Access*, vol. 7, DOI 10.1109/ACCESS.2019.2909267, pp. 45 831–45 840, 2019. 33, 34

[101] Z. Chen, F. Han, L. Wu, J. Yu, S. Cheng, P. Lin, and H. Chen, "Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents," *Energy Conversion and Management*, vol. 178, DOI 10.1016/j.enconman.2018.10.040, no. August, pp. 250–264, 2018. 33, 34

[102] E. Garoudja, A. Chouder, K. Kara, and S. Silvestre, "An enhanced machine learning based approach for failures detection and diagnosis of PV systems," *Energy Conversion and Management*, vol. 151, DOI 10.1016/j.enconman.2017.09.019, no. September, pp. 496–513, 2017. 33, 34

[103] Y. Huo, G. Prasad, L. Lampe, and V. C. Leung, "Advanced Smart Grid Monitoring: Intelligent Cable Diagnostics using Neural Networks," *2020 IEEE International Symposium on Power Line Communications and its Applications, ISPLC 2020*, DOI 10.1109/ISPLC48789.2020.9115403, no. Mv, 2020. 33, 34

[104] N. A. Letizia and A. M. Tonello, "Supervised Fault Detection in Energy Grids Measuring Electrical Quantities in the PLC Band," *2020 IEEE International Symposium on Power Line Communications and its Applications, ISPLC 2020*, DOI 10.1109/ISPLC48789.2020.9115408, 2020. 34

[105] L. Gigoni, A. Betti, M. Tucci, and E. Crisostomi, "A Scalable Predictive Maintenance Model for Detecting Wind Turbine Component Failures Based on SCADA Data," *IEEE Power and Energy Society General Meeting*, vol. 2019-Augus, DOI 10.1109/PESGM40551.2019.8973898, 2019. 35

[106] S. Biswal and G. R. Sabareesh, "Design and development of a wind turbine test rig for condition monitoring studies," *2015 International Conference on Industrial Instrumentation and Control, ICIC 2015*, DOI 10.1109/IIC.2015.7150869, no. Icic, pp. 891–896, 2015. 35

[107] H. De Faria, J. G. S. Costa, and J. L. M. Olivas, "A review of monitoring methods for predictive maintenance of electric power transformers based on dissolved gas analysis," *Renewable and Sustainable Energy Reviews*, vol. 46, DOI 10.1016/j.rser.2015.02.052, pp. 201–209, 2015. 35

[108] A. U. Adoghe, C. O. A. Awosope, and J. C. Ekeh, "Asset maintenance planning in electric power distribution network using statistical analysis of outage data," *International Journal of Electrical Power and Energy Systems*, vol. 47, DOI 10.1016/j.ijepes.2012.10.061, no. 1, pp. 424–435, 2013. 35, 36

[109] S. Eke, B. Lyon, I. D. Lyon, T. Aka-ngnui, B. Lyon, I. D. Lyon, B. Lyon, and I. D. Lyon, "Characterization of the Operating Periods of a Power Transformer by Clustering the Dissolved Gas Data," *IEEE*, 2017. 35, 36

[110] F. Kabir, B. Foggo, and N. Yu, "Data driven predictive maintenance of distribution transformers," *China International Conference on Electricity Distribution, CICED*, DOI 10.1109/CICED.2018.8592417, no. 201805280000166, pp. 312–316, 2018. 35, 36

[111] J. L. Velasquez-Contreras, M. A. Sanz-Bobi, and S. Galceran Arellano, "General asset management model in the context of an electric utility: Application to power transformers," *Electric Power Systems Research*, vol. 81, DOI 10.1016/j.epsr.2011.06.007, no. 11, pp. 2015–2037, 2011. 35, 36

[112] R. Venkataswamy, K. Uma Rao, and P. Meena, "Internet of things based metaheuristic reliability centered maintenance of distribution transformers," *IOP Conference Series: Earth and Environmental Science*, vol. 463, DOI 10.1088/1755-1315/463/1/012047, no. 1, 2020. 35, 36

[113] T. R. Chaves, M. A. Martins, and B. A. Pacheco, "MV/LV Overhead Transformer Monitoring and Hot Spot Temperature Estimation," *2020 Intermountain Engineering, Technology and Computing, IETC 2020*, DOI 10.1109/IETC47856.2020.9249225, pp. 6–10, 2020. 35

[114] T. P. Carvalho, F. A. Soares, R. Vita, R. d. P. Francisco, J. P. Basto, and S. G. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers and Industrial Engineering*, vol. 137, DOI 10.1016/j.cie.2019.106024, no. April, p. 106024, 2019. 35

[115] "Predictive maintenance for Power & Utility Industry – Neuron soundware." [Online]. Available: https://www.neuronsw.com/ 35

[116] "Predictive layer." [Online]. Available: https://www.predictivelayer.com/ 35

[117] T. Huuhtanen and A. Jung, "Predictive maintenance of photovoltaic panels via deep learning," *2018 IEEE Data Science Workshop, DSW 2018 - Proceedings*, DOI 10.1109/DSW.2018.8439898, pp. 66–70, 2018. 36

[118] M. R. Skydt, M. Bang, and H. R. Shaker, "A probabilistic sequence classification approach for early fault prediction in distribution grids using long short-term memory neural networks," *Measurement: Journal of the International Measurement Confederation*, vol. 170, DOI 10.1016/j.measurement.2020.108691, no. July 2020, p. 108691, 2021. 36

[119] J. L. Viegas, P. R. Esteves, R. Melício, V. M. Mendes, and S. M. Vieira, "Solutions for detection of non-technical losses in the electricity grid: A review," *Renewable and Sustainable Energy Reviews*, vol. 80, DOI 10.1016/j.rser.2017.05.193, no. May, pp. 1256–1268, 2017. 36, 37

[120] I. Monedero, F. Biscarri, C. León, J. I. Guerrero, J. Biscarri, and R. Millán, "Detection of frauds and other non-technical losses in a power utility using Pearson coefficient, Bayesian networks and decision trees," *International Journal of Electrical Power and Energy Systems*, vol. 34, DOI 10.1016/j.ijepes.2011.09.009, no. 1, pp. 90–98, 2012. 36, 38

[121] "Endesa detected approximately 65,000 cases of electricity fraud in 2018 - Endesa." [Online]. Available: https://www.endesa.com/en/press/press-room/news/customers/endesa-detected-approximately-65000-cases-of-electricity-fraud-in-2018 36

[122] J. Hu and A. V. Vasilakos, "Energy Big Data Analytics and Security: Challenges and Opportunities," *IEEE Transactions on Smart Grid*, vol. 7, DOI 10.1109/TSG.2016.2563461, no. 5, pp. 2432–2436, 2016. 37

[123] K. M. U. Ghori, M. Imran, A. Nawaz, R. A. Abbasi, A. Ullah, and L. Szathmary, "Performance analysis of machine learning classifiers for non-technical loss detection," *Journal of Ambient Intelligence and Humanized Computing*, DOI 10.1007/s12652-019-01649-9, no. 0123456789, 2020. 37

[124] J. L. Viegas, P. R. Esteves, and S. M. Vieira, "Clustering-based novelty detection for identification of non-technical losses," *International Journal of Electrical Power and Energy Systems*, vol. 101, DOI 10.1016/j.ijepes.2018.03.031, no. February, pp. 301–310, 2018. 37, 38

[125] P. Jokar, N. Arianpoo, and V. C. Leung, "Electricity theft detection in AMI using customers' consumption patterns," *IEEE Transactions on Smart Grid*, vol. 7, DOI 10.1109/TSG.2015.2425222, no. 1, pp. 216–226, 2016. 37, 38, 41

[126] E. Villar-Rodriguez, J. Del Ser, I. Oregi, M. N. Bilbao, and S. Gil-Lopez, "Detection of non-technical losses in smart meter data based on load curve profiling and time series analysis," *Energy*, vol. 137, DOI 10.1016/j.energy.2017.07.008, pp. 118–128, 2017. 37, 38

[127] R. G. Rodríguez, J. J. Mares, and C. G. Quintero, "Computational intelligent approaches for non-technical losses management of electricity," *Energies*, vol. 13, DOI 10.3390/en13092393, no. 9, 2020. 37, 38

[128] P. Glauner, A. Boechat, L. Dolberg, R. State, F. Bettinger, Y. Rangoni, and D. Duarte, "Large-scale detection of non-technical losses in imbalanced data sets," *2016 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2016*, DOI 10.1109/ISGT.2016.7781159, pp. 1–5, 2016. 37, 38

[129] G. Figueroa, Y. S. Chen, N. Avila, and C. C. Chu, "Improved practices in machine learning algorithms for NTL detection with imbalanced data," *IEEE Power and Energy Society General Meeting*, vol. 2018-Janua, DOI 10.1109/PESGM.2017.8273852, pp. 1–5, 2018. 37, 38

[130] H. Qin, H. Zhou, and J. Cao, "Imbalanced learning algorithm based intelligent abnormal electricity consumption detection," *Neurocomputing*, vol. 402, DOI 10.1016/j.neucom.2020.03.085, pp. 112–123, 2020. 37, 38

[131] D. Martino, F. Decia, J. Molinelli, and A. Fern, "Improving Electric Fraud Detection using class imbalance strategies," no. March, pp. 135–141, 2010. 37, 38

[132] M. Zanetti, E. Jamhour, M. Pellenz, M. Penna, V. Zambenedetti, and I. Chueiri, "A Tunable Fraud Detection System for Advanced Metering Infrastructure Using Short-Lived Patterns," *IEEE Transactions on Smart Grid*, vol. 10, DOI 10.1109/TSG.2017.2753738, no. 1, pp. 830–840, 2019. 37, 38

[133] X. Lu, Y. Zhou, Z. Wang, Y. Yi, L. Feng, and F. Wang, "Knowledge embedded semi-supervised deep learning for detecting non-technical losses in the smart grid," *Energies*, vol. 12, DOI 10.3390/en12183452, no. 18, pp. 1–18, 2019. 37, 38

[134] J. A. Meira, P. Glauner, R. State, P. Valtchev, L. Dolberg, F. Bettinger, and D. Duarte, "Distilling provider-independent data for general detection of non-technical losses," *2017 IEEE Power and Energy Conference at Illinois, PECI 2017*, DOI 10.1109/PECI.2017.7935765, pp. 3–7, 2017. 38

[135] S. Shekara, S. Reddy, L. Wang, and V. Devabhaktuni, "Support Vector Machine Based Data Classification for Detection of Electricity Theft," *IEEE Xplore*, pp. 1–8, 2011. 38

[136] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and F. Nagi, "Improving SVM-based nontechnical loss detection in power utility using the fuzzy inference system," *IEEE Transactions on Power Delivery*, vol. 26, DOI 10.1109/TPWRD.2010.2055670, no. 2, pp. 1284–1285, 2011. 38

[137] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad, "Nontechnical loss detection for metered customers in power utility using support vector machines," *IEEE Transactions on Power Delivery*, vol. 25, DOI 10.1109/TPWRD.2009.2030890, no. 2, pp. 1162–1171, 2010. 38

[138] A. Amara Korba and N. El Islem Karabadji, "Smart Grid Energy Fraud Detection Using SVM," *Proceedings - ICNAS 2019: 4th International Conference on Networking and Advanced Systems*, DOI 10.1109/ICNAS.2019.8807832, pp. 1–6, 2019. 38

[139] J. E. Cabral, J. O. Pinto, and A. M. Pinto, "Fraud detection system for high and low voltage electricity consumers based on data mining," *2009 IEEE Power and Energy Society General Meeting, PES '09*, DOI 10.1109/PES.2009.5275809, pp. 1–5, 2009. 38

[140] J. Li and F. Wang, "Non-technical loss detection in power grids with statistical profile images based on semi- supervised learning," *Sensors (Switzerland)*, vol. 20, DOI 10.3390/s20010236, no. 1, 2020. 38

[141] M. Nazmul Hasan, R. N. Toma, A. A. Nahid, M. M. Manjurul Islam, and J. M. Kim, "Electricity theft detection in smart grid systems: A CNN-LSTM based approach," *Energies*, vol. 12, DOI 10.3390/en12173310, no. 17, pp. 1–18, 2019. 38

[142] R. R. Bhat, R. D. Trevizan, R. Sengupta, X. Li, and A. Bretas, "Identifying nontechnical power loss via spatial and temporal deep learning," *Proceedings - 2016 15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016*, DOI 10.1109/ICMLA.2016.107, pp. 272–279, 2017. 38

[143] N. J. Johannesen, M. Kolhe, and M. Goodwin, "Relative evaluation of regression tools for urban area electrical energy demand forecasting," *Journal of Cleaner Production*, vol. 218, DOI 10.1016/j.jclepro.2019.01.108, pp. 555–564, 2019. 39, 40

[144] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, "Deep learning for estimating building energy consumption," *Sustainable Energy, Grids and Networks*, vol. 6, DOI 10.1016/j.segan.2016.02.005, pp. 91–99, 2016. 39

[145] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani, "Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches," *Energies*, vol. 11, DOI 10.3390/en11071636, no. 7, 2018. 39

[146] M. Bourdeau, X. qiang Zhai, E. Nefzaoui, X. Guo, and P. Chatellier, "Modeling and forecasting building energy consumption: A review of data-driven techniques," 2019. 39

[147] S. Mujeeb, N. Javaid, and S. Javaid, "Data Analytics for Price Forecasting in Smart Grids: A Survey," *Proceedings of the 21st International Multi Topic Conference, INMIC 2018*, DOI 10.1109/INMIC.2018.8595571, pp. 1–10, 2018. 39, 45

[148] B. Luo and S. Miao, "A Literature Survey on Electricity Price Forecasting in Deregulated Markets," *iSPEC 2019 - 2019 IEEE Sustainable Power and Energy Conference: Grid Modernization for Energy Revolution, Proceedings*, DOI 10.1109/iSPEC48194.2019.8974939, pp. 359–365, 2019. 39, 45

