ANNEX 82

EBC
Energy in Building and Communities Programme

International Energy Agency

Methodologies and evaluations of energy flexibility for clusters of buildings

Energy in Buildings and Communities Technology Collaboration Programme

July 2025

ANNEX 82

EBC

Energy in Building and Communities Programme

International Energy Agency

Methodologies and evaluations of energy flexibility for clusters of buildings

Energy in Buildings and Communities Technology Collaboration Programme

June 2025

Editors

Rui Amaral Lopes, School of Science and Technology, UNINOVA-CTS and LASI, NOVA University Lisbon, Portugal, rm.lopes@fct.unl.pt

Jérôme Le Dréau, LaSIE UMR CNRS 7356, La Rochelle University, France, jledreau@univ-lr.fr

Gregor Henze, (1) Department of Civil, Environmental and Architectural Engineering, University of Colorado, Boulder, (2) Renewable and Sustainable Energy Institute, Boulder, United States, gregor.henze@colorado.edu

Michaël Kummert, Department of Mechanical Engineering, Polytechnique Montréal, Montréal, Québec, Canada, michael.kummert@polymtl.ca

Authors

Rui Amaral Lopes, School of Science and Technology, UNINOVA-CTS and LASI, NOVA University Lisbon, Portugal, rm.lopes@fct.unl.pt

Jérôme Le Dréau, LaSIE UMR CNRS 7356, La Rochelle University, France, jledreau@univ-lr.fr

Gregor Henze, (1) Department of Civil, Environmental and Architectural Engineering, University of Colorado, Boulder, (2) Renewable and Sustainable Energy Institute, Boulder, United States, gregor.henze@colorado.edu

Michaël Kummert, Department of Mechanical Engineering, Polytechnique Montréal, Montréal, Québec, Canada, michael.kummert@polymtl.ca
Alessia Arteconi, (1) Department of Mechanical Engineering, KU Leuven, 3000, Leuven, Belgium; (2) Dipartimento di Ingegneria Industriale e Scienze Matematiche, Università Politecnica delle Marche, 60131, Ancona, Italy, a.arteconi@univpm.it
Hanmin Cai, Urban Energy Systems Laboratory, Swiss Federal Laboratories for Material Science and Technology (Empa), Dubendorf, Switzerland, hanmin.cai@empa.ch
George Dawes, School of Architecture, Building and Civil Engineering, Loughborough University, UK, G.Dawes@lboro.ac.uk
Monika Hall, Institute of Sustainability and Energy in Construction, University of Applied Sciences and Arts Northwestern Switzerland, Switzerland, monika.hall@fhnw.ch
Hussain Kazmi, Department of Electrical Engineering (ELECTA-ESAT), KU Leuven, Belgium, hussain.kazmi@kuleuven.be
Tuğçin Kırant-Mitić, Department of Building Physics and Technical Services, Faculty of Architecture and Civil Engineering, Wuppertal University, Germany, kirantmitic@uni-wuppertal.de
Rongling Li, Department of Civil and Mechanical Engineering, Technical University of Denmark, Brovej Building 118, 2800 Kgs. Lyngby, Denmark, liron@dtu.dk
Henrik Madsen, Department of Applied Mathematics and Computer Science, Technical University of Denmark, Building 324, 2800 Kgs. Lyngby, Denmark, hmad@dtu.dk
Reza Mokhtari, Department of Civil and Mechanical Engineering, Technical University of Denmark, Brovej Building 118, 2800 Kgs. Lyngby, Denmark, remok@dtu.dk
Zheng O'Neill, J Mike Walker' 66 Department of Mechanical Engineering, Texas A&M University, College Station, TX, 77843, USA, zoneill@tamu.edu
Subbu Sethuvenkatraman, CSIRO Energy Centre, PO Box 330, Newcastle, NSW 2300, Australia, Subbu.Sethuvenkatraman@csiro.au
Shahab Tohidi, Department of Applied Mathematics and Computer Science Technical University of Denmark, Building 324, 2800 Kgs. Lyngby, Denmark, sshto@dtu.dk
Vojtech Zavrel, University Centre for Energy Efficient Buildings, Czech Technical University in Prague, Třinecká 1024, 273 43 Buštěhrad, Czech Republic vojtech.zavrel@cvut.cz
Sara Willems, KU Leuven, Belgium, sara.willems@kuleuven.be
Stephen White, CSIRO Energy Centre, PO Box 330, Newcastle, NSW 2300, Australia, Stephen.D.White@csiro.au

Published by Technical University of Denmark, Anker Engelunds Vej 10, 2800 Kongens Lyngby, Denmark

ISBN: 978-87-7475-795-5

Participating countries in the EBC TCP: Australia, Austria, Belgium, Brazil, Canada, P.R. China, Czech Republic, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Republic of Korea, the Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, Türkiye, United Kingdom and the United States of America.

Additional copies of this report may be obtained from: EBC Executive Committee Support Services Unit (ESSU), C/o AECOM Ltd, The Colmore Building, Colmore Circus Queensway, Birmingham B4 6AT, United Kingdom
www.iea-ebc.org
essu@iea-ebc.org

Preface

The International Energy Agency

The International Energy Agency (IEA) was established in 1974 within the framework of the Organisation for Economic Co-operation and Development (OECD) to implement an international energy programme. A basic aim of the IEA is to foster international co-operation among the 30 IEA participating countries and to increase energy security through energy research, development and demonstration in the fields of technologies for energy efficiency and renewable energy sources.

The IEA Energy in Buildings and Communities Programme

The IEA co-ordinates international energy research and development (R&D) activities through a comprehensive portfolio of Technology Collaboration Programmes (TCPs). The mission of the IEA Energy in Buildings and Communities (IEA EBC) TCP is to support the acceleration of the transformation of the built environment towards more energy efficient and sustainable buildings and communities, by the development and dissemination of knowledge, technologies and processes and other solutions through international collaborative research and open innovation. (Until 2013, the IEA EBC Programme was known as the IEA Energy Conservation in Buildings and Community Systems Programme, ECBCS.)

The high priority research themes in the EBC Strategic Plan 2019-2024 are based on research drivers, national programmes within the EBC participating countries, the Future Buildings Forum (FBF) Think Tank Workshop held in Singapore in October 2017 and a Strategy Planning Workshop held at the EBC Executive Committee Meeting in November 2017. The research themes represent a collective input of the Executive Committee members and Operating Agents to exploit technological and other opportunities to save energy in the buildings sector, and to remove technical obstacles to market penetration of new energy technologies, systems and processes. Future EBC collaborative research and innovation work should have its focus on these themes.

At the Strategy Planning Workshop in 2017, some 40 research themes were developed. From those 40 themes, 10 themes of special high priority have been extracted, taking into consideration a score that was given to each theme at the workshop. The 10 high priority themes can be separated in two types namely 'Objectives' and 'Means'. These two groups are distinguished for a better understanding of the different themes.

Objectives - The strategic objectives of the EBC TCP are as follows:

– reinforcing the technical and economic basis for refurbishment of existing buildings, including financing, engagement of stakeholders and promotion of co-benefits;

– improvement of planning, construction and management processes to reduce the performance gap between design stage assessments and real-world operation;

– the creation of 'low tech', robust and affordable technologies;

– the further development of energy efficient cooling in hot and humid, or dry climates, avoiding mechanical cooling if possible;

– the creation of holistic solution sets for district level systems taking into account energy grids, overall performance, business models, engagement of stakeholders, and transport energy system implications.