[149] T. Ahmad, H. Zhang, and B. Yan, "A review on renewable energy and electricity requirement forecasting models for smart grid and buildings," *Sustainable Cities and Society*, vol. 55, DOI 10.1016/j.scs.2020.102052, no. April 2019, p. 102052, 2020. 39, 44, 59

[150] M. Q. Raza and A. Khosravi, "A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings," *Renewable and Sustainable Energy Reviews*, vol. 50, DOI 10.1016/j.rser.2015.04.065, pp. 1352–1372, 2015. 39

[151] L. Wu and M. Shahidehpour, "A hybrid model for integrated day-ahead electricity price and load forecasting in smart grid," *IET Generation, Transmission and Distribution*, vol. 8, DOI 10.1049/iet-gtd.2013.0927, no. 12, pp. 1937–1950, 2014. 39

[152] R. K. Agrawal, F. Muchahary, and M. M. Tripathi, "Ensemble of relevance vector machines and boosted trees for electricity price forecasting," *Applied Energy*, vol. 250, DOI 10.1016/j.apenergy.2019.05.062, no. April, pp. 540–548, 2019. 39, 49

[153] S. Mujeeb, N. Javaid, M. Ilahi, Z. Wadud, F. Ishmanov, and M. K. Afzal, "Deep long short-term memory: A new price and load forecasting scheme for big data in smart cities," *Sustainability (Switzerland)*, vol. 11, DOI 10.3390/su11040987, no. 4, pp. 1–29, 2019. 39, 43, 45, 50

[154] R. Khalid, N. Javaid, F. A. Al-zahrani, K. Aurangzeb, E. U. H. Qazi, and T. Ashfaq, "Electricity load and price forecasting using jaya-long short term memory (JLSTM) in smart grids," *Entropy*, vol. 22, DOI 10.3390/e22010010, no. 1, p. 10, 2020. 39, 45, 50

[155] S. Mujeeb and N. Javaid, "ESAENARX and DE-RELM: Novel schemes for big data predictive analytics of electricity load and price," *Sustainable Cities and Society*, vol. 51, DOI 10.1016/j.scs.2019.101642, no. May, p. 101642, 2019. 39, 48

[156] M. Tan, S. Yuan, S. Li, Y. Su, H. Li, and F. H. He, "Ultra-Short-Term Industrial Power Demand Forecasting Using LSTM Based Hybrid Ensemble Learning," *IEEE Transactions on Power Systems*, vol. 35, DOI 10.1109/TPWRS.2019.2963109, no. 4, pp. 2937–2948, 2020. 39, 43

[157] I. P. Panapakidis and A. S. Dagoumas, "Day-ahead electricity price forecasting via the application of artificial neural network based models," *Applied Energy*, vol. 172, DOI 10.1016/j.apenergy.2016.03.089, pp. 132–151, 2016. 39, 45

[158] F. Zhang and H. Fleyeh, "A Review of single artificial neural network models for electricity spot price forecasting," *International Conference on the European Energy Market, EEM*, vol. 2019-Septe, DOI 10.1109/EEM.2019.8916423, pp. 1–6, 2019. 39, 45

[159] H. Jahangir, H. Tayarani, S. Baghali, A. Ahmadian, A. Elkamel, M. A. Golkar, and M. Castilla, "A Novel Electricity Price Forecasting Approach Based on Dimension Reduction Strategy and Rough Artificial Neural Networks," *IEEE Transactions on Industrial Informatics*, vol. 16, DOI 10.1109/TII.2019.2933009, no. 4, pp. 2369–2381, 2020. 39, 49

[160] M. Askari and F. Keynia, "Mid-term electricity load forecasting by a new composite method based on optimal learning MLP algorithm," *IET Generation, Transmission and Distribution*, vol. 14, DOI 10.1049/iet-gtd.2019.0797, no. 5, pp. 845–852, 2020. 39, 42

[161] F. Ziel and R. Steinert, "Probabilistic mid- and long-term electricity price forecasting," *Renewable and Sustainable Energy Reviews*, vol. 94, DOI 10.1016/j.rser.2018.05.038, no. June, pp. 251–266, 2018. 39, 45

[162] N. Wei, C. Li, X. Peng, F. Zeng, and X. Lu, "Conventional models and artificial intelligence-based models for energy consumption forecasting: A review," *Journal of Petroleum Science and Engineering*, vol. 181, DOI 10.1016/j.petrol.2019.106187, no. June, p. 106187, 2019. 40

[163] C. Kuster, Y. Rezgui, and M. Mourshed, "Electrical load forecasting models: A critical systematic review," *Sustainable Cities and Society*, vol. 35, DOI 10.1016/j.scs.2017.08.009, no. July, pp. 257–270, 2017. 40

[164] A. A. Mamun, M. Sohel, N. Mohammad, M. S. Haque Sunny, D. R. Dipta, and E. Hossain, "A Comprehensive Review of the Load Forecasting Techniques Using Single and Hybrid Predictive Models," *IEEE Access*, vol. 8, DOI 10.1109/ACCESS.2020.3010702, pp. 134 911–134 939, 2020. 40

[165] L. Suganthi and A. A. Samuel, "Energy models for demand forecasting - A review," *Renewable and Sustainable Energy Reviews*, vol. 16, DOI 10.1016/j.rser.2011.08.014, no. 2, pp. 1223–1240, 2012. 40, 41

[166] S. N. Fallah, M. Ganjkhani, S. Shamshirband, and K. wing Chau, "Computational intelligence on short-term load forecasting: A methodological overview," *Energies*, vol. 12, DOI 10.3390/en12030393, no. 3, 2019. 40

[167] S. Member, S. Member, and E. Systems, "A Generalized Knowledge-Based Short-Term Load-Forecasting Technique," vol. 8, no. 2, pp. 508–514, 1993. 40

[168] H. Ahmed, "Applying Big Data Analytics for Energy Efficiency," *Master's Programme in International Design Business Management Hussnain*, p. 85, 2014. 40, 41

[169] L. Cheng and T. Yu, "Game-Theoretic Approaches Applied to Transactions in the Open and Ever-Growing Electricity Markets from the Perspective of Power Demand Response: An Overview," *IEEE Access*, vol. 7, DOI 10.1109/ACCESS.2019.2900356, pp. 25 727–25 762, 2019. 40

[170] L. Guo, S. Wang, H. Chen, and Q. Shi, "A Load Identification Method Based on Active Deep Learning and Discrete Wavelet Transform," *IEEE Access*, vol. 8, DOI 10.1109/access.2020.3003778, pp. 1–1, 2020. 41

[171] V. Singhal, J. Maggu, and A. Majumdar, "Simultaneous Detection of Multiple Appliances from Smart-Meter Measurements via Multi-Label Consistent Deep Dictionary Learning and Deep Transform Learning," *IEEE Transactions on Smart Grid*, vol. 10, DOI 10.1109/TSG.2018.2815763, no. 3, pp. 2969–2978, 2019. 41

[172] J. G. Kim and B. Lee, "Appliance classification by power signal analysis based on multi-feature combination multi-layer LSTM," *Energies*, vol. 12, DOI 10.3390/en12142804, no. 14, 2019. 41

[173] X. Wu, Y. Gao, and D. Jiao, "Multi-label classification based on Random Forest algorithm for non-intrusive load monitoring system," *Processes*, vol. 7, DOI 10.3390/pr7060337, no. 6, 2019. 41

[174] T. T. H. Le, H. Kang, and H. Kim, "Household appliance classification using lower odd-numbered harmonics and the bagging decision tree," *IEEE Access*, vol. 8, DOI 10.1109/ACCESS.2020.2981969, pp. 55 937–55 952, 2020. 41

[175] S. Welikala, C. Dinesh, M. P. B. Ekanayake, R. I. Godaliyadda, and J. Ekanayake, "Incorporating Appliance Usage Patterns for Non-Intrusive Load Monitoring and Load Forecasting," *IEEE Transactions on Smart Grid*, vol. 10, DOI 10.1109/TSG.2017.2743760, no. 1, pp. 448–461, 2019. 41

[176] C. Chelmis, J. Kolte, and V. K. Prasanna, "Big data analytics for demand response: Clustering over space and time," *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, DOI 10.1109/BigData.2015.7364011, pp. 2223–2232, 2015. 41

[177] S. Singh and A. Yassine, "Big data mining of energy time series for behavioral analytics and energy consumption forecasting," *Energies*, vol. 11, DOI 10.3390/en11020452, no. 2, 2018. 41

[178] Y. Wang, Q. Chen, C. Kang, and Q. Xia, "Clustering of Electricity Consumption Behavior Dynamics Toward Big Data Applications," *IEEE Transactions on Smart Grid*, vol. 7, DOI 10.1109/TSG.2016.2548565, no. 5, pp. 2437–2447, 2016. 41

[179] S. Singh and A. Yassine, "Mining energy consumption behavior patterns for households in smart grid," *IEEE Transactions on Emerging Topics in Computing*, vol. 7, DOI 10.1109/TETC.2017.2692098, no. 3, pp. 404–419, 2019. 41

[180] T. K. Wijaya, M. Vasirani, S. Humeau, and K. Aberer, "Cluster-based aggregate forecasting for residential electricity demand using smart meter data," *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, DOI 10.1109/BigData.2015.7363836, pp. 879–887, 2015. 41

[181] A. Satre-Meloy, M. Diakonova, and P. Grünewald, "Cluster analysis and prediction of residential peak demand profiles using occupant activity data," *Applied Energy*, vol. 260, DOI 10.1016/j.apenergy.2019.114246, no. November 2019, p. 114246, 2020. 41

[182] H. Chen, S. Wang, S. Wang, and Y. Li, "Day-ahead aggregated load forecasting based on two-terminal sparse coding and deep neural network fusion," *Electric Power Systems Research*, vol. 177, DOI 10.1016/j.epsr.2019.105987, no. August, 2019. 41, 42

[183] S. V. Oprea and A. Bara, "Machine Learning Algorithms for Short-Term Load Forecast in Residential Buildings Using Smart Meters, Sensors and Big Data Solutions," *IEEE Access*, vol. 7, DOI 10.1109/ACCESS.2019.2958383, pp. 177 874–177 889, 2019. 41, 42

[184] R. Machlev, D. Tolkachov, Y. Levron, and Y. Beck, "Dimension reduction for NILM classification based on principle component analysis," *Electric Power Systems Research*, vol. 187, DOI 10.1016/j.epsr.2020.106459, no. February 2019, p. 106459, 2020. 41

[185] A. F. Da Veloso, R. G. De Oliveira, A. A. Rodrigues, R. A. Rabelo, and J. J. Rodrigues, "Cognitive smart plugs for signature identification of residential home appliance load using machine learning: From theory to practice," *2019 IEEE International Conference on Communications Workshops, ICC Workshops 2019 - Proceedings*, DOI 10.1109/ICCW.2019.8756885, 2019. 41, 52

[186] O. Adeoye and C. Spataru, "Modelling and forecasting hourly electricity demand in West African countries," *Applied Energy*, vol. 242, DOI 10.1016/j.apenergy.2019.03.057, no. November 2018, pp. 311–333, 2019. 41, 42

[187] M. Bessec and J. Fouquau, "Short-run electricity load forecasting with combinations of stationary wavelet transforms," *European Journal of Operational Research*, vol. 264, DOI 10.1016/j.ejor.2017.05.037, no. 1, pp. 149–164, 2018. 42

[188] D. Wei, J. Wang, K. Ni, and G. Tang, "Research and application of a novel hybrid model based on a deep neural network combined with fuzzy time series for energy forecasting," *Energies*, vol. 12, DOI 10.3390/en12183588, no. 18, 2019. 42

[189] J. Zhang, Y. M. Wei, D. Li, Z. Tan, and J. Zhou, "Short term electricity load forecasting using a hybrid model," *Energy*, vol. 158, DOI 10.1016/j.energy.2018.06.012, pp. 774–781, 2018. 42

[190] C. Tong, J. Li, C. Lang, F. Kong, J. Niu, and J. J. Rodrigues, "An efficient deep model for day-ahead electricity load forecasting with stacked denoising auto-encoders," *Journal of Parallel and Distributed Computing*, vol. 117, DOI 10.1016/j.jpdc.2017.06.007, pp. 267–273, 2018. 42

[191] W. Ahmad, N. Ayub, T. Ali, M. Irfan, M. Awais, M. Shiraz, and A. Glowacz, "Towards short term electricity load forecasting using improved support vector machine and extreme learning machine," *Energies*, vol. 13, DOI 10.3390/en13112907, no. 11, pp. 1–17, 2020. 42

[192] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, DOI 10.1145/3298981, no. 2, pp. 1–19, 2019. 42, 133, 134

[193] S. R. Khuntia, J. L. Rueda, and M. A. Van der Meijden, "Long-term electricity load forecasting considering volatility using multiplicative error model," *Energies*, vol. 11, DOI 10.3390/en11123308, no. 12, 2018. 42

[194] Q. Yang, Y. Liu, Y. Cheng, Y. Kang, T. Chen, and H. Yu, "Federated Learning," *Synth. Lect. Artif. Intell. Mach. Learn.*, vol. 13, DOI 10.2200/S00960ED2V01Y201910AIM043, no. 3, pp. 1–207, 2020. 43, 134

[195] J. Pórtoles, C. González, and J. M. Moguerza, "Electricity Price Forecasting with Dynamic Trees: A Benchmark Against the Random Forest Approach," *Energies*, vol. 11, DOI 10.3390/en11061588, no. 6, 2018. 43, 48, 49

[196] C. Forecasting and U. Deep, "HousEEC : Day-Ahead Household Electrical Energy." 43

[197] H. Shi, M. Xu, and R. Li, "Deep Learning for Household Load Forecasting-A Novel Pooling Deep RNN," *IEEE Transactions on Smart Grid*, vol. 9, DOI 10.1109/TSG.2017.2686012, no. 5, pp. 5271–5280, 2018. 43

[198] S. K. Acharya, Y. M. Wi, and J. Lee, "Short-term load forecasting for a single household based on convolution neural networks using data augmentation," *Energies*, vol. 12, DOI 10.3390/en12183560, no. 18, 2019. 43

[199] Z. Guo, K. Zhou, X. Zhang, and S. Yang, "A deep learning model for short-term power load and probability density forecasting," *Energy*, vol. 160, DOI 10.1016/j.energy.2018.07.090, pp. 1186–1200, 2018. 43

[200] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network," *IEEE Transactions on Smart Grid*, vol. 10, DOI 10.1109/TSG.2017.2753802, no. 1, pp. 841–851, 2019. 43

[201] D. Solyali, "A Comparative Analysis of Machine Learning Approaches for Short-/Long-Term Electricity Load Forecasting in Cyprus," *Sustainability*, vol. 12, DOI 10.3390/su12093612, no. 9, p. 3612, 2020. 43, 45

[202] X. J. Luo, "A novel clustering-enhanced adaptive artificial neural network model for predicting day-ahead building cooling demand," *Journal of Building Engineering*, vol. 32, DOI 10.1016/j.jobe.2020.101504, no. March, p. 101504, 2020. 43

[203] N. M. Shah, "Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques," *IET Renewable Power Generation*, vol. 13, no. 7, pp. 1009–1023(14), 2019. 44

[204] S. Sobri, S. Koohi-Kamali, and N. A. Rahim, "Solar photovoltaic generation forecasting methods: A review," *Energy Conversion and Management*, vol. 156, DOI 10.1016/j.enconman.2017.11.019, no. May 2017, pp. 459–497, 2018. 44

[205] A. Zendehboudi, M. A. Baseer, and R. Saidur, "Application of support vector machine models for forecasting solar and wind energy resources: A review," *Journal of Cleaner Production*, vol. 199, DOI 10.1016/j.jclepro.2018.07.164, pp. 272–285, 2018. 44

[206] M. Sharifzadeh, A. Sikinioti-Lock, and N. Shah, "Machine-learning methods for integrated renewable power generation: A comparative study of artificial neural networks, support vector regression, and Gaussian Process Regression," *Renewable and Sustainable Energy Reviews*, vol. 108, DOI 10.1016/j.rser.2019.03.040, no. July 2018, pp. 513–538, 2019. 44

[207] L. Gigoni, A. Betti, E. Crisostomi, A. Franco, M. Tucci, F. Bizzarri, and D. Mucci, "Day-Ahead Hourly Forecasting of Power Generation from Photovoltaic Plants," *IEEE Transactions on Sustainable Energy*, vol. 9, DOI 10.1109/TSTE.2017.2762435, no. 2, pp. 831–842, 2018. 44

[208] K. Wang, X. Qi, and H. Liu, "A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network," *Applied Energy*, vol. 251, DOI 10.1016/j.apenergy.2019.113315, no. May, p. 113315, 2019. 44

[209] J. Jeong and H. Kim, "Multi-site photovoltaic forecasting exploiting space-time convolutional neural network," *Energies*, vol. 12, DOI 10.3390/en12234490, no. 23, 2019. 44, 46

[210] A. Tascikaraoglu and M. Uzunoglu, "A review of combined approaches for prediction of short-term wind speed and power," *Renewable and Sustainable Energy Reviews*, vol. 34, DOI 10.1016/j.rser.2014.03.033, pp. 243–254, 2014. 44

[211] A. P. Marugán, F. P. G. Márquez, J. M. P. Perez, and D. Ruiz-Hernández, "A survey of artificial neural network in wind energy systems," *Applied Energy*, vol. 228, DOI 10.1016/j.apenergy.2018.07.084, no. June, pp. 1822–1836, 2018. 44

[212] H. zhi Wang, G. qiang Li, G. bing Wang, J. chun Peng, H. Jiang, and Y. tao Liu, "Deep learning based ensemble approach for probabilistic wind power

forecasting," *Applied Energy*, vol. 188, DOI 10.1016/j.apenergy.2016.11.111, pp. 56–70, 2017. 45

[213] A. T. Eseye, J. Zhang, and D. Zheng, "Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information," *Renewable Energy*, vol. 118, DOI 10.1016/j.renene.2017.11.011, pp. 357–367, 2018. 45, 46

[214] H. Liu, X. wei Mi, and Y. fei Li, "Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network," *Energy Conversion and Management*, vol. 156, DOI 10.1016/j.enconman.2017.11.053, no. September 2017, pp. 498–514, 2018. 45, 47

[215] M. Q. Raza, M. Nadarajah, J. Li, K. Y. Lee, and H. B. Gooi, "An Ensemble Framework For Day-Ahead Forecast of PV Output in Smart Grids," *IEEE Transactions on Industrial Informatics*, DOI 10.1109/TII.2018.2882598, pp. 1–10, 2018. 45, 46

[216] H. Shaker, H. Zareipour, and D. Wood, "Estimating Power Generation of Invisible Solar Sites Using Publicly Available Data," *IEEE Transactions on Smart Grid*, vol. 7, DOI 10.1109/TSG.2016.2533164, no. 5, pp. 2456–2465, 2016. 45

[217] X. Qing and Y. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM," *Energy*, vol. 148, DOI 10.1016/j.energy.2018.01.177, pp. 461–468, 2018. 46

[218] B. Chen, P. Lin, Y. Lai, S. Cheng, Z. Chen, and L. Wu, "Very-short-term power prediction for PV power plants using a simple and effective RCC-LSTM model based on short term multivariate historical datasets," *Electronics (Switzerland)*, vol. 9, DOI 10.3390/electronics9020289, no. 2, 2020. 46

[219] M. Abdel-Nasser and K. Mahmoud, "Accurate photovoltaic power forecasting models using deep LSTM-RNN," *Neural Computing and Applications*, vol. 31, DOI 10.1007/s00521-017-3225-z, no. 7, pp. 2727–2740, 2019. 46

[220] H. Zang, L. Cheng, T. Ding, K. W. Cheung, Z. Liang, Z. Wei, and G. Sun, "Hybrid method for short-term photovoltaic power forecasting based on deep convolutional neural network," *IET Generation, Transmission and Distribution*, vol. 12, DOI 10.1049/iet-gtd.2018.5847, no. 20, pp. 4557–4567, 2018. 46

[221] L. Liu, Y. Zhao, D. Chang, J. Xie, Z. Ma, Q. Sun, H. Yin, and R. Wennersten, "Prediction of short-term PV power output and uncertainty analysis," *Applied Energy*, vol. 228, DOI 10.1016/j.apenergy.2018.06.112, no. November, pp. 700–711, 2018. 46

[222] D. Lee and K. Kim, "Recurrent neural network-based hourly prediction of photovoltaic power output using meteorological information," *Energies*, vol. 12, DOI 10.3390/en12020215, no. 2, 2019. 46

[223] F. Rodríguez, A. Fleetwood, A. Galarza, and L. Fontán, "Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control," *Renewable Energy*, vol. 126, DOI 10.1016/j.renene.2018.03.070, pp. 855–864, 2018. 46

[224] H. Zhou, Y. Zhang, L. Yang, Q. Liu, K. Yan, and Y. Du, "Short-Term photovoltaic power forecasting based on long short term memory neural network and attention mechanism," *IEEE Access*, vol. 7, DOI 10.1109/ACCESS.2019.2923006, pp. 78 063–78 074, 2019. 46

[225] Y. Ju, G. Sun, Q. Chen, M. Zhang, H. Zhu, and M. U. Rehman, "A model combining convolutional neural network and lightgbm algorithm for ultra-short-term wind power forecasting," *IEEE Access*, vol. 7, DOI 10.1109/ACCESS.2019.2901920, pp. 28 309–28 318, 2019. 47

[226] A. Lahouar and J. Ben Hadj Slama, "Hour-ahead wind power forecast based on random forests," *Renewable Energy*, vol. 109, DOI 10.1016/j.renene.2017.03.064, pp. 529–541, 2017. 47

[227] J. Chen, G. Q. Zeng, W. Zhou, W. Du, and K. D. Lu, "Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization," *Energy Conversion and Management*, vol. 165, DOI 10.1016/j.enconman.2018.03.098, no. March, pp. 681–695, 2018. 47

[228] B. Zhou, X. Ma, Y. Luo, and D. Yang, "Wind Power Prediction Based on LSTM Networks and Nonparametric Kernel Density Estimation," *IEEE Access*, vol. 7, DOI 10.1109/ACCESS.2019.2952555, no. Mcc, pp. 165 279–165 292, 2019. 47

[229] A. S. Qureshi, A. Khan, A. Zameer, and A. Usman, "Wind power prediction using deep neural network based meta regression and transfer learning," *Applied Soft Computing Journal*, vol. 58, DOI 10.1016/j.asoc.2017.05.031, pp. 742–755, 2017. 47

[230] J. Nowotarski and R. Weron, "Recent advances in electricity price forecasting: A review of probabilistic forecasting," *Renewable and Sustainable Energy Reviews*, vol. 81, DOI 10.1016/j.rser.2017.05.234, no. March 2017, pp. 1548–1568, 2018. 45

[231] J. Lago, F. De Ridder, and B. De Schutter, "Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms," *Applied Energy*, vol. 221, DOI 10.1016/j.apenergy.2018.02.069, no. April, pp. 386–405, 2018. 45, 49

[232] S. Demir, K. Mincev, K. Kok, and N. G. Paterakis, "Introducing Technical Indicators to Electricity Price Forecasting: A Feature Engineering Study for Linear, Ensemble, and Deep Machine Learning Models," *Applied Sciences*, vol. 10, DOI 10.3390/app10010255, no. 1, p. 255, 2019. 45

[233] P. H. Kuo and C. J. Huang, "An electricity price forecasting model by hybrid structured deep neural networks," *Sustainability (Switzerland)*, vol. 10, DOI 10.3390/su10041280, no. 4, pp. 1–17, 2018. 45, 49

[234] S. Atef and A. B. Eltawil, "A Comparative Study Using Deep Learning and Support Vector Regression for Electricity Price Forecasting in Smart Grids," *2019 IEEE 6th International Conference on Industrial Engineering and Applications, ICIEA 2019*, DOI 10.1109/IEA.2019.8715213, pp. 603–607, 2019. 45, 49

[235] S. Luo and Y. Weng, "A two-stage supervised learning approach for electricity price forecasting by leveraging different data sources," *Applied Energy*, vol. 242, DOI 10.1016/j.apenergy.2019.03.129, no. February, pp. 1497–1512, 2019. 45, 49

[236] R. A. Chinnathambi, S. J. Plathottam, T. Hossen, A. S. Nair, and P. Ranganathan, "Deep Neural Networks (DNN) for Day-Ahead Electricity Price Markets," *2018 IEEE Electrical Power and Energy Conference, EPEC 2018*, DOI 10.1109/EPEC.2018.8598327, 2018. 45, 49

[237] U. Ugurlu, I. Oksuz, and O. Tas, "Electricity price forecasting using recurrent neural networks," *Energies*, vol. 11, DOI 10.3390/en11051255, no. 5, pp. 1–23, 2018. 45, 48, 50

[238] T. Windler, J. Busse, and J. Rieck, "One month-ahead electricity price forecasting in the context of production planning," *Journal of Cleaner Production*, vol. 238, DOI 10.1016/j.jclepro.2019.117910, p. 117910, 2019. 45, 50

[239] T. Wilcox, N. Jin, P. Flach, and J. Thumim, "A Big Data platform for smart meter data analytics," *Computers in Industry*, vol. 105, DOI 10.1016/j.compind.2018.12.010, pp. 250–259, 2019. 45

[240] A. Shiri, M. Afshar, A. Rahimi-Kian, and B. Maham, "Electricity price forecasting using Support Vector Machines by considering oil and natural gas price impacts," *International Conference on Smart Energy Grid Engineering, SEGE 2015*, DOI 10.1109/SEGE.2015.7324591, pp. 1–5, 2015. 45, 48, 50

[241] G. Díaz, J. Coto, and J. Gómez-Aleixandre, "Prediction and explanation of the formation of the Spanish day-ahead electricity price through machine learning regression," *Applied Energy*, vol. 239, DOI 10.1016/j.apenergy.2019.01.213, no. February, pp. 610–625, 2019. 48, 49

[242] C. Xiao, D. Sutanto, K. M. Muttaqi, and M. Zhang, "Online Sequential Extreme Learning Machine Algorithm for Better Prediction of the Real-time Electricity Price under Dynamic Environmental Changes," *2019 IEEE Industry Applications Society Annual Meeting, IAS 2019*, DOI 10.1109/IAS.2019.8912333, pp. 1–6, 2019. 48, 49

[243] J. Wang, W. Yang, P. Du, and T. Niu, "Outlier-robust hybrid electricity price forecasting model for electricity market management," *Journal of Cleaner Production*, vol. 249, DOI 10.1016/j.jclepro.2019.119318, p. 119318, 2020. 48, 49

[244] R. Bisoi, P. K. Dash, and P. P. Das, "Short-term electricity price forecasting and classification in smart grids using optimized multikernel extreme learning machine," *Neural Computing and Applications*, vol. 32, DOI 10.1007/s00521-018-3652-5, no. 5, pp. 1457–1480, 2020. 48, 49

[245] J. Wang, F. Liu, Y. Song, and J. Zhao, "A novel model: Dynamic choice artificial neural network (DCANN) for an electricity price forecasting system," *Applied Soft Computing Journal*, vol. 48, DOI 10.1016/j.asoc.2016.07.011, pp. 281–297, 2016. 49

[246] A. Alshejari, V. S. Kodogiannis, and S. Leonidis, "Development of Neurofuzzy Architectures for Electricity Price Forecasting," DOI 10.3390/en13051209, pp. 1–25, 2020. 49

[247] R. A. De Marcos, J. Reneses, and A. Bello, "Long-term Spanish electricity market price forecasting with cointegration and VEC models," *2016 International Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2016 - Proceedings*, DOI 10.1109/PMAPS.2016.7764158, 2016. 50

[248] Universal Smart Energy Framework (USEF), "USEF: The framework explained," Tech. Rep., 2015. [Online]. Available: www.usef.energy/download-the-framework/ 48