Means - The strategic objectives of the EBC TCP will be achieved by the means listed below:

– the creation of tools for supporting design and construction through to operations and maintenance, including building energy standards and life cycle analysis (LCA);

– benefitting from 'living labs' to provide experience of and overcome barriers to adoption of energy efficiency measures;

– improving smart control of building services technical installations, including occupant and operator interfaces;

– addressing data issues in buildings, including non-intrusive and secure data collection;

– the development of building information modelling (BIM) as a game changer, from design and construction through to operations and maintenance.

The themes in both groups can be the subject for new Annexes, but what distinguishes them is that the 'objectives' themes are final goals or solutions (or part of) for an energy efficient built environment, while the 'means' themes are instruments or enablers to reach such a goal. These themes are explained in more detail in the EBC Strategic Plan 2019-2024.

The Executive Committee

Overall control of the IEA EBC Programme is maintained by an Executive Committee, which not only monitors existing projects, but also identifies new strategic areas in which collaborative efforts may be beneficial. As the Programme is based on a contract with the IEA, the projects are legally established as Annexes to the IEA EBC Implementing Agreement. At the present time, the following projects have been initiated by the IEA EBC Executive Committee, with completed projects identified by (*) and joint projects with the IEA Solar Heating and Cooling Technology Collaboration Programme by (☼):

Annex 1: Load Energy Determination of Buildings (*)
Annex 2: Ekistics and Advanced Community Energy Systems (*)
Annex 3: Energy Conservation in Residential Buildings (*)
Annex 4: Glasgow Commercial Building Monitoring (*)
Annex 5: Air Infiltration and Ventilation Centre
Annex 6: Energy Systems and Design of Communities (*)
Annex 7: Local Government Energy Planning (*)
Annex 8: Inhabitants Behaviour with Regard to Ventilation (*)
Annex 9: Minimum Ventilation Rates (*)
Annex 10: Building HVAC System Simulation (*)
Annex 11: Energy Auditing (*)
Annex 12: Windows and Fenestration (*)
Annex 13: Energy Management in Hospitals (*)
Annex 14: Condensation and Energy (*)
Annex 15: Energy Efficiency in Schools (*)
Annex 16: BEMS 1- User Interfaces and System Integration (*)
Annex 17: BEMS 2- Evaluation and Emulation Techniques (*)
Annex 18: Demand Controlled Ventilation Systems (*)
Annex 19: Low Slope Roof Systems (*)
Annex 20: Air Flow Patterns within Buildings (*)
Annex 21: Thermal Modelling (*)
Annex 22: Energy Efficient Communities (*)
Annex 23: Multi Zone Air Flow Modelling (COMIS) (*)
Annex 24: Heat, Air and Moisture Transfer in Envelopes (*)
Annex 25: Real time HVAC Simulation (*)
Annex 26: Energy Efficient Ventilation of Large Enclosures (*)
Annex 27: Evaluation and Demonstration of Domestic Ventilation Systems (*)
Annex 28: Low Energy Cooling Systems (*)
Annex 29: ☼ Daylight in Buildings (*)
Annex 30: Bringing Simulation to Application (*)
Annex 31: Energy-Related Environmental Impact of Buildings (*)
Annex 32: Integral Building Envelope Performance Assessment (*)
Annex 33: Advanced Local Energy Planning (*)
Annex 34: Computer-Aided Evaluation of HVAC System Performance (*)
Annex 35: Design of Energy Efficient Hybrid Ventilation (HYBVENT) (*)
Annex 36: Retrofitting of Educational Buildings (*)
Annex 37: Low Exergy Systems for Heating and Cooling of Buildings (LowEx) (*)
Annex 38: ☼ Solar Sustainable Housing (*)
Annex 39: High Performance Insulation Systems (*)

Annex 40: Building Commissioning to Improve Energy Performance (*)
Annex 41: Whole Building Heat, Air and Moisture Response (MOIST-ENG) (*)
Annex 42: The Simulation of Building-Integrated Fuel Cell and Other Cogeneration Systems (FC+COGEN-SIM) (*)
Annex 43: ☼ Testing and Validation of Building Energy Simulation Tools (*)
Annex 44: Integrating Environmentally Responsive Elements in Buildings (*)
Annex 45: Energy Efficient Electric Lighting for Buildings (*)
Annex 46: Holistic Assessment Tool-kit on Energy Efficient Retrofit Measures for Government Buildings (EnERGo) (*)
Annex 47: Cost-Effective Commissioning for Existing and Low Energy Buildings (*)
Annex 48: Heat Pumping and Reversible Air Conditioning (*)
Annex 49: Low Exergy Systems for High Performance Buildings and Communities (*)
Annex 50: Prefabricated Systems for Low Energy Renovation of Residential Buildings (*)
Annex 51: Energy Efficient Communities (*)
Annex 52: ☼ Towards Net Zero Energy Solar Buildings (*)
Annex 53: Total Energy Use in Buildings: Analysis and Evaluation Methods (*)
Annex 54: Integration of Micro-Generation and Related Energy Technologies in Buildings (*)
Annex 55: Reliability of Energy Efficient Building Retrofitting - Probability Assessment of Performance and Cost (RAP-RETRO) (*)
Annex 56: Cost Effective Energy and CO_2 Emissions Optimization in Building Renovation (*)
Annex 57: Evaluation of Embodied Energy and CO_2 Equivalent Emissions for Building Construction (*)
Annex 58: Reliable Building Energy Performance Characterisation Based on Full Scale Dynamic Measurements (*)
Annex 59: High Temperature Cooling and Low Temperature Heating in Buildings (*)
Annex 60: New Generation Computational Tools for Building and Community Energy Systems (*)
Annex 61: Business and Technical Concepts for Deep Energy Retrofit of Public Buildings (*)
Annex 62: Ventilative Cooling (*)
Annex 63: Implementation of Energy Strategies in Communities (*)
Annex 64: LowEx Communities - Optimised Performance of Energy Supply Systems with Exergy Principles (*)
Annex 65: Long-Term Performance of Super-Insulating Materials in Building Components and Systems (*)
Annex 66: Definition and Simulation of Occupant Behavior in Buildings (*)
Annex 67: Energy Flexible Buildings (*)
Annex 68: Indoor Air Quality Design and Control in Low Energy Residential Buildings (*)
Annex 69: Strategy and Practice of Adaptive Thermal Comfort in Low Energy Buildings (*)
Annex 70: Energy Epidemiology: Analysis of Real Building Energy Use at Scale (*)
Annex 71: Building Energy Performance Assessment Based on In-situ Measurements (*)
Annex 72: Assessing Life Cycle Related Environmental Impacts Caused by Buildings (*)
Annex 73: Towards Net Zero Energy Resilient Public Communities (*)
Annex 74: Competition and Living Lab Platform (*)
Annex 75: Cost-effective Building Renovation at District Level Combining Energy Efficiency and Renewables (*)
Annex 76: ☼ Deep Renovation of Historic Buildings Towards Lowest Possible Energy Demand and CO_2 Emissions (*)
Annex 77: ☼ Integrated Solutions for Daylight and Electric Lighting (*)
Annex 78: Supplementing Ventilation with Gas-phase Air Cleaning, Implementation and Energy Implications
Annex 79: Occupant-Centric Building Design and Operation
Annex 80: Resilient Cooling
Annex 81: Data-Driven Smart Buildings
Annex 82: Energy Flexible Buildings Towards Resilient Low Carbon Energy Systems
Annex 83: Positive Energy Districts