[249] S. Dalhues, Y. Zhou, O. Pohl, F. Rewald, F. Erlemeyer, D. Schmid, Z. Hagemann, C. Wagner, D. M. Gonzalez, and H. Liu, "Towards research and practice of flexibility in distribution systems: A review," *CSEE Journal of Power and Energy Systems*, vol. 5, DOI 10.17775/cseejpes.2019.00170, no. 3, pp. 285–294, 2019. 51

[250] R. Ahmadiahangar, T. Häring, A. Rosin, T. Korõtko, and J. Martins, "Residential Load Forecasting for Flexibility Prediction Using Machine Learning-Based Regression Model," *Proceedings - 2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe, EEEIC/I and CPS Europe 2019*, DOI 10.1109/EEEIC.2019.8783634, 2019. 51

[251] Vallés Mercedes, "Efficient Implementation and Potential Benefits of Demand Response in Electricity Distribution Networks," 2017. 51

[252] BEUC, "Electricity Aggregators: Starting off on the right foot with consumers," vol. 32, no. 9505781573, pp. 0–11, 2018. 51

[253] C. Heinrich, C. Ziras, A. L. Syrri, and H. W. Bindner, "EcoGrid 2.0: A large-scale field trial of a local flexibility market," *Applied Energy*, vol. 261, DOI 10.1016/j.apenergy.2019.114399, no. December 2019, p. 114399, 2020. 51

[254] J. Iria, F. Soares, and M. Matos, "Optimal bidding strategy for an aggregator of prosumers in energy and secondary reserve markets," *Applied Energy*, vol. 238, DOI 10.1016/j.apenergy.2019.01.191, no. July 2018, pp. 1361–1372, 2019. 51

[255] S. Ø. Ottesen, A. Tomasgard, and S.-E. Fleten, "Multi market bidding strategies for demand side flexibility aggregators in electricity markets," *European Journal of Operational Research*, vol. 149, DOI 10.1016/j.energy.2018.01.187, pp. 120–134, 2017. 51, 57

[256] A. Lucas, L. Jansen, N. Andreadou, E. Kotsakis, and M. Masera, "Load Flexibility Forecast for DR Using Non-Intrusive Load Monitoring in the Residential Sector," *Energies*, vol. 12, DOI 10.3390/en12142725, no. 14, p. 2725, 2019. 51

[257] G. Krishnadas and A. Kiprakis, "A machine learning pipeline for demand response capacity scheduling," *Energies*, vol. 13, DOI 10.3390/en13071848, no. 7, pp. 1–25, 2020. 51

[258] G. Kotsis, I. Moschos, C. Corchero, and M. Cruz-Zambrano, “Demand aggregator flexibility forecast: Price incentives sensitivity assessment,” *International Conference on the European Energy Market, EEM*, vol. 2015-Augus, DOI 10.1109/EEM.2015.7216756, pp. 1–5, 2015. 51

[259] M. Vellei, J. Le Dréau, and S. Y. Abdelouadoud, “Predicting the demand flexibility of wet appliances at national level: The case of France,” *Energy and Buildings*, vol. 214, DOI 10.1016/j.enbuild.2020.109900, p. 109900, 2020. 51

[260] R. Pinto, R. J. Bessa, and M. A. Matos, “Multi-period flexibility forecast for low voltage prosumers,” *Energy*, vol. 141, DOI 10.1016/j.energy.2017.11.142, pp. 2251–2263, 2017. 51

[261] J. Ponocko and J. V. Milanovic, “Forecasting Demand Flexibility of Aggregated Residential Load Using Smart Meter Data,” *IEEE Transactions on Power Systems*, vol. 33, DOI 10.1109/TPWRS.2018.2799903, no. 5, pp. 5446–5455, 2018. 51

[262] C. Ziras, C. Heinrich, M. Pertl, and H. W. Bindner, “Experimental flexibility identification of aggregated residential thermal loads using behind-the-meter data,” *Applied Energy*, vol. 242, DOI 10.1016/j.apenergy.2019.03.156, no. March, pp. 1407–1421, 2019. 51

[263] S. Barja-Martinez, M. Aragüés-Peñalba, Í. Munné-Collado, P. Lloret-Gallego, E. Bullich-Massagué, and R. Villafafila-Robles, “Artificial intelligence techniques for enabling Big Data services in distribution networks: A review,” *Renew. Sustain. Energy Rev.*, vol. 150, DOI 10.1016/j.rser.2021.111459, p. 111459, 2021. 52, 131

[264] I. Antonopoulos, V. Robu, B. Couraud, D. Kirli, S. Norbu, A. Kiprakis, D. Flynn, S. Elizondo-Gonzalez, and S. Wattam, “Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review,” *Renewable and Sustainable Energy Reviews*, vol. 130, DOI 10.1016/j.rser.2020.109899, no. June, p. 109899, 2020. 52, 53, 54, 57

[265] H. Shareef, M. S. Ahmed, A. Mohamed, and E. Al Hassan, “Review on Home Energy Management System Considering Demand Responses, Smart Technologies, and Intelligent Controllers,” *IEEE Access*, vol. 6, DOI 10.1109/ACCESS.2018.2831917, pp. 24 498–24 509, 2018. 52

[266] A. Kumari, R. Gupta, S. Tanwar, and N. Kumar, “Blockchain and AI amalgamation for energy cloud management: Challenges, solutions, and future directions,” *Journal of Parallel and Distributed Computing*, vol. 143, DOI 10.1016/j.jpdc.2020.05.004, pp. 148–166, 2020. 52

[267] Z. Khalid, G. Abbas, M. Awais, T. Alquthami, and M. B. Rasheed, “A novel load scheduling mechanism using artificial neural network based customer profiles in smart grid,” *Energies*, vol. 13, DOI 10.3390/en13051062, no. 5, pp. 1–23, 2020. 52, 55, 56

[268] B. Yuce, Y. Rezgui, and M. Mourshed, “ANN-GA smart appliance scheduling for optimised energy management in the domestic sector,” *Energy and*

Buildings, vol. 111, DOI 10.1016/j.enbuild.2015.11.017, pp. 311–325, 2016. 52, 55

[269] F. Smarra, A. Jain, T. de Rubeis, D. Ambrosini, A. D'Innocenzo, and R. Mangharam, "Data-driven model predictive control using random forests for building energy optimization and climate control," *Applied Energy*, vol. 226, DOI 10.1016/j.apenergy.2018.02.126, no. February, pp. 1252–1272, 2018. 52, 55

[270] R. Lu, S. H. Hong, and M. Yu, "Demand Response for Home Energy Management Using Reinforcement Learning and Artificial Neural Network," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6629–6639, 2019. 53, 55, 56

[271] M. Ahmed, A. Mohamed, R. Homod, and H. Shareef, "Hybrid LSA-ANN Based Home Energy Management Scheduling Controller for Residential Demand Response Strategy," *Energies*, vol. 9, DOI 10.3390/en9090716, no. 9, p. 716, 2016. 53, 55

[272] S. Bourhnane, M. R. Abid, R. Lghoul, K. Zine-Dine, N. Elkamoun, and D. Benhaddou, "Machine learning for energy consumption prediction and scheduling in smart buildings," *SN Applied Sciences*, vol. 2, DOI 10.1007/s42452-020-2024-9, no. 2, pp. 1–10, 2020. 53, 55

[273] D. Liu, Y. Sun, Y. Qu, B. Li, and Y. Xu, "Analysis and Accurate Prediction of User's Response Behavior in Incentive-Based Demand Response," *IEEE Access*, vol. 7, DOI 10.1109/ACCESS.2018.2889500, pp. 3170–3180, 2019. 53, 55

[274] J. R. Vázquez-Canteli and Z. Nagy, "Reinforcement learning for demand response: A review of algorithms and modeling techniques," *Applied Energy*, vol. 235, DOI 10.1016/j.apenergy.2018.11.002, no. November 2018, pp. 1072–1089, 2019. 53, 54

[275] X. Huang, S. H. Hong, M. Yu, Y. DIng, and J. Jiang, "Demand Response Management for Industrial Facilities: A Deep Reinforcement Learning Approach," *IEEE Access*, vol. 7, DOI 10.1109/ACCESS.2019.2924030, pp. 82 194–82 205, 2019. 53, 55, 56

[276] F. Ruelens, B. J. Claessens, S. Quaiyum, B. De Schutter, R. Babuška, and R. Belmans, "Reinforcement Learning Applied to an Electric Water Heater: From Theory to Practice," *IEEE Transactions on Smart Grid*, vol. 9, DOI 10.1109/TSG.2016.2640184, no. 4, pp. 3792–3800, 2018. 53, 55, 56

[277] Z. Wan, H. L. H. He, and D. Prokhorov, "Model-Free Real-Time EV Charging Scheduling Based on Deep Reinforcement Learning," *IEEE Transactions on Smart Grid*, vol. 10, DOI 10.1109/TSG.2018.2879572, no. 5, pp. 5246–5257, 2018. 53, 55, 56

[278] X. Xu, Y. Jia, Y. Xu, Z. Xu, S. Chai, and C. S. Lai, "A Multi-Agent Reinforcement Learning-Based Data-Driven Method for Home Energy Management," *IEEE Transactions on Smart Grid*, vol. 11, DOI 10.1109/tsg.2020.2971427, no. 4, pp. 3201–3211, 2020. 53, 55

[279] R. Lu, S. H. Hong, and X. Zhang, "A Dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach," *Applied Energy*, vol. 220, DOI 10.1016/j.apenergy.2018.03.072, no. February, pp. 220–230, 2018. 53, 55, 56

[280] L. Wen, K. Zhou, J. Li, and S. Wang, "Modified deep learning and reinforcement learning for an incentive-based demand response model," *Energy*, vol. 205, DOI 10.1016/j.energy.2020.118019, p. 118019, 2020. 53, 55

[281] E. Mocanu, D. C. Mocanu, P. H. Nguyen, A. Liotta, M. E. Webber, M. Gibescu, and J. G. Slootweg, "On-Line Building Energy Optimization Using Deep Reinforcement Learning," *IEEE Transactions on Smart Grid*, vol. 10, DOI 10.1109/TSG.2018.2834219, no. 4, pp. 3698–3708, 2019. 53, 55, 59

[282] E. Foruzan, L. K. Soh, and S. Asgarpoor, "Reinforcement Learning Approach for Optimal Distributed Energy Management in a Microgrid," *IEEE Transactions on Power Systems*, vol. 33, DOI 10.1109/TPWRS.2018.2823641, no. 5, pp. 5749–5758, 2018. 53, 55

[283] F. Ruelens, B. J. Claessens, S. Vandael, B. De Schutter, R. Babuska, and R. Belmans, "Residential Demand Response of Thermostatically Controlled Loads Using Batch Reinforcement Learning," *IEEE Transactions on Smart Grid*, vol. 8, DOI 10.1109/TSG.2016.2517211, no. 5, pp. 2149–2159, 2017. 53, 55

[284] T. Remani, E. A. Jasmin, and T. P. Ahamed, "Residential Load Scheduling with Renewable Generation in the Smart Grid: A Reinforcement Learning Approach," *IEEE Systems Journal*, vol. 13, DOI 10.1109/JSYST.2018.2855689, no. 3, pp. 3283–3294, 2019. 53, 55, 56

[285] Sascha Lange, Thomas Gabel, M. Riedmiller, and Abstract, "Batch Reinforcement Learning Sascha," *International Conference in Machine Learning*, vol. 4, DOI 10.1038/nature14236, no. 7540, p. 14, 2015. 53

[286] A. M. Carreiro, H. M. Jorge, and C. H. Antunes, "Energy management systems aggregators: A literature survey," *Renewable and Sustainable Energy Reviews*, vol. 73, DOI 10.1016/j.rser.2017.01.179, no. December 2016, pp. 1160–1172, 2017. 56

[287] Eurelectric, "Flexibility and aggregation - requirements for their interaction in the market," *Ieee Transactions on Power Systems*, vol. 30, DOI 10.1109/TPWRS.2014.2328865, no. January, p. 13, 2015. 56

[288] J. Iria and F. Soares, "A cluster-based optimization approach to support the participation of an aggregator of a larger number of prosumers in the day-ahead energy market," *Electric Power Systems Research*, vol. 168, DOI 10.1016/j.epsr.2018.11.022, no. July 2018, pp. 324–335, 2019. 56, 57

[289] P. Olivella-Rosell, F. Rullan, P. Lloret-Gallego, E. Prieto-Araujo, R. Ferrer-San-José, S. Barja-Martinez, S. Bjarghov, V. Lakshmanan, A. Hentunen, J. Forsström, S. Ottesen, R. Villafafila-Robles, and A. Sumper, "Centralised and Distributed Optimization for Aggregated Flexibility Services Provision," pp. 1–10, 2019. 56, 57

[290] J. Iria, P. Scott, and A. Attarha, "Network-constrained bidding optimization strategy for aggregators of prosumers," *Energy*, vol. 207, DOI 10.1016/j.energy.2020.118266, 2020. 56, 57

[291] F. Shen, Q. Wu, X. Jin, B. Zhou, C. Li, and Y. Xu, "ADMM-based market clearing and optimal flexibility bidding of distribution-level flexibility market for day-ahead congestion management of distribution networks," *International Journal of Electrical Power and Energy Systems*, vol. 123, DOI 10.1016/j.ijepes.2020.106266, no. May, p. 106266, 2020. 56, 57

[292] S. Huang, Q. Wu, M. Shahidehpour, and Z. Liu, "Distributed Optimization-Based Dynamic Tariff for Congestion Management in Distribution Networks," *IEEE Transactions on Smart Grid*, vol. 10, DOI 10.1109/TSG.2018.2790638, no. 2, pp. 2148–2157, 2019. 56, 57

[293] F. Lezama, J. Soares, B. Canizes, and Z. Vale, "Flexibility management model of home appliances to support DSO requests in smart grids," *Sustainable Cities and Society*, vol. 55, DOI 10.1016/j.scs.2020.102048, no. November 2018, p. 102048, 2020. 56