Annex 84: Demand Management of Buildings in Thermal Networks
Annex 85: Indirect Evaporative Cooling
Annex 86: Energy Efficient Indoor Air Quality Management in Residential Buildings
Annex 87: Energy and Indoor Environmental Quality Performance of Personalised Environmental Control Systems
Annex 88: Evaluation and Demonstration of Actual Energy Efficiency of Heat Pump Systems in Buildings
Annex 89: Ways to Implement Net-zero Whole Life Carbon Buildings
Annex 90: ☼ EBC Annex 90 / SHC Task 70 Low Carbon, High Comfort Integrated Lighting
Annex 91: Open BIM for Energy Efficient Buildings
Annex 92: Smart Materials for Energy-efficient Heating, Cooling and IAQ Control in Residential Buildings
Annex 93: Energy Resilience of the Buildings in Remote Cold Regions
Annex 94: Validation and Verification of In-situ Building Energy Performance
Annex 95: Human-centric Building Design and Operation for a Changing Climate
Annex 96: Grid Integrated Control of Buildings
Annex 97: Sustainable Cooling in Cities

Working Group - Energy Efficiency in Educational Buildings (*)
Working Group - Indicators of Energy Efficiency in Cold Climate Buildings (*)
Working Group - Annex 36 Extension: The Energy Concept Adviser (*)
Working Group - HVAC Energy Calculation Methodologies for Non-residential Buildings (*)
Working Group - Cities and Communities (*)
Working Group - Building Energy Codes

Summary

Decarbonization of the electricity grid and the electrification of transport, heat and industry related energy consumption are at the center of many governments' policies to reach ambitious targets to reduce greenhouse gas emissions. The integration of variable renewable energy sources into power grids and the growing electricity demand reinforce the need for energy flexibility: buildings should adapt their energy demand and/or production to weather conditions, to their users' needs, and to the requirements of the electricity grid.

This report summarizes the work performed within Annex 82, which includes the review of state-of-the-art methodologies for the evaluation of energy flexibility at the building cluster level. Addressing energy flexibility at the building cluster level remains a challenge, with technical and non-technical barriers to adoption (non-technical barriers are discussed in more detail in another report of Annex 82, which presents recommendations for policy makers and government entities). Technical barriers for the early planning, design, and operation phases include the development of integrated modelling tools, control strategies addressing the building cluster or portfolio level, and the development of quantitative methodologies and indicators. More effective strategies and policies must be developed to design and operate resilient and sustainable communities. Dominant factors for energy flexibility are related to occupants, building characteristics, energy systems and storage, control systems, and external factors (e.g., weather and market).

Research teams involved in Annex 82 also investigated building-grid interaction signals, a generalization of "price signals" or "penalty signals". These signals must be aligned with the demand response objectives for the correct use of available energy flexibility. To characterize this flexibility, building load prediction plays a crucial role in the evaluation of baseline scenarios and available demand-response potential. Key performance indicators for energy flexibility were also assessed and developed within the Annex. In particular, the flexibility index and flexibility function were tested and improved.

One of the main collaborative efforts within Annex 82 consisted in the "Common Exercise", which aimed at enabling various research groups to tackle a joint challenge by specifying common boundary conditions, while allowing teams to use their own simulation tools, control algorithms, and datasets. A common building-grid interaction signal (a "price signal" in this case) was defined and all teams were asked to select a building portfolio and to assess the energy flexibility at the portfolio level. Common performance indicators were assessed by all teams, which allowed a comparison across widely different building portfolios and climate conditions. Both reactive and predictive strategies led to a worsening of monitored metrics such as the load factor or system ramping and generate significant rebounds (as well as "prebounds") before and/or after the flexibility events. All results illustrate the need for mitigation strategies to obtain a positive impact of local energy flexibility strategies at the global (i.e., portfolio) level. Energy flexibility strategies that strictly maintain an equivalent comfort level also led to an overall energy increase, which may be justified by the variable "value" of electricity, whether it is assessed by cost or GHG emissions, for example.

Finally, Annex 82 work highlighted the scarcity of field studies demonstrating energy flexibility at the building cluster or portfolio level. Two field implementations conducted within Annex 82 are described in the report: the use of a novel control algorithm (signal matrix model predictive control) to control space heating, domestic hot water heating, and a stationary battery, and the coordinated use of the different assets in a fully equipped occupied building for flexibility. Annex 82 has advanced the state-of-the-art in methods to characterize, model, and harness energy flexibility of building clusters and portfolios, which will enable more demonstration projects to confirm the promising results of those two field studies.

Table of content

Acronyms

Abbreviations	Meaning
DER	Distributed Energy Resources
DeePC	Data-Enabled Predictive Control
DHW	Domestic Hot Water
DR	Demand Response
DSO	Distribution System Operator
ESCOs	Energy Service Companies
EU	European Union
EV	Electric Vehicle
FERC	Federal Energy Regulatory Commission
FF	Flexibility Function
FI	Flexibility Index
HVAC	Heating, Ventilation and Air Conditioning
GTA	Global Temperature Adjustment
HP	Heat Pump
KPI	Key Performance Indicator
MTSR	Monash Time Series Repository
MPC	Model Predictive Control
OPC	Open Platform Communications
OCP	Optimal Control Problem
P2P	Peer-to-Peer
PV	Photovoltaic
PLC	Programmable Logic Controller
RBC	Rule-Based Control
REC	Renewable Energy Community
RES	Renewable Energy Resources
SSSS	Sub-keyword Synonym Subtopics Searching
SEOP	Smart Energy Operating Systems
SMM-PC	Signal Matrix Model Predictive Control
SH	Space Heating
SOC	State of Charge
SPC	Subspace Predictive Control
TSO	Transmission System Operator
US	United States
BGI	Building-Grid Interaction Signal

1. Introduction

Decarbonization of the electricity grid and the electrification of transport, heat and industry related energy consumption are two of the primary means targeted at reducing greenhouse gas emissions. The goal is to meet the target of limiting global warming to below 1.5°C, as agreed in the Paris Agreement (United Nations Framework Convention on Climate Change, 2015) and confirmed at the recent COP26 summit (United Nations Framework Convention on Climate Change, 2022). Decarbonization of the energy grids through increased deployment of renewable energy sources requires measures such as active management of the power grids to always balance energy supply and demand. Buildings, which account for about 35 % of the global energy use (United Nations Environment Programme; Global Alliance for Buildings and Construction, 2020), have a significant potential for the development of Demand Response (DR) strategies (M. Evans et al., 2021) using existing systems for energy flexibility. **Figure 1** shows the importance of distributed flexibility resources in the coming years. In their Net Zero Emissions by 2050 Scenario (IEA, 2023), the IEA is calling for a tenfold increase in demand response availability from buildings between 2020 and 2030 (**Figure 2**). This flexibility is defined as the ability of a building to change its short-term (a few hours or a couple of days) energy demand and/or energy generation, according to weather conditions, user needs and energy network requirements, without jeopardizing the technical capabilities of the building or occupant comfort (H. Li et al., 2021). Practical examples of flexible loads in buildings include storage heaters (space and water), heat pumps, air conditioners and other relevant energy systems. Interest in energy flexibility in buildings has significantly gained momentum in the last decade, partially driven by increased penetration of renewable energy systems, coupled with increases in energy prices, as well as consumer focus on energy costs.

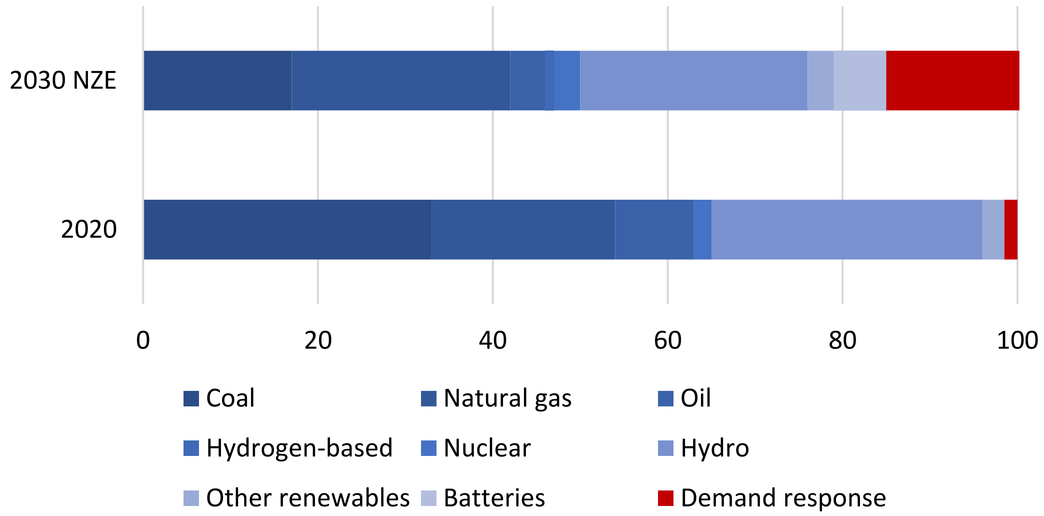


Figure 1: Flexibility from batteries and demand response will grow from 3% in 2020 to 22% in 2030 (IEA, 2021).

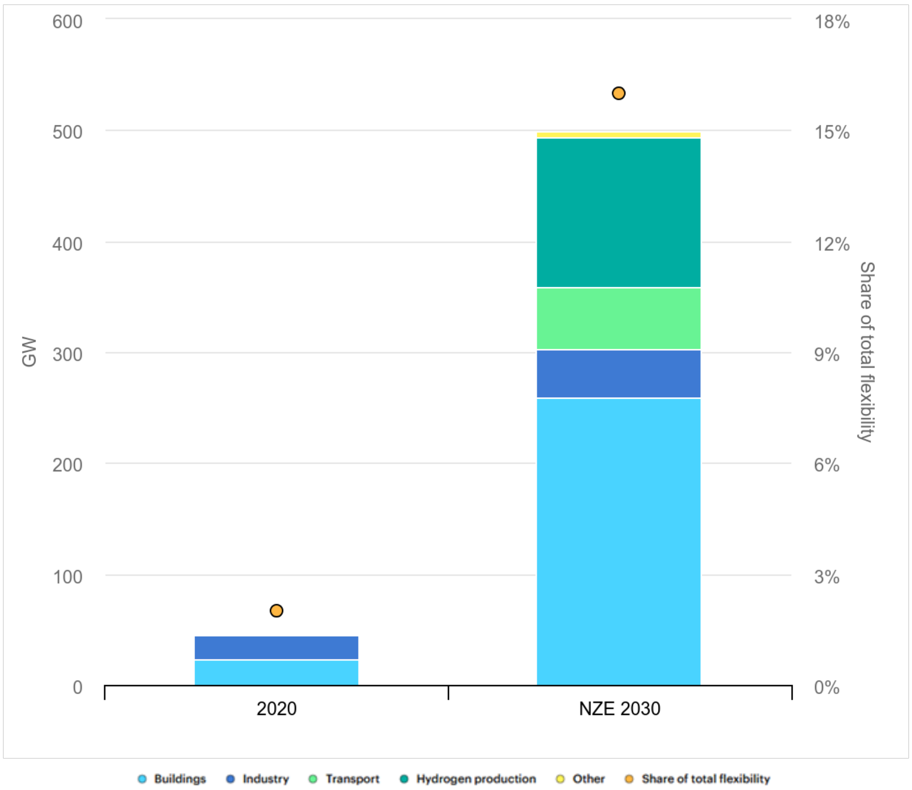


Figure 2: Demand response availability in 2020 (contracted capacity accross capacity, balancing and frequency response) and estimated potential for 2030 in the Net Zero Emissions (NZE) by 2050 scenario – Geographical area: Europe, North America, Australisia.

Energy flexibility from single buildings can be impractical to harness due to small quantities available at the individual building level, the diverse small sources of flexibility, and stochasticity in occupant behavior (Sala et al., 2021). Aggregation at scale is viewed as a solution to foster the development of flexibility as it creates critical mass providing larger quantities of flexible loads and reduces uncertainty due to occupant behavior by increasing diversity. This leads to greater attractiveness for aggregators further providing the flexible capacity to TSOs/DSOs and utilities (Martinez et al., 2022). Moreover, new opportunities may develop at district scale, e.g., shared community production systems (Foteinaki et al., 2020; R. Li et al., 2022), thereby reducing redundancy requirements and decreasing reliance on fossil fuel backup generation. However, the practical exploitation of distributed flexibility resources is hindered, in part, by the lack of scalable modeling and control procedures. This need is emphasized by the large number of appliances to be controlled in coordinated manner, as seen in **Figure 3** according to the estimation of the IEA.

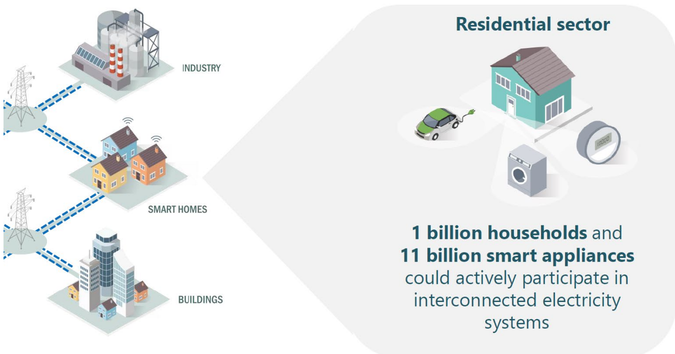


Figure 3: Active participation of smart appliances in system management (IEA, 2017)

This deliverable focuses on methodologies for assessing the energy flexibility of building clusters, based on the work carried out under the Annex 82. It is a combination of literature reviews and original work (both simulation and experimental) that promotes the development of flexibility at an aggregated scale. This deliverable focuses mainly on the technical characterization and use of flexibility and does not cover the legal, social or economic challenges. These aspects are covered in more detail in a second deliverable from the Annex 82 entitled "*Review and assessment of market, policy and stakeholder participation in energy flexibility of buildings*".