[294] P. Macdougall, B. Ran, M. Klever, and G. Deconinck, "Value assessment of aggregated energy flexibility when traded on multiple markets," *International Conference on the European Energy Market, EEM*, DOI 10.1109/EEM.2017.7981960, 2017. 56

[295] S. Huang and Q. Wu, "Real-time congestion management in distribution networks by flexible demand swap," *IEEE Transactions on Smart Grid*, vol. 9, DOI 10.1109/TSG.2017.2655085, no. 5, pp. 4346–4355, 2018. 56, 57

[296] BDEW German Association of Energy and Water Industries, "Smart Grid Traffic Light Concept - Design of the amber phase," BDEW German Association of Energy and Water Industries, Berlin, Tech. Rep. March, 2015. 57

[297] P. Olivella-Rosell, P. Lloret-Gallego, Í. Munné-Collado, R. Villafafila-Robles, A. Sumper, S. Ottessen, J. Rajasekharan, and B. Bremdal, "Local Flexibility Market Design for Aggregators Providing Multiple Flexibility Services at Distribution Network Level," *Energies*, vol. 11, DOI 10.3390/en11040822, no. 4, p. 822, 2018. 57

[298] K. Paridari and L. Nordström, "Flexibility prediction, scheduling and control of aggregated TCLs," *Electric Power Systems Research*, vol. 178, DOI 10.1016/j.epsr.2019.106004, no. February 2019, p. 106004, 2020. 57

[299] C. Ziras, C. Heinrich, M. Pertl, and H. W. Bindner, "Experimental flexibility identification of aggregated residential thermal loads using behind-the-meter data," *Applied Energy*, vol. 242, DOI 10.1016/j.apenergy.2019.03.156, no. October 2018, pp. 1407–1421, 2019. 57

[300] K. Amasyali, M. Olama, and A. Perumalla, "A machine learning-based approach to predict the aggregate flexibility of HVAC systems," *2020 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2020*, DOI 10.1109/ISGT45199.2020.9087695, pp. 0–4, 2020. 57

[301] M. Short, S. Rodriguez, R. Charlesworth, T. Crosbie, and N. Dawood, "Optimal dispatch of aggregated HVAC units for demand response: An industry 4.0 approach," *Energies*, vol. 12, DOI 10.3390/en12224320, no. 22, 2019. 57

[302] C. A. Correa-Florez, A. Michiorri, and G. Kariniotakis, "Comparative analysis of adjustable robust optimization alternatives for the participation of aggregated residential prosumers in electricity markets," *Energies*, vol. 12, DOI 10.3390/en12061019, no. 6, 2019. 57

[303] A. Camero, G. Luque, Y. Bravo, and E. Alba, "Customer segmentation based on the electricity demand signature: The andalusian case," *Energies*, vol. 11, DOI 10.3390/en11071788, no. 7, pp. 1–14, 2018. 57

[304] I. R. U. Qhuj, H. Xvlqj, L. Q. J. Rulv, and H. H. Oh, "Comparison of Clustering Techniques for Residential Energy Behavior using Smart Meter Data," in *AAAI Workshops - Artificial Intelligence for Smart Grids and Buildings, March 2017*, pp. 260–266, San Francisco, CA, 2017. 57

[305] A. Rajabi, M. Eskandari, M. J. Ghadi, L. Li, J. Zhang, and P. Siano, "A comparative study of clustering techniques for electrical load pattern segmentation," *Renewable and Sustainable Energy Reviews*, vol. 120, DOI 10.1016/j.rser.2019.109628, no. December 2019, 2020. 57

[306] L. Hernández, C. Baladrón, J. M. Aguiar, B. Carro, and A. Sánchez-Esguevillas, "Classification and clustering of electricity demand patterns in industrial parks," *Energies*, vol. 5, DOI 10.3390/en5125215, no. 12, pp. 5215–5228, 2012. 57

[307] P. Siano, G. De Marco, A. Rolan, and V. Loia, "A Survey and Evaluation of the Potentials of Distributed Ledger Technology for Peer-to-Peer Transactive Energy Exchanges in Local Energy Markets," *IEEE Systems Journal*, vol. 13, DOI 10.1109/JSYST.2019.2903172, no. 3, pp. 3454–3466, 2019. 58

[308] M. B. Luigi D'Oriano, Giuseppe Mastandrea, Giuseppe Rana, Giuseppe Raveduto, Vincenzo Croce, Matteo Verber, "Smart Grids : Requirements engineering and use cases," *IEEE International Conference and Workshop in Óbuda on Electrical and Power Engineering*, pp. 39–44, 2020. 58

[309] P. Ratanasopitkul, "Blockchain - Revolutionize Green Energy Management," *Proceedings of the Conference on the Industrial and Commercial Use of Energy, ICUE*, vol. 2018-Octob, DOI 10.23919/ICUE-GESD.2018.8635666, no. October, pp. 1–6, 2019. 58

[310] C. Pop, T. Cioara, M. Antal, I. Anghel, I. Salomie, and M. Bertoncini, "Blockchain based decentralized management of demand response programs in smart energy grids," *Sensors (Switzerland)*, vol. 18, DOI 10.3390/s18010162, no. 1, 2018. 58

[311] M. Troncia, M. Galici, M. Mureddu, E. Ghiani, and F. Pilo, "Distributed ledger technologies for peer-to-peer local markets in distribution networks," *Energies*, vol. 12, DOI 10.3390/en12173249, no. 17, 2019. 58

[312] K. Zhang, S. Troitzsch, S. Hanif, and T. Hamacher, "Coordinated Market Design for Peer-to-Peer Energy Trade and Ancillary Services in

Distribution Grids," *IEEE Transactions on Smart Grid*, vol. 11, DOI 10.1109/TSG.2020.2966216, no. 4, pp. 2929–2941, 2020. 58

[313] M. Foti, D. Greasidis, and M. Vavalis, "Viability analysis of a decentralized energy market based on blockchain," *International Conference on the European Energy Market, EEM*, vol. 2018-June, DOI 10.1109/EEM.2018.8469906, pp. 1–5, 2018. 58

[314] M. A. Ferrag and L. Maglaras, "DeepCoin: A Novel Deep Learning and Blockchain-Based Energy Exchange Framework for Smart Grids," *IEEE Transactions on Engineering Management*, vol. 67, DOI 10.1109/TEM.2019.2922936, no. 4, pp. 1285–1297, 2020. 58

[315] S. Zhou, Z. Hu, W. Gu, M. Jiang, and X.-P. Zhang, "Artificial intelligence based smart energy community management: A reinforcement learning approach," *CSEE Journal of Power and Energy Systems*, vol. 5, DOI 10.17775/cseejpes.2018.00840, no. 1, pp. 1–10, 2019. 59

[316] T. Chen and S. Bu, "Realistic Peer-to-Peer Energy Trading Model for Microgrids using Deep Reinforcement Learning," *Proceedings of 2019 IEEE PES Innovative Smart Grid Technologies Europe, ISGT-Europe 2019*, DOI 10.1109/ISGTEurope.2019.8905731, 2019. 59

[317] F. Jamil, N. Iqbal, Imran, S. Ahmad, and D. H. Kim, "Peer-to-Peer Energy Trading Mechanism based on Blockchain and Machine Learning for Sustainable Electrical Power Supply in Smart Grid," *IEEE Access*, DOI 10.1109/ACCESS.2021.3060457, pp. 39 193–39 217, 2021. 59

[318] K. Chen, J. Lin, and Y. Song, "Trading strategy optimization for a prosumer in continuous double auction-based peer-to-peer market: A prediction-integration model," *Applied Energy*, vol. 242, DOI 10.1016/j.apenergy.2019.03.094, no. March, pp. 1121–1133, 2019. 59

[319] V. D. Mei, "Paper 0303 ARTIFICIAL INTELLIGENCE FOR MICROGRID PLANNING CIRED Workshop - Ljubljana , 7-8 June 2018 Paper 0303," no. 0303, pp. 7–8, 2018. 59

[320] X. Fu, Q. Guo, and H. Sun, "Statistical Machine Learning Model for Stochastic Optimal Planning of Distribution Networks Considering a Dynamic Correlation and Dimension Reduction," *IEEE Transactions on Smart Grid*, vol. 11, DOI 10.1109/tsg.2020.2974021, no. 4, pp. 2904–2917, 2020. 59

[321] M. Resener, S. Haffner, L. A. Pereira, and P. M. Pardalos, "Optimization techniques applied to planning of electric power distribution systems: a bibliographic survey," *Energy Systems*, vol. 9, DOI 10.1007/s12667-018-0276-x, no. 3, pp. 473–509, 2018. 59

[322] S. Klyapovskiy, S. You, H. Cai, and H. W. Bindner, "Incorporate flexibility in distribution grid planning through a framework solution," *International Journal of Electrical Power and Energy Systems*, vol. 111, DOI 10.1016/j.ijepes.2019.03.069, no. February, pp. 66–78, 2019. 59

[323] J. Contreras, M. Asensio, P. Meneses, G. Munoz-Delgado, and S. Montoyo-Bueno, *Joint RES and Distribution Network Expansion Planning under a Demand Response Framework*, 2016. 59

[324] National Renewable Energy Laboratory and Arara Blue Energy Group, "VOICES of Experience - Leveraging AMI Networks and Data," U.S. Department of Energy, Tech. Rep., 2019. [Online]. Available: https://www.smartgrid.gov/document/VOE{_}Leveraging{_}AMI{_}Networks{_}Data 66

[325] R. Gold, C. Waters, and D. York, "Leveraging Advanced Metering Infrastructure to Save Energy," American Council for an Energy-Efficient Economy (ACEEE), Tech. Rep., 2020. [Online]. Available: https://aceee.org/research-report/u2001 66

[326] S. Sagiroglu, R. Terzi, Y. Canbay, and I. Colak, "Big data issues in smart grid systems," *2016 IEEE International Conference on Renewable Energy Research and Applications, ICRERA 2016*, vol. 5, DOI 10.1109/ICRERA.2016.7884486, pp. 1007–1012, 2017. 66

[327] Big Data Value Association (BDVA), "European Big Data Value Strategic Research and Innovation Agenda - BDV SRIA," Tech. Rep., 2017. [Online]. Available: www.bdva.eu/sites/default/files/BDVA{_}SRIA{_}v4{_}Ed1.1.pdf 67

[328] Expert Group for Standards and Interoperability for Smart Grids Deployment, "Interoperability, Standards and Functionalities applied in the large scale roll out of smart metering," *Smart Grids Task Force*, no. October, pp. 1–31, 2015. 67

[329] European Network of Transmission System Operators for Electricity (ENTSO-E), "Annual Report 2015: Electricity without borders," p. 22, 2015. 68

[330] M. Chui, J. Manyika, M. Miremadi, N. Henke, R. Chung, P. Nel, and S. Malhotra, "Notes from the AI Frontier: Applications and Value of Deep Learning," 2018. 68

[331] M. Kufeoglu, Sinan ; Gaomin Liu; Pollitt, "Digitalisation and New Business Models in Energy Sector," *EPRG Working Paper . Energy Policy Research Group, University of Cambridge.*, vol. 1920, no. June, pp. 1–39, 2019. 68

[332] C. o. t. E. Union and E. Parliament, "Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union," 2016. [Online]. Available: https://op.europa.eu/en/publication-detail/-/publication/d2912aca-4d75-11e6-89bd-01aa75ed71a1 69

[333] European Parliament and Council of the European Union, "Regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (Data Protection Directive)," pp. 1–88, 2016. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02016R0679-20160504{&}from=EN 69

[334] D. Markopoulou, V. Papakonstantinou, and P. de Hert, "The new EU cybersecurity framework: The NIS Directive, ENISA's role and the General Data Protection Regulation," *Computer Law and Security Review*, vol. 35, DOI 10.1016/j.clsr.2019.06.007, no. 6, p. 105336, 2019. 69

[335] International Energy Agency, "Global CO2 emissions in 2019 – Analysis," https://www.iea.org/articles/global-co2-emissions-in-2019, 2020, [Online; accessed September-2020]. 73, 74

[336] DNV GL, "Energy transition Outlook 2019 - A global and regional forecast to 2050." DNV GL's Group Technology & Research, Tech. Rep., 2019. 73

[337] IEA, "Perspectives for the Clean Energy Transition- The Critical Role of Buildings," p. 117, 2019. 74

[338] Y. Kwon, T. Kim, K. Baek, and J. Kim, "Multi-objective optimization of home appliances and electric vehicle considering customer's benefits and offsite shared photovoltaic curtailment," *Energies*, vol. 13, DOI 10.3390/en13112852, no. 11, 2020. 74, 75

[339] J. M. Veras, I. R. S. Silva, P. R. Pinheiro, R. A. Rabêlo, A. F. S. Veloso, F. A. S. Borges, and J. J. Rodrigues, "A multi-objective demand response optimization model for scheduling loads in a home energy management system," *Sensors (Switzerland)*, vol. 18, DOI 10.3390/s18103207, no. 10, 2018. 74, 75

[340] R. Reghukumar, S. Sambhu, and V. Ravikumar Pandi, "Multi-objective optimization for efficient home energy management system using differential evolution algorithm," *3rd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, RTEICT 2018 - Proceedings*, DOI 10.1109/RTEICT42901.2018.9012604, pp. 1157–1162, 2018. 74, 75

[341] X. Wang, X. Mao, and H. Khodaei, "A multi-objective home energy management system based on internet of things and optimization algorithms," *Journal of Building Engineering*, vol. 33, DOI 10.1016/j.jobe.2020.101603, no. May 2020, p. 101603, 2021. 74, 75, 135

[342] M. C. Di Piazza, G. La Tona, M. Luna, and A. Di Piazza, "A two-stage Energy Management System for smart buildings reducing the impact of demand uncertainty," *Energy and Buildings*, vol. 139, DOI 10.1016/j.enbuild.2017.01.003, pp. 1–9, 2017. 74

[343] G. Huang, J. Yang, and C. Wei, "Cost-effective and comfort-aware electricity scheduling for home energy management system," *Proceedings - 2016 IEEE International Conferences on Big Data and Cloud Computing, BDCloud 2016*, DOI 10.1109/BDCloud-SocialCom-SustainCom.2016.73, pp. 453–460, 2016. 74