The report is structured as follows: section 2 summarizes various reviews focusing on the barriers to developing flexibility at the aggregated level as well as the drivers and the relationship between flexibility and resilience; section 3 presents various original methods for characterizing and activating flexibility as well as for forecasting short-term loads; section 4 presents various simulation case studies demonstrating how flexibility can be activated in clusters of buildings; finally, section 5 presents some case studies to highlight the practical issues of managing a group of assets.

2. General energy flexibility concepts

This section presents general energy flexibility concepts addressed during Annex 82, providing information on: i) the barriers and research gaps for the design and development of energy flexibility related projects at aggregated level (Section 2.1); ii) the relation between resilience and energy flexibility (Section 2.2); and iii) the factors influencing the characterization and exploitation of energy flexibility (Section 2.3).

2.1 Barriers and research gaps at district scale

This section addresses the main barriers and research gaps across three design and development phases of projects focusing on energy flexibility, namely: market and policy, early planning and design, and operation. The information presented here is based on the review article "*Developing energy flexibility in clusters of buildings: A critical analysis of barriers from planning to operation*", published in the journal Energy and Buildings (Le Dréau et al., 2023). Relevant sources were selected after an initial screening phase on the topic of both aggregated flexibility (keywords cluster, aggregate*, district, community, groups of buildings) and individual building flexibility. The key journals identified from this search were as follows: Energy and Buildings, Applied Energy, Energy and IEEE Series. Additional documents such as research project reports, guidelines and legislation were also reviewed, as they provide complementary information on the development of flexibility.

2.1.1 Barriers

The challenges of enabling demand-side flexibility in buildings are numerous, including the need for innovative business models, supportive legislation and regulations, and technological development, while operational evaluation of real performance is hampered by the lack of demonstration projects. **Figure 4** highlights the identified key barriers to this development (Le Dréau et al., 2023), grouped into five categories (policy, economic, technical, professional, social) and link them to the development phases of an energy flexibility exploitation project (market and policy, early planning, design and operation).

In terms of **policy**, the main barriers are related to the legal and regulatory framework, the limited access to the wholesale electricity market, and the privacy and cybersecurity concerns. The legislative and regulatory framework has developed rapidly in recent years and the landscape will continue to evolve. However, the rapid pace of change has sometimes led to a lack of customer awareness. The development of energy flexibility is also often linked to the development of energy communities and DERs, which have been initiated differently in the three jurisdictions studied. Two types of Energy Communities (i.e., Renewable Energy Communities (RECs) and Citizen Energy Communities (CECs)) have been explicitly legislated for by the EU (Hoicka et al., 2021), but the member state's implementation at local level differs and the design of effective supports for energy flexibility remains unclear. Increasing participation in energy markets of energy flexible resources is more advanced in the US through measures, such as Federal Energy Regulatory Commission (FERC), permitting DERs to participate in wholesale electricity markets (Forrester & Cappers, 2021). This, coupled with dynamic or real-time pricing at retail level, thereby reflecting actual renewable generation output, is a key enabler for flexibility in both the US (at state level) and the EU (through Directive 2019/944) (Hanny et al., 2022; Satchwell et al., 2020). In Australia, an overall review of the electricity market is underway and further initiatives to increase energy flexibility are expected. Resource planning is starting to incorporate the energy flexibility capabilities of DERs, such as through Australia's Demand Management Incentive Scheme, and it would be beneficial if such initiatives were replicated in other jurisdictions.

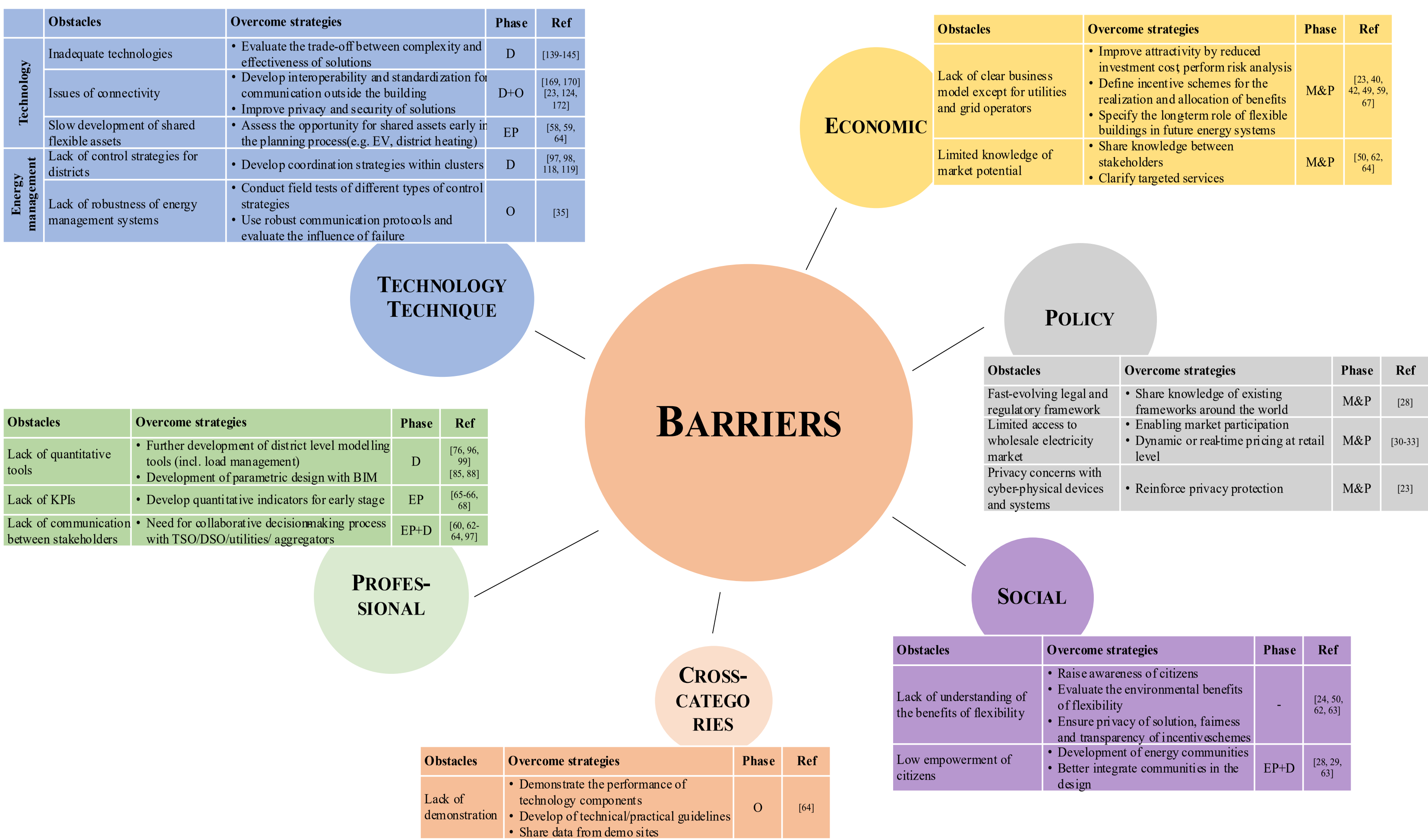


Figure 4: Overview of barriers to the development of energy flexibility in building clusters (acronyms: M&P market and policy, EP early-planning, D design, O operation). Source: (Le Dréau et al., 2023).

In terms of **economy**, the financing of DR primarily comes from flexibility activation payments and tariff optimization, but additional approaches that are being developed include virtual net metering, flexibility tenders and leveraging the collective power of Energy Communities. However, significant economic barriers still exist, in particular the lack of clarity around the value capture from multiple small sources of flexibility and specific financing mechanisms for flexibility within Energy Communities (Olgyay et al., 2020). The lack of a standardized building-to-grid assessment framework limits the ability of stakeholders and industry to quantify the value of flexibility. A standardized building-to-grid assessment framework could include (but not be limited to) a list of flexible systems with some quantitative indicators (installed power, availability, constraints), as well as information on the building's connectivity to the grid and the degree of interoperability. In terms of energy flexibility trading, P2P is emerging as a possible solution for small scale prosumers, and while trading barriers still exist (e.g., lack of low-cost retrofit solutions), it has been demonstrated successfully in Italy (Antal et al., 2021). Business models such as cooperatives, ESCOs or public–private partnerships have been most viable in a microgrid configuration to date. The business model canvas developed by Hamwai et al. (Hamwi et al., 2021) provides a starting point to valorize other approaches, but further work is needed on the value proposition motivating households and smaller participants in flexibility services.

In terms of **technology and techniques**, barriers exist at the technology and energy management levels. Cost-effective and reliable technologies should be developed to enable the activation of flexible assets. The communication of these flexible loads with the grid is also a cornerstone in the development of DR in buildings. A suite of technologies needs to work in harmony to control flexible loads, local generation or energy storage and create value for building owners and TSOs/DSOs. A reliable and secure two-way communication with a relatively high sampling rate is usually required and any failure in the chain of control or actuation may result in loss of signal transmission. Such issues may arise from databases, hardware, and technologies beyond building levels. Communication failure in operational projects can occur more frequently than planned in a design stage (Brinsmead et al., 2021). Interoperability and standardization should help improve reliability, but the robustness of solutions to communication failure should be tested. At the energy management level, the low diffusion of Building Automation System (BAS) and the lack of standards and seamless cross-domain data exchange solutions represent some of the main barriers for the implementation of energy flexibility services (Pallonetto et al., 2020; Satchwell et al., 2021). During the design phase, the data exchange between the cross-domain applications required to perform these processes also suffers from interoperability and standardization challenges (Y. Li et al., 2020; Tchouanguem Djuedja et al., 2019). Studies on semantic web technologies have made progress on this topic (Fierro et al., 2020; Pauwels et al., 2017; Pritoni et al., 2021). However, there is a lack of application of these studies dedicated to energy flexibility with standardized and replicable workflows.

Barriers to the integration of flexibility that **professionals** will experience during the design stage were also identified. The development of quantitative and qualitative methodologies (such as Smart Readiness Indicator and the Grid Optimal Buildings LEED pilot credit) should be pursued in the early planning stages to assess the flexibility potential of projects with low levels of information. At the design stage, building energy simulations (BES) are often limited in their ability to incorporate flexibility and load management strategies. For a district, several single-building models must be connected and coordinated, a feature which is not currently part of commercially available BES modeling software. Therefore, modeling flexibility in single buildings and districts with BES often requires development of external algorithms, co-simulation, pre-and post-processing. Together the co-simulation environment, complex energy system modeling and prediction horizon might cause a time-consuming model set-up, numerical problems, and long simulation times regarding its complexity. There is therefore a need for further development of district level modeling tools capable of testing control strategies for building clusters at early design stages. In terms of stakeholders, there is a very limited cooperation between the building and the energy sector, with the two working in silos. The energy infrastructure and building development are considered separately due to differing industry practices, stakeholders, project timelines, and regulatory frameworks within each sector. Overcoming this barrier will require improved collaboration between sectors to better consider their interrelated impacts and optimize solutions at the interface between buildings and the grid. This will require the energy sector to be more present at the local level to enable collaborative decision making.

In terms of **social** barriers, limited end-user knowledge of flexibility, and even energy in general, is one of the main limitations. Energy Communities can be used as a common ground to promote discussion at local level and to raise citizens' awareness of the concept of flexibility. The environmental and societal benefits of energy flexibility should also be emphasized, and the design of DR-programmes should account for the diversity of end-users. More studies are also needed to evaluate the relevance of price-based DR-programs to decrease GHG emissions, as a mismatch can be observed in some countries (Fleschutz et al., 2021). Moreover, there remain some privacy and security concerns for customers as cyber-physical devices and systems need to be integrated to enable smart management of homes and communities (Asghar et al., 2017; W. Li et al., 2021). Perceived consequences include potential leakage of personal information, losing control of devices and causing financial losses (W. Li et al., 2021).

Finally, the review performed highlights the lack of information from field studies, which should inform to the development of new technologies and the design of appropriate strategies. The experience gained in previous pilot projects is usually not publicly available and a steep learning curve may be required to reproduce the studies. There is also a lack of follow-up projects partially or fully reusing the infrastructure developed during the pilots. Therefore, transferability of pilot study learnings needs to be improved.

2.1.2 Research gaps

Based on the barriers identified in the previous part, several research directions can be formulated for the different development phases (**Figure 5**). In the following, we will describe some of these research directions and highlight possible solutions to overcome the barriers. However, we do not intend to be exhaustive, as various research directions can be formulated and this is an active field of research.

Assessing the role of building flexibility in long-term energy planning is seen as a key factor in encouraging investments. Energy planning and building planning have been and still are anchored in very different sectors and regional scales. To fill this gap, Thorvaldsen et al. (Thorvaldsen et al., 2022) highlighted the long-term value of building flexibility and the potential impact on price structures in this context. In addition, Chantzis et al. (Chantzis et al., 2023) have shown that the policy and regulatory aspects have a strong influence on the long-term contribution of demand response to decarbonization targets in the building sector. Key research questions in this context relate to the regulatory framework and its impact on the rapid uptake of flexibility measures in future developments, as well as pricing structures related to the potential uptake of building energy storage and flexibility in existing and new buildings.