[344] K. Dittawit and F. A. Aagesen, "Home energy management system for electricity cost savings and comfort preservation," *IEEE International Conference on Consumer Electronics - Berlin, ICCE-Berlin*, vol. 2015-Febru, DOI 10.1109/ICCE-Berlin.2014.7034267, no. February, pp. 309–313, 2015. 74

[345] N. Javaid, F. Ahmed, I. Ullah, S. Abid, W. Abdul, A. Alamri, and A. S. Almogren, "Towards cost and comfort based hybrid optimization for residential load scheduling in a smart grid," *Energies*, vol. 10, DOI 10.3390/en10101546, no. 10, pp. 1–27, 2017. 74

[346] M. Waseem, Z. Lin, S. Liu, I. A. Sajjad, and T. Aziz, "Optimal GWCSO-based home appliances scheduling for demand response considering end-users comfort," *Electric Power Systems Research*, vol. 187, DOI 10.1016/j.epsr.2020.106477, no. January, p. 106477, 2020. 74

[347] T. Sattarpour, D. Nazarpour, and S. Golshannavaz, "A multi-objective HEM strategy for smart home energy scheduling: A collaborative approach to support microgrid operation," *Sustainable Cities and Society*, vol. 37, DOI 10.1016/j.scs.2017.09.037, no. September 2017, pp. 26–33, 2018. 74, 75

[348] M. A. Fotouhi Ghazvini, J. Soares, O. Abrishambaf, R. Castro, and Z. Vale, "Demand response implementation in smart households," *Energy and Buildings*, vol. 143, DOI 10.1016/j.enbuild.2017.03.020, pp. 129–148, 2017. 74, 75

[349] S. Barja-Martinez, I. Munne-Collado, P. Lloret-Gallego, M. Aragues-Penalba, and R. Villafafila-Robles, "A Novel Home Energy Management System Environmental-based with LCA Minimization," in *2020 IEEE International Conference on Environment and Electrical Engineering (EEEIC / I&CPS Europe)*, DOI 10.1109/eeeic/icpseurope49358.2020.9160619, pp. 1–6. Madrid, Spain: IEEE procedings, 2020. 74, 75, 79

[350] T. Xing, H. Lin, Z. Tan, and L. Ju, "Coordinated energy management for micro energy systems considering carbon emissions using multi-objective optimization," *Energies*, vol. 12, DOI 10.3390/en12234414, no. 23, 2019. 74, 75

[351] A. Mahmood, F. Baig, N. Alrajeh, U. Qasim, Z. A. Khan, and N. Javaid, "An enhanced system architecture for optimized demand side management in smart grid," *Applied Sciences*, vol. 6, DOI 10.3390/app6050122, no. 5, 2016. 75

[352] X. Jin, K. Baker, D. Christensen, and S. Isley, "Foresee: A user-centric home energy management system for energy efficiency and demand response," *Applied Energy*, vol. 205, DOI 10.1016/j.apenergy.2017.08.166, no. June, pp. 1583–1595, 2017. 75

[353] M. Shakeri, M. Shayestegan, H. Abunima, S. M. Reza, M. Akhtaruzzaman, A. R. Alamoud, K. Sopian, and N. Amin, "An intelligent system architecture in home energy management systems (HEMS) for efficient demand response in smart grid," *Energy and Buildings*, vol. 138, DOI 10.1016/j.enbuild.2016.12.026, pp. 154–164, 2017. 75

[354] H. T. Nguyen, D. T. Nguyen, and L. B. Le, "Energy management for households with solar assisted thermal load considering renewable energy and price uncertainty," *IEEE Transactions on Smart Grid*, vol. 6, DOI 10.1109/TSG.2014.2350831, no. 1, pp. 301–314, 2015. 75

[355] F. Wang, B. Xiang, K. Li, X. Ge, H. Lu, J. Lai, and P. Dehghanian, "Smart households' aggregated capacity forecasting for load aggregators under incentive-based demand response programs," *IEEE Transactions on Industry Applications*, vol. 56, DOI 10.1109/TIA.2020.2966426, no. 2, pp. 1086–1097, 2020. 75

[356] M. L. DI Silvestre, P. Gallo, E. R. Sanseverino, G. Sciume, and G. Zizzo, "Aggregation and Remuneration in Demand Response with a Blockchain-Based Framework," *IEEE Transactions on Industry Applications*, vol. 56, DOI 10.1109/TIA.2020.2992958, no. 4, pp. 4248–4257, 2020. 75

[357] I. Munné-Collado, F. M. Aprà, P. Olivella-Rosell, and R. Villafáfila-Robles, "The potential role of flexibility during peak hours on greenhouse gas emissions: A life cycle assessment of five targeted national electricity grid mixes," *Energies*, vol. 12, DOI 10.3390/en12234443, no. 23, 2019. 78

[358] E. Redondo-Iglesias, P. Venet, and S. Pelissier, "Modelling lithium-ion battery ageing in electric vehicle applications—calendar and cycling ageing combination effects," *Batteries*, vol. 6, DOI 10.3390/batteries6010014, no. 1, pp. 1–18, 2020. 81

[359] J. Schmalstieg, S. Käbitz, M. Ecker, and D. U. Sauer, "A holistic aging model for Li(NiMnCo)O2 based 18650 lithium-ion batteries," *Journal of Power Sources*, vol. 257, DOI 10.1016/j.jpowsour.2014.02.012, pp. 325–334, 2014. 81, 82

[360] F. Rücker, I. Bremer, S. Linden, J. Badeda, and D. U. Sauer, "Development and Evaluation of a Battery Lifetime Extending Charging Algorithm for an Electric Vehicle Fleet," *Energy Procedia*, vol. 99, DOI 10.1016/j.egypro.2016.10.118, pp. 285–291, 2016. 82, 83

[361] P. S. Inc., "Dataport database," https://www.pecanstreet.org/dataport/, 2020, [Online; accessed July-2020]. 87

[362] "Energy and the green deal." [Online]. Available: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal/energy-and-green-deal_en 101

[363] "Comunidades energéticas — idae," https://www.idae.es/ayudas-y-financiacion/comunidades-energeticas, (Accessed on 04/08/2023). 102

[364] E. Commission and D.-G. for Energy, *Clean energy for all Europeans*. Publications Office, 2019. 103

[365] A. European Commision. Caramizaru and A. Uihlein, *Energy communities: an overview of energy and social innovation Social aspects of the energy transition*, 2020. [Online]. Available: https://ec.europa.eu/jrc 103

[366] "Directive (eu) 2019/ 944 of the european parliament and of the council - of 5 june 2019 - on common rules for the internal market for electricity and amending directive 2012/ 27/ eu," https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L0944, (Accessed on 04/08/2023). 103

[367] "Directive (eu) 2018/ 2001 of the european parliament and of the council - of 11 december 2018 - on the promotion of the use of energy from renewable sources," https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018L2001, (Accessed on 04/08/2023). 103

[368] M. E. Biresselioglu, S. A. Limoncuoglu, M. H. Demir, J. Reichl, K. Burgstaller, A. Sciullo, and E. Ferrero, "Legal provisions and market conditions for energy

communities in austria, germany, greece, italy, spain, and turkey: A comparative assessment," *Sustain.*, vol. 13, DOI 10.3390/su132011212, no. 20, 2021. 104

[369] "Real decreto 244/2019 autoconsumo de energía eléctrica. disposición 5089 del boe núm. 83 de 2019," https://www.boe.es/boe/dias/2019/04/06/pdfs/BOE-A-2019-5089.pdf, (Accessed on 04/08/2023). 105, 115

[370] H. Abb and P. Grids, "Digital transformation of the distribution grid." 131

[371] F. Tounquet and C. Alaton, *Benchmarking Smart Metering Deployment in EU-28*, no. December, 2020. [Online]. Available: https://op.europa.eu/en/publication-detail/-/publication/b397ef73-698f-11ea-b735-01aa75ed71a1/language-en 132

[372] European Parliament and Council of the European Union, "Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation)," Official Journal of the European Union, 2016. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj 132, 133

[373] Y. Wang, I. L. Bennani, X. Liu, M. Sun, and Y. Zhou, "Electricity Consumer Characteristics Identification: A Federated Learning Approach," *IEEE Trans. Smart Grid*, vol. 12, DOI 10.1109/TSG.2021.3066577, no. 4, pp. 3637–3647, 2021. 132, 135, 140

[374] M. Jia, Y. Wang, C. Shen, and G. Hug, "Privacy-Preserving Distributed Clustering for Electrical Load Profiling," *IEEE Trans. Smart Grid*, vol. 12, DOI 10.1109/TSG.2020.3031007, no. 2, pp. 1429–1444, 2021. 132

[375] Y. Li, J. Li, and Y. Wang, "Privacy-Preserving Spatiotemporal Scenario Generation of Renewable Energies: A Federated Deep Generative Learning Approach," *IEEE Trans. Ind. Informatics*, vol. 18, DOI 10.1109/TII.2021.3098259, no. 4, pp. 2310–2320, 2022. 132

[376] Y. L. Tun, K. Thar, C. M. Thwal, and C. S. Hong, "Federated learning based energy demand prediction with clustered aggregation," *Proc. - 2021 IEEE Int. Conf. Big Data Smart Comput. BigComp 2021*, DOI 10.1109/BigComp51126.2021.00039, pp. 164–167, 2021. 132, 135, 140

[377] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Aguera y Arcas, "Federated learning of deep networks using model averaging," *arXiv preprint arXiv:1602.05629*, 2016. 132

[378] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings *et al.*, "Advances and open problems in federated learning," *arXiv preprint arXiv:1912.04977*, 2019. 132

[379] U. Zafar, S. Bayhan, and A. Sanfilippo, "Home Energy Management System Concepts, Configurations, and Technologies for the Smart Grid," *IEEE Access*, vol. 8, DOI 10.1109/ACCESS.2020.3005244, pp. 119 271–119 286, 2020. 132

[380] Y. Zhao, K. Zhou, Y. Li, Y. Qian, X. Lu, and J. Zhang, "A survey on federated learning in energy systems," *Applied Energy*, vol. 279, p. 115807, 2020. 132

[381] Y. Wang, "Personalized federated learning for individual consumer load forecasting," *CSEE J. Power Energy Syst.*, DOI 10.17775/cseejpes.2021.07350, no. February, 2022. 132, 136

[382] M. H. Albadi and E. F. El-Saadany, "Demand response in electricity markets: An overview," in *2007 IEEE Power Engineering Society General Meeting*, DOI 10.1109/PES.2007.385728, pp. 1–5, 2007. 132

[383] C. W. Granger, "Outline of forecast theory using generalized cost functions," *Spanish Economic Review*, vol. 1, DOI 10.1007/s101080050007, pp. 161–173, 1999. 133, 140

[384] J. Zhang, Y. Wang, and G. Hug, "Cost-oriented load forecasting," *Electr. Power Syst. Res.*, vol. 205, DOI 10.1016/j.epsr.2021.107723, no. October 2021, p. 107723, 2022. 133, 140, 141

[385] G. Li and H. D. Chiang, "Toward Cost-Oriented Forecasting of Wind Power Generation," *IEEE Trans. Smart Grid*, vol. 9, DOI 10.1109/TSG.2016.2614341, no. 4, pp. 2508–2517, 2018. 133, 140

[386] Y. Wang and L. Wu, "Improving economic values of day-ahead load forecasts to real-time power system operations," *IET Generation, Transmission & Distribution*, vol. 11, DOI https://doi.org/10.1049/iet-gtd.2017.0517, no. 17, pp. 4238–4247, 2017. 133

[387] A. Tajer, S. M. Perlaza, and H. V. Poor, *Advanced Data Analytics for Power Systems*, 2021. 133

[388] T. Zhang, L. Gao, C. He, M. Zhang, B. Krishnamachari, and A. S. Avestimehr, "Federated learning for the internet of things: Applications, challenges, and opportunities," *IEEE Internet of Things Magazine*, vol. 5, DOI 10.1109/IOTM.004.2100182, no. 1, pp. 24–29, 2022. 133

[389] S. Banabilah, M. Aloqaily, E. Alsayed, N. Malik, and Y. Jararweh, "Federated learning review: Fundamentals, enabling technologies, and future applications," *Inf. Process. Manag.*, vol. 59, DOI 10.1016/j.ipm.2022.103061, no. 6, p. 103061, 2022. 134

[390] S. Lee and D.-H. Choi, "Federated reinforcement learning for energy management of multiple smart homes with distributed energy resources," *IEEE Transactions on Industrial Informatics*, vol. 18, DOI 10.1109/TII.2020.3035451, no. 1, pp. 488–497, 2022. 134

[391] R. Firouzi, R. Rahmani, and T. Kanter, "Federated learning for distributed reasoning on edge computing," *Procedia Comput. Sci.*, vol. 184, DOI 10.1016/j.procs.2021.03.053, no. 2020, pp. 419–427, 2021. 135

[392] R. A. Sater and A. B. Hamza, "A Federated Learning Approach to Anomaly Detection in Smart Buildings," *ACM Trans. Internet Things*, vol. 2, DOI 10.1145/3467981, no. 4, pp. 1–23, 2021. 135

[393] B. Shubyn, D. Kostrzewa, P. Grzesik, P. Benecki, T. Maksymyuk, V. Sunderam, J.-H. Syu, J. C.-W. Lin, and D. Mrozek, "Federated learning for improved prediction of failures in autonomous guided vehicles," *Journal of Computational Science*, DOI https://doi.org/10.1016/j.jocs.2023.101956, p. 101956, 2023. 135

[394] J. Lin, J. Ma, and J. Zhu, "Privacy-Preserving Household Characteristic Identification With Federated Learning Method," *IEEE Trans. Smart Grid*, vol. 13, DOI 10.1109/TSG.2021.3125677, no. 2, pp. 1088–1099, 2022. 135, 140