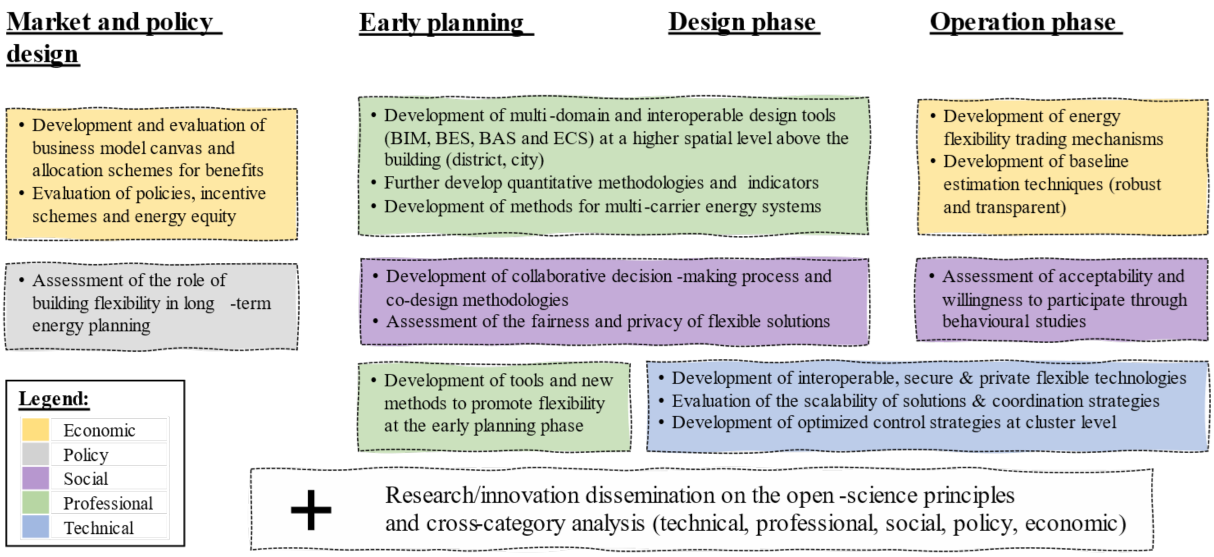


Figure 5: Summary of links between research gaps and phases. Source: (Le Dréau et al., 2023).

Data privacy and security is also one of the main concerns when implementing DR. This barrier can be tackled not only by establishing a legal framework to reinforce privacy protection, such as the General Data Protection Regulation (GDPR) in Europe or the California Consumer Privacy Act (CCPA) in the U.S., but also by developing technical solutions that minimize, or completely avoid, the use of personal data or data related to consumption and building occupancy. Therefore, technical innovations that allow decentralized control without sharing private user data (Tang et al., 2019) and/or solutions that can rely on aggregated data (Junker et al., 2018a), which increases the difficulty to obtain personal information (Yu et al., 2023), represent active areas of research. The co-design methodology within Energy Communities is also seen as a solution to overcome this problem (Crosbie et al., 2018).

Another barrier related to end-users is their acceptance and willingness to participate in DR programmes. Many technical studies ignore the social or economic aspects, although they are key to the success of flexibility (Ellabban & Abu-Rub, 2016). In the commercial and industrial sector, (Lashmar et al., 2022) conducted a literature review and interviewed end-users. They found out that DR participants were mainly motivated by financial benefits, but many were unaware of system and community benefits. They also identified newly reported barriers, such as the lack of trust between DR service providers and consumers, the resistance to change and the lack of interest in energy. In the residential sector, (Naghiyev et al., 2022) tested different user interface designs for automated washing machines and highlighted that DR incentivization should focus on convenience rather than money. Equity is also an active area of research, in particular to assess the potential price risk for consumers who do not respond to price signals (Blaschke, 2022). (Guo & Kontou, 2021) assessed the equity of EV purchase rebates across income groups and disadvantaged communities and highlighted the importance of income cap policies to improve equity.

To overcome the barrier of harnessing flexibility at scale, new strategies for coordinating and controlling building clusters need to be developed and tested (R. Li et al., 2022). As mentioned by (Kaspar et al., 2022), this manifests itself in the implementation of effective control at both building and cluster level. However, this is an open problem from an algorithmic perspective. On the one hand, activating flexibility in a centralized manner requires access to consumer electricity demand data, which may impact privacy as mentioned above. Such an approach also suffers from potential issues with communication and introduces a single point of failure. While techniques such as federated learning and learning from encrypted data have recently been introduced to energy flexible assets by (Balint et al., 2023), it is unclear what privacy protection they actually offer and how they will be adopted by industry players. On the other hand, truly decentralized flexibility activation requires both system identification and state estimation by each node. This is not only financially challenging, but it also introduces the risk of overshoot, i.e., situations where too much flexibility is activated due to poorly constructed price signals. The fact that this overshoot may not be observable in real time further complicates the problem.

To ensure better coordination of buildings, the multi-agent framework has recently been developed in research for different types of control strategies (centralized, decentralized or distributed). (Cai et al., 2020) exploited the flexibility from a cluster of buildings to alleviate network congestion issues in district heating systems by means of a coordinator to ensure that the collective response does not adversely impact the system. The method is capable of auto-correction with real-time demand and weather data, allowing optimization with a rolling horizon method to reduce the impact of prediction uncertainties. (Pinto et al., 2022) describe a multi-agent system for managing flexibility in building clusters at district level. Two multi-agent reinforcement learning methods are explored: a centralized (coordinated) controller and a decentralized (cooperative) controller, which are benchmarked against a rule-based controller. (Mazzarino et al., 2021) describes a modular multi-agent framework platform, which was tested numerically on a case study of 1,000 buildings, performing an analysis of the effects of small temperature deviations in buildings on the primary grid substation balancing problem. The results show the flexibility of the platform in testing different strategies. (Nweye et al., 2022) describe the CityLearn environment, an OpenAI Gym environment for the easy implementation of reinforcement learning agents in a demand response setting to reshape the aggregated curve of electricity demand by controlling the energy storage of a diverse set of buildings in a district. As seen in the previous examples, agent-based coordination has shown an interesting potential for harnessing flexibility at scale. However, it requires a stableand reliable two-way communication between the different agents and the sensitivity to network conditions should be evaluated (Zhang et al., 2023).

The challenge of coordination and control becomes even more complex when considering multi-carrier energy systems, where optimizing energy flows between different energy sources adds to the complexity of the problem. (Gholinejad et al., 2020) present a hierarchical home energy management system for energy hubs based on multi-agent reinforcement learning to schedule the flexible loads, the storage systems and the combined heat and power units. (Srithapon & Månsson, 2023) examined coordination approaches for multi-energy systems incorporating electricity and heating systems. (Zheng et al., 2022) describe a distributed multi-energy demand response method for the optimal coordinated operation of smart building clusters, which exploits a hierarchical building-aggregator framework.

Finally, in terms of dissemination based on open science principles, several initiatives can be highlighted. Among them, we can mention the publication of open datasets by different groups of researchers (e.g., three flexibility-related datasets made available in (Sartori et al., 2023), 16 in (H. Li et al., 2023), and four in (Energy.Gov, 2024)). Better dissemination of project results can also be seen in more recent projects, such as the Connected Communities program in the U.S. (U.S. Department of Energy, 2024) or the Smart-Grid program in France (Berthelon et al., 2020).

2.2 Influencing factors

The information presented in this section is based on the review article "*Energy flexibility at multi-building scales: A review of the dominant factors and their uncertainties*" (Dawes et al., 2025). The developed work employs a systematic search methodology called Sub-keyword Synonym Subtopics Searching (SSSS). The keywords used are first "flexibility", "demand response" and "energy flexibility", and then "stochastic", "probabilistic", "probability", "risk" and "uncertainty". From this systematic search, and considering the Google Scholar database, 264 papers were screened and 121 were selected for detailed analysis (filter based on a citation threshold of 5 and ranking of the 10 most cited papers). This review study categorizes uncertainties impacting energy flexibility into aleatory uncertainty (i.e., randomness of real-world actions data, e.g., weather or occupancy patterns) and epistemic uncertainty (i.e., a lack of complete knowledge, e.g., incomplete data on building characteristics or energy systems).