[395] A. Taik and S. Cherkaoui, "Electrical Load Forecasting Using Edge Computing and Federated Learning," *IEEE Int. Conf. Commun.*, vol. 2020-June, DOI 10.1109/ICC40277.2020.9148937, 2020. 135, 136, 137, 139

[396] N. Gholizadeh and P. Musilek, "Federated learning with hyperparameter-based clustering for electrical load forecasting," *Internet of Things (Netherlands)*, vol. 17, DOI 10.1016/j.iot.2021.100470, no. November 2021, p. 100470, 2022. 135, 140

[397] Y. Li, X. Li, G. Li, and Z. Li, "Privacy Protection in Prosumer Energy Management Based on Federated Learning," *IEEE Access*, vol. 9, DOI 10.1109/ACCESS.2021.3053573, no. Iid, pp. 16 707–16 715, 2021. 135

[398] M. Wen, R. Xie, K. Lu, L. Wang, and K. Zhang, "FedDetect: A Novel Privacy-Preserving Federated Learning Framework for Energy Theft Detection in Smart Grid," *IEEE Internet Things J.*, vol. 9, DOI 10.1109/JIOT.2021.3110784, no. 8, pp. 6069–6080, 2022. 135

[399] Y. M. Saputra, D. T. Hoang, D. N. Nguyen, E. Dutkiewicz, M. D. Mueck, and S. Srikanteswara, "Energy demand prediction with federated learning for electric vehicle networks," *2019 IEEE Glob. Commun. Conf. GLOBECOM 2019 - Proc.*, DOI 10.1109/GLOBECOM38437.2019.9013587, 2019. 135

[400] Z. Teimoori, A. Yassine, and M. Shamim Hossain, "A Secure Cloudlet-Based Charging Station Recommendation for Electric Vehicles Empowered by Federated Learning," *IEEE Trans. Ind. Informatics*, vol. 18, DOI 10.1109/TII.2022.3148997, no. 9, pp. 6464–6473, 2022. 135

[401] J. Lin, J. Ma, and J. Zhu, "A Privacy-Preserving Federated Learning Method for Probabilistic Community-Level Behind-the-Meter Solar Generation Disaggregation," *IEEE Trans. Smart Grid*, vol. 13, DOI 10.1109/TSG.2021.3115904, no. 1, pp. 268–279, 2022. 135, 136

[402] X. Zhang, F. Fang, and J. Wang, "Probabilistic solar irradiation forecasting based on variational bayesian inference with secure federated learning," *IEEE Transactions on Industrial Informatics*, vol. 17, DOI 10.1109/TII.2020.3035807, no. 11, pp. 7849–7859, 2021. 135

[403] T. Toubeau, Jean-François; Teng, Fei; Wang, Yi; Von Krannichfeldt, Leandro; Morstyn, "Privacy-Preserving Probabilistic Voltage Forecasting in Local Energy Communities," *TechRxiv*, DOI 10.1109/TSG.2022.3187557, no. January, pp. 0–10, 2022. 135

[404] Y. Wang and L. Wu, "Improving economic values of day-ahead load forecasts to real-time power system operations," *IET Gener. Transm. Distrib.*, vol. 11, DOI 10.1049/iet-gtd.2017.0517, no. 17, pp. 4238–4247, 2017. 140, 141

[405] "Smart meters in london," www.kaggle.com/datasets/, accessed: 2022-10-01. 144

[406] S. Barja-Martinez, F. Rücker, and M. Aragüés-Peñalba, "A novel hybrid home energy management system considering electricity cost and greenhouse gas emissions minimization," *IEEE Transactions on Industry Applications*, vol. 57, DOI 10.1109/TIA.2021.3057014, no. 3, pp. 2782–2790, 2021. 146

[407] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *NIPS-W*, 2017. 146

[408] A. Ziller and A. T. et al., "Pysyft: A library for easy federated learning," 2021. 146

[409] F. Pedregosa and V. et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. 148

[410] D. J. Stekhoven and P. Buehlmann, "Missforest - non-parametric missing value imputation for mixed-type data," *Bioinformatics*, vol. 28, no. 1, pp. 112–118, 2012. 213

Appendix A

Publications

This appendix presents the journal articles and conference proceedings that the author has contributed to. Some of these publications are directly related to the chapters of this thesis, while others are from additional research in which the author has participated.

Journal papers

Included in the thesis

J1 S. Barja-Martinez, M. Aragüés-Peñalba, Í. Munné-Collado, P. Lloret-Gallego, E. Bullich-Massagué, R. Villafafila-Robles, "Artificial intelligence techniques for enabling Big Data services in distribution networks: A review", *Renewable and Sustainable Energy Reviews*, Volume 150, 2021.

J2 S. Barja-Martinez, F. Rücker, M. Aragüés-Peñalba, R. Villafafila-Robles, Í. Munné-Collado and P. Lloret-Gallego,"A Novel Hybrid Home Energy Management System Considering Electricity Cost and Greenhouse Gas Emissions Minimization", in *IEEE Transactions on Industry Applications*, vol. 57, no. 3, pp. 2782-2790, May-June 2021.

J3 S. Barja-Martinez, Fei Teng, Adrià Junyent-Ferré and Mònica Aragüés-Peñalba. "Personalized Federated Learning with Cost-Oriented Load Forecasting for Home Energy Management Systems", *IEEE Transactions on Smart Grids*, Submitted, March 2023.

J4 S. Barja-Martinez, A. Sumper, R. Villafáfila-Robles, M. Aragüés-Peñalba, D. Heredero-Peris and J.F Forero-Quintero "Local Energy Community Centralized-Approach with Cost and $C0_2$ minimization considering PV and Storage Sharing", May 2023, Pending to submit.

Not included in the thesis

J5 P. Olivella-Rosell (...) S. Barja-Martinez et al., "Centralised and Distributed Optimization for Aggregated Flexibility Services Provision," March 2023, submitted to: *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3257-3269, July 2020.

J6 J.F. Forero-Quintero, R. Villafáfila-Robles, S. Barja-Martinez, I. Munné-Collado, P. Olivella-Rosell, D. Montesinos-Miracle, "Profitability analysis on demand-side flexibility: A review", *Renewable and Sustainable Energy Reviews*, Volume 169, 2022.

J7 Adriano Caprara, S. Barja-Martinez, Mònica Aragués Peñalba "Flexibility Services for a Balance Responsible Party : A Day-ahead Portfolio Optimization Approach", March 2023, Submitted to: *IEEE Transactions on Smart Grids.*

J8 J. F. Forero-Quintero, R. Villafafila-Robles, D. Montesinos-Miracle, S. Barja-Martinez, M. Codina-Escolar "A Flexibility Management System for Behind-the-meter Flexibiity with Distributed Energy Resources: A Sensitivity Analysis", March 2023, Submitted to: *Sustainable Energy Technologies and Assessments, Elsevier.*

Conference papers

Included in the thesis

C1 Sara Barja-Martinez, Pol Olivella-Rosell, Pau Lloret-Gallego and Roberto Villafafila-Robles "Intelligent Flexibility Management for Prosumers: Development of Algorithms for the Energy Management of Electric Vehicles, Loads, Generators and Batteries", *I Congreso Iberoamericano de Ciudades Inteligentes (ICSC-CITIES 2018).* Soria, Spain. September 2018.

C2 S. Barja-Martinez, I. Munné-Collado, P. Lloret-Gallego, M. Aragüés and R. Villafafila-Robles, "A Novel Home Energy Management System Environmental-based with LCA Minimization" *2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)*, Madrid, Spain, 2020.

Not included in the thesis

C3 P. Olivella-Rosell, P. Lloret-Gallego,S. Barja-Martinez and R. Villafafila-Robles. "How to provide flexibility to prosumers, BRPs and DSOs from storage and electric vehicles?". XI International Conference on Energy Innovation, Barcelona, Spain. November 2018.

C4 Sara Barja-Martinez, Pol Olivella-Rosell, Pau Lloret-Gallego, Roberto Villafafila-Robles "Centralized flexibility services for Distribution System Operators through distributed flexible resources", *II Congreso Iberoamericano de Ciudades Inteligentes* (ICSC-CITIES 2019). Soria, Spain. October 2019.

C5 S. Barja-Martinez, P. Olivella-Rosell, P. Lloret-Gallego, R. Villafafila-Robles and A. Sumper, "A scheduling optimization model of electric water heaters for electricity cost minimization with limited information". *IEEE International Conference on Modern Power Systems* 2019 (MPS2019), Cluj-Napoca, Romania. May 2019.

C6 P. Olivella-Rosell, P. Lloret-Gallego, S. Barja-Martinez, S. Bjarghov, V. Lakshmanan, S. Odegaard Ottesen, N. Refas, F. Geerts, R. Villafafila-Robles and F. Diaz-Gonzalez "INVADE Flexibility Centralized Algorithm to Manage Electric Vehicles under DSO Requests in Buildings with Limited Information", Journal of transactions of Smart Grid". *ISGT Europe conference*, September 2019.

C7 V. Palma Costa, R. Gallart Fernandez, P. Lloret-Gallego, P. Olivella-Rosell, S. Barja-Martinez, I. Munne Collado, R. Villafafila-Robles, "Servicios de flexibilidad para el dso y brp: el piloto espanol del proyecto invade", *VI Congreso Smart Grids*. Madrid, Spain. Diciembre 2019.

Conferences oral presentations

P-C1 Presentation of "A scheduling optimization model of electric water heaters for electricity cost minimization with limited information" in *IEEE International Conference on Modern Power Systems*, Cluj-Napoca, Romania. May 2019.

P-C2 Presentation of "A Novel Home Energy Management System Environmental-based with LCA Minimization" in *IEEE International Conference on Environment and Electrical Engineering (EEEIC)*, Madrid, Spain, June 2020.

AppendixAppendix B

Appendix B

Detection of measurement errors tool

In this thesis, a tool for error detection is developed to avoid errors when using raw datasets. The tool is designed to identify and solve data anomalies, such as outliers or missing values in the raw data sources used in this study, in order to guarantee optimal data quality and algorithm effectiveness. Depending on the anomaly detected, an automatic correction is performed according to the preferences set, using analytical tools ranging from Statistics to Artificial Intelligence, including Big Data processing techniques. Additionally, it is possible to report the detected anomalies. This appendix provides a brief overview of this tool and its functionalities.

B.1 Tool description

For the deployment of smart grids, a large amount of metering and sensing devices are being installed at different voltage levels, exponentially increasing the amount of data available. These data are vital for the operation and control of the grid; however, they may contain anomalies such as outliers or missing records. Therefore, this tool ensures high data quality during the operation of power system-related services performed in this thesis.

The Detection of Measurement Errors service (from now on, DoME) offers the possibility of analyzing time-series data sources coming from a great variety of fields and identifying eventual anomalies through the automatic application of Machine Learning, Statistics and Big Data processing methods. Depending on the user's preferences and the type of anomaly detected, a specific correction is implemented in the data set. This ensures accurate, insightful, and high-quality input data for services utilizing DoME. DoME works on-demand when the energy service starts running and gets the input data. When imputing values to missing data, random forest and K-Nearest Neighbors (KNN) techniques are used. This tool is also designed to be used with Spark, as presented in Figure B.1. This service can clean and process large amounts of data, thus enabling data scalability.

The most suitable technique to address missing data imputations or anomalies might vary depending on the service requirements. AI methods are a powerful tool for assigning predictive values in place of missing values. The main advantage of using AI and Big Data analysis is that it automates and improves the error detection process of the ever-increasing energy-related measured data. However, in certain cases when the data needs to be imputed and analyzed fast, Statistics benchmark methods like mean and median are applied.

B.2 Methodology for the tool development

The DoME tool follows the workflow presented in Figure B.1. The first diagram shows the available options for non-Big Data files cleaning process. The algorithm makes sure that no data is missing in the case of time series. However, if the dataset needed is used for training tasks (data for training models option), the missing data rows are eliminated instead of imputing a non-true value since the missing data records account for a small amount of the total. In data for operation, the next step notifies the service user of the row/columns with missing data, if requested in the DoME setup. The following step is data analysis to detect outliers that do not correspond with the rest of the data. Depending on the service requesting the DoME, the anomaly is eliminated and/or reported. For example, the fraud detection or predictive maintenance service will choose to be notified of anomalous data, which may imply that fraud is being committed in the network or the malfunctioning of a piece of equipment. Then, the DoME performs the task of imputing or deleting rows with missing values, depending on whether the dataset is used for training tasks. By imputing values, the dataset size remains the same. Only the empty records which were either missing or deleted due to anomalies are now filled with predicted values. The second diagram describes the DoME process for Big Data files. Spark is an open-source parallel processing framework for running large-scale data analytics applications across clustered computers. It can handle both batch and real-time analytics and data processing workloads. In this context, DoME uses Spark to process and clean the data, returning clean datasets with the required quality.

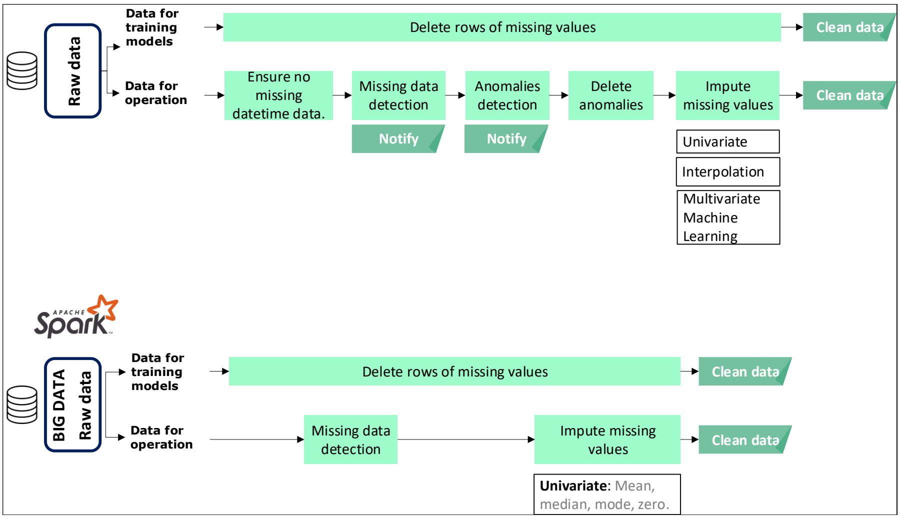

Figure B.1: Error detection tool flowchart.

Some functionalities have been eliminated when dealing with Big Data files since not all the libraries used for smaller files (first diagram) allow to work computationally in a distributed manner, as Spark requires.

B.2.1 Missing data imputation methods

The missing values should be replaced with rational records to offer the data-driven services a completed dataset. The approach to handling missing values is called Imputation. Several Imputation techniques are applied and offered in this service, ranging from Statistics to Artificial Intelligence.

B.2.1.1 Statistics methods

Univariate imputation

The following methods are used as the imputation value for the Statistical estimation of the missing data.

- **Mean**. This method replaces missing values with the mean of the available records in that column. It is a fast imputation method; however, it can provide imputations far from the true value if the data do not follow a normal distribution. In addition, mean values may be influenced by outliers.
- **Median**. Similar to mean imputation, this method replaces missing values with the median of the available values in that column. It is a more robust method than mean imputation since it is not affected by outliers.
- **Most frequent value**. This method replaces missing values with the most frequent value in that column. This method is typically used for categorical data.
- **Zero values**. This method replaces missing values with zeros. It can be useful when missing values indicate that the variable did not occur or have a value of zero. For instance, PV generation at night.

Interpolation: Interpolation is a mathematical method that adjusts a function to the dataset and uses this function to extrapolate the missing data. The simplest type is linear interpolation, but polynomial is also available in DME, indicating the degree.

- Linear. This method assumes a linear relationship between known data points and calculates the missing values accordingly.
- Polynomial. This method fits a polynomial function to the available data and uses it to estimate the missing values. It is useful when there is a non-linear relationship between the variables.

B.2.1.2 Artificial Intelligence methods

Multivariate Imputation (Machine Learning techniques)
For this imputation technique, a distributed set of observed data estimates a set of imputation values for the missing data. A Multiple Imputation by Chained Equations (MICE method) is applied in this method.

- **Light Gradient Boosting Machine**. It is a Multiple Imputation by Chained Equations (MICE) missing data in a dataset through an iterative series of predictive models. In each iteration, each specified variable in the dataset is imputed using the other variables in the dataset. These iterations should be run until it appears that convergence has been met.
- **Random Forest**. This method fits a random forest model on the observed part and then predicts the missing part. [410]
- **KNN**. Each sample of missing values is imputed using the mean value from k nearest neighbors found in the training set. Two samples are close if the features that neither is missing are close.

There are trade-offs between the proposed missing imputation options. Using one method or another will depend on the service's necessary specifications that need clean data. The key points are the following:

- Execution time (best to worst): Univariate, Interpolation, Multivariate Imputation.
- Imputation quality (best to worst): Multivariate Imputation, Interpolation, Univariate.

B.2.2 Outliers detection methods

The error detection tool offers two methods to detect outliers.

IQR method: The Interquartile Range (IQR) is often used to detect outliers in data, based on the concept of the quartiles of a distribution. It follows these steps:

1. First, calculate the IQR range for the data, which is calculated as the difference between the third quartile (Q3) and the first quartile (Q1).

$$IQR = Q3 - Q1 \tag{B.1}$$

2. Calculate the lower and upper bounds for detecting outliers are defined. Any data point that falls outside these bounds is considered an outlier. The choice of 1.5 as the multiplier is arbitrary and can be modified as needed in this tool.
3. Upper bound: Add 1.5IQR to the third quartile (Q3). Any number more significant than this is considered an outlier.

$$Upperbound = Q3 + (1.5IQR) \tag{B.2}$$

4. Lower bound: Substract 1.5IQR to the first quartile (Q1). Any number less significant than this is considered an outlier.

$$Lowerbound = Q1 - (1.5IQR) \tag{B.3}$$

The IQR method is robust to outliers and can be useful for detecting them in datasets with skewed distributions or outliers that are not extreme. However, it may not work well for datasets with extreme outliers or small sample sizes.

Mean-std difference method: This method deletes the outliers that are above and below the mean value minus the standard deviation, multiplied by a parameter that can be modified:

$$Outlier = (mean \pm std) * Constantparameter \tag{B.4}$$

B.2.3 DoME workflow

The operational input parameters for this tool are listed and described in Table B.1. Many have already been explained in the prior section, but others are defined here for the first time.

Figure B.2 illustrates the execution process of the DoME tool. Depending on the selected input parameters (see Table B.1), the tool will perform differently. In this visual example, the selected input variables follow the light blue arrows until reaching the output file, which contains clean and error-free data.

Table B.1: DoME parameters description.

Parameters	Description
Pre-training models	(bool) If True, missing data will be deleted
Outliers detection	(bool) If True, the outlier detection is performed and saved in a text file
missing detection	(bool) If True, the missing data detection is performed and saved in a text file
outlier methods	(simple, IQR). simple: easy way of detecting outliers using the standard deviation. 'IQR' method.
IQR coefficient	(float >1) Value needed to calculate the IQR method.
Simple coefficient	(float) Value that sets the outlier detection range for the 'simple' outlier detection method. The bigger the coefficient is, the stricter the outlier detection becomes.
Outliers deletion	(bool) If True, outliers are deleted, and a missing value is imputed afterwards.
Missing imputation method	(mean, median, most frequent, micerf, multivariate, KNN, linear, poly) The "micerf" imputes the missing data using Random Forests and the other columns as input training data to estimate the missing value smartly and automatically. If the dataset has only one column, micerf can not be selected (linear imputation runs instead). Micerf, multivariate and KNN use ML techniques for missing data imputation.
order	(int) If missing method = 'poly', this parameter indicates the polynomial order.
plot visualization	(bool) If True, plots of each step are saved (raw data, anomalies deleted)
missing data visualization	(bool) If True, plots regarding missing data in the raw dataset are saved.
missing threshold	(float <1) Parameter that sets the maximum acceptable limit of missing data in the input raw dataset. If the missing data threshold = 0.1, the dataset should have less than 10% of missing records in the dataset. If higher, a warning sign/text pops up.
outliers threshold	(=float <1) Parameter that sets the maximum acceptable limit of outliers in the raw input dataset. If the outliers threshold = 0.05, the dataset should have less than 5% of outliers in the dataset. If higher, a warning sign/text pops up.

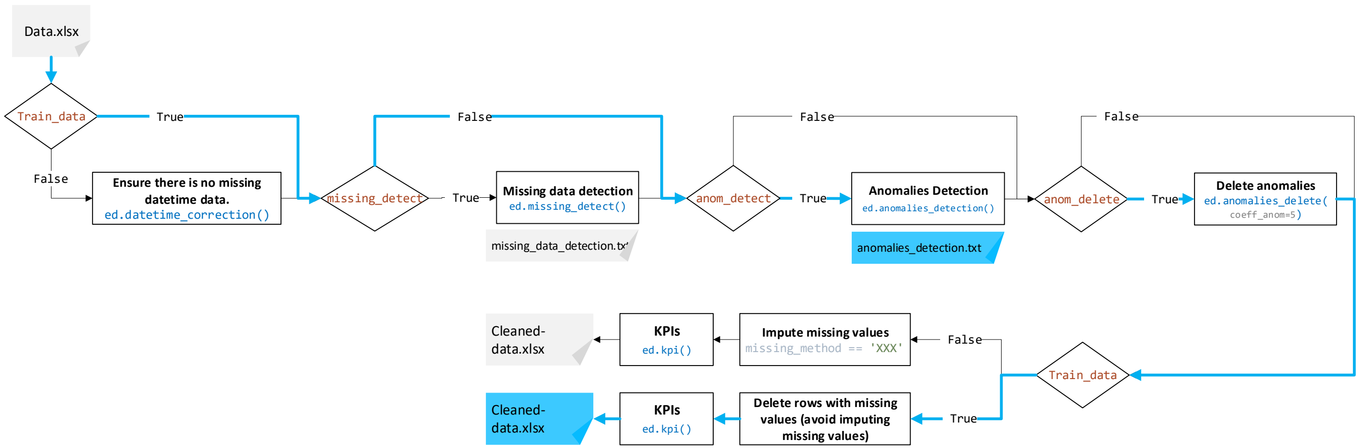

Figure B.2: Dataflow detection of measurement errors

B.3 Example of tool application in thesis

This section provides a use case to help understand the performance and functionality of DoME during the development of this thesis.

Specifically, the case study is taken from Chapter 4, where DoME was used to clean data for several scheduled generation sources in Spain, which were necessary to calculate the total hourly GWP of the national electricity mix. To provide context, the objective was to calculate the hourly GWP of the Spanish electrical grid from March to December 2022, in order to use these indicators as input for the environmental and hybrid-based LEC optimization approach.

For this case study, two datasets were chosen from the Spanish Electricity Market database: anthracite and lignite coal generation. However, it was discovered that a significant number of data points were missing since the database did not provide information for hours when there was no generation. This is logical as the share of carbon generation units in the energy mix is decreasing due to the pollutants they release into the atmosphere. It is common for databases to delete the date and time if the values are 0 in order to conserve storage space.

Figures B.3 and B.5 show the raw data extracted from the database for hard coal and lignite, respectively. The results or KPIs obtained after the data cleaning for each dataset are presented in Tables B.2 and B.3.

Table B.2: KPIs obtained for Lignite data cleaning with DoME.

KPIs	**Value**
Total data	7147.00
Missing data	2057.00
Share missing data	28.78%

To handle the missing datetime values, DoME applies the zero-value imputation method since it is known that the missing values were due

Table B.3: KPIs obtained for Hard coal data cleaning with DoME.

KPIs	Value
Total data	7299.00
Missing data	1942.00
Share missing data	26.61%

to the actual value being zero.

Finally, Figures B.4 and B.6 display the resultant cleaned data files, which are now ready for further utilization.

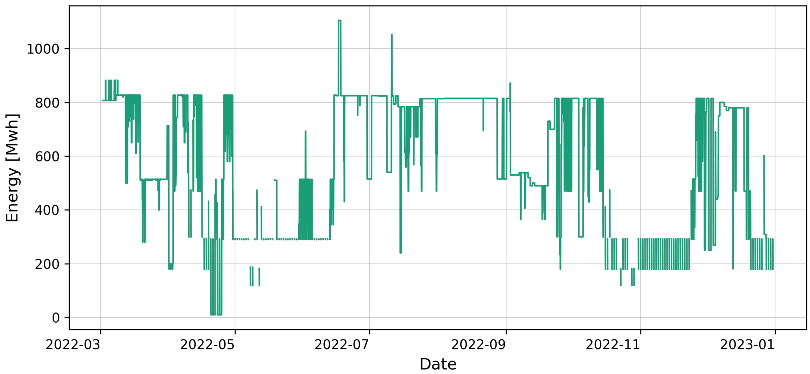

Figure B.3: Raw hard coal generation dataset.

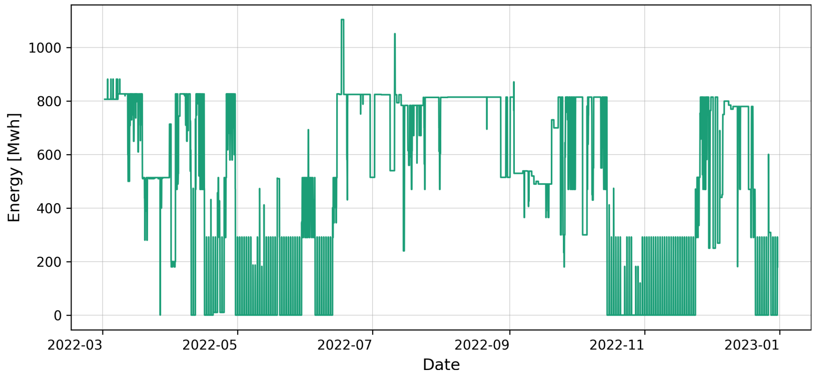

Figure B.4: Cleansed hard coal generation dataset.

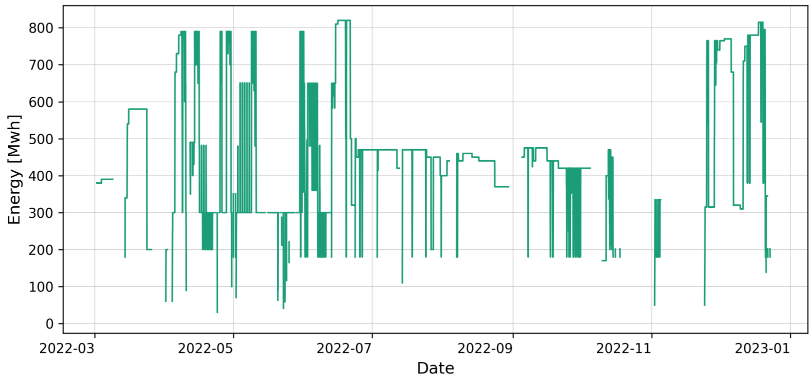

Figure B.5: Raw lignite generation dataset.

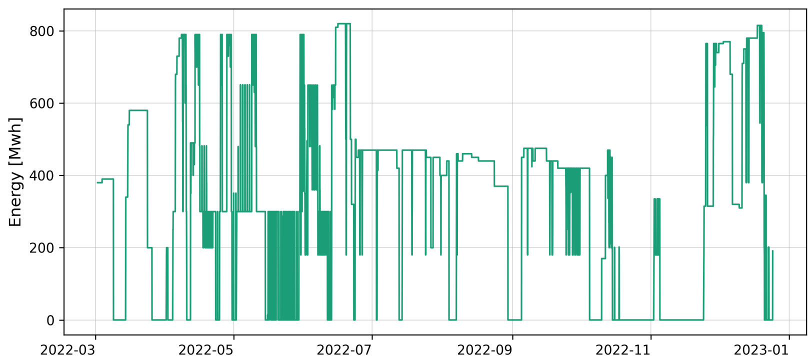

Figure B.6: Cleansed lignite generation dataset.