2.2.1 Dominant factors

Dominant factors affecting energy flexibility from the built environment can be categorized into 5 segments:

- Occupant and customer behavior, which introduces both aleatory and epistemic uncertainties due to unpredictable human actions (such as appliance use) and the lack of comprehensive data on energy usage patterns. Addressing these requires better data collection and modelling approaches. Effective strategies must account for individual behaviors, engagement with energy flexibility services, and external factors like dynamic pricing and weather conditions.
- Building characteristics, which refers to the thermal mass of a building, its heat loss rate and its materials when focusing on buildings' thermal properties. Uncertainties in thermal properties and retrofitting measures (i.e. the building performance gap) lead to challenges in predicting a building's energy flexibility potential. Addressing this requires improved empirical data, real-time monitoring, and advanced models for evaluating the energy performance of different building archetypes and retrofit outcomes.
- Technology and HVAC, which includes energy storage (thermal and electrical). Here we discuss the critical nature of fast/active storage for energy flexibility, alongside self-generation and use of controls to deliver energy flexibility. Uncertainties arise, at this level, from incomplete knowledge about their configuration and interaction with other systems, as well as from external variability like weather conditions. To reduce the barriers and limit the uncertainty improved empirical data and real-time monitoring are key, as well as complete information on the systems and their capabilities.
- Control systems and optimization that utilize building energy management systems (BEMS) and advanced control algorithms. They play a significant role in managing energy flexibility due to their role of interpreting fuzzy or complex data on the grid/network side and turning this information into system actuation to deliver flexibility. However, uncertainties arise from the lack of accurate real-time data and the complexity of predicting energy demand and renewable generation. Improving data quality and system coordination is critical to enable and improve accurate energy flexibility quantification.
- External factors and interactions, such as weather conditions, energy market fluctuations, and interactions with other buildings or energy networks can significantly influence the overall levels of exploitable energy flexibility. Challenges and barriers in managing energy flexibility are mostly due to aleatory sources such as unpredictability in weather data/predications, fluctuating energy prices, and limited coordination between buildings and building clusters. Communication delays and insufficient data further complicate building-grid interactions and introduce epistemic uncertainty. Effective strategies to reduce these impacts include real-time data integration, improving forecasting accuracy, and regulatory adjustments to incentivize flexible energy behavior.

These five segments interact when activating flexibility in multi-building systems. Uncertainty propagation in energy flexibility systems arises when uncertainties from different sources interact and amplify throughout the modelling, control, and optimization processes, as illustrated in **Figure 6**. For example, a lack of knowledge on occupants' presence might lead to the wrong control decision and possibility a dissatisfaction from end-users. Similarly, it is not just the level of insulation in a building that is important in predicting energy storage, but also the amount and configuration of thermal mass.

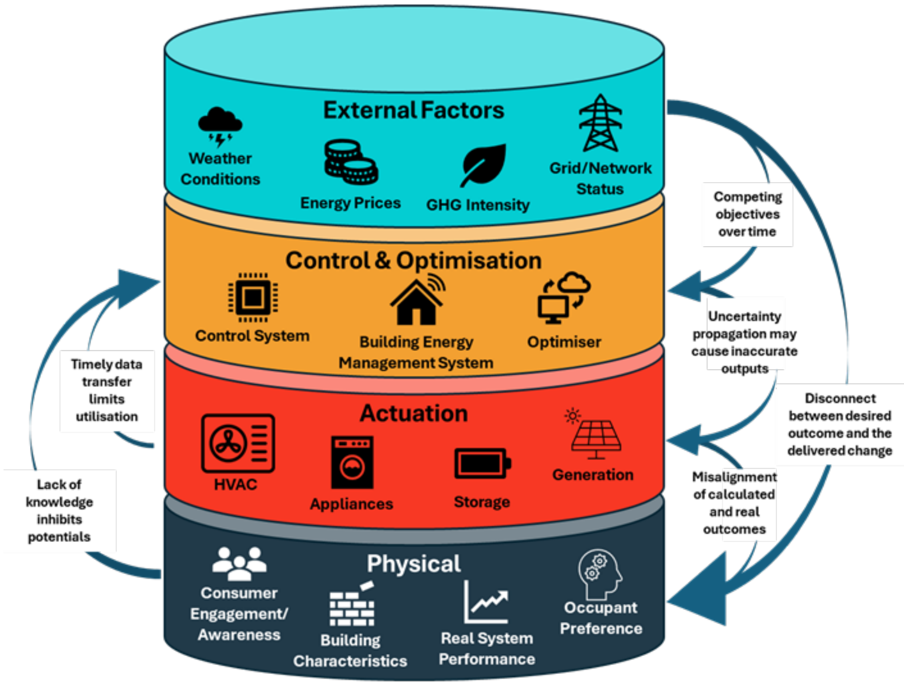


Figure 6: A physical representation of the interrelationships between a four layered feedback loop, considering the dominant factors of energy flexibility and the propagation of uncertainty.

2.2.2 Strategies for Mitigating Uncertainty

These dominant factors and their associated uncertainties can hinder the use of flexibility in practice and solutions need to be found to mitigate them. Reducing or managing these uncertainties will improve the confidence of the different stakeholders (e.g. better revenues for end-users, less risk for aggregators, more appropriate tariff structure for utilities). The following are examples of how some epistemic and aleatory uncertainties are managed when applying flexibility:

- Specific examples of aleatory uncertainty explored in the review work include weather variability, occupant and consumer behaviors and market dynamics. To address aleatory uncertainty, methods such as robust optimization techniques, scenario analysis, and adaptive control systems could be employed. For example, predictive control models that incorporate stochastic disturbances can provide insights into the range of outcomes, enabling more targeted and robust management strategies to leverage energy flexibility.
- Epistemic uncertainty is prevalent across many of the dominant factors discussed in the review, but mainly covered by data limitations, model simplifications and assumptions, and system interactions. To reduce epistemic uncertainty and better characterize the flexibility of buildings, a combination of data-driven approaches and advanced modelling techniques is essential. Studies have shown that continuously updating models with real-time data, using advanced sensing technologies, and employing sensitivity analysis can significantly enhance model reliability.

To realize energy flexibility at scale, aggregation is required which can help mitigating both aleatory and epistemic uncertainties by averaging individual variability, such as differences in energy consumption patterns or weather responses. For example, while one building may experience peak demand due to occupancy, another might have reduced load, thereby balancing the aggregate demand profile. An example of such an aggregation effect is shown in **Figure 7**, where the load shedding during a single flexibility event is evaluated for different occupant characteristics and different cluster sizes. The shaded area corresponds to the estimated confidence interval evaluated by Monte Carlo simulations. From these three numerical studies, it can be observed that the uncertainty is large (over 50-100%) for aggregations of 10 households, whereas it decreases to a few percent for aggregations of 1000 households. However, this aggregation can obscure significant differences that influence the overall energy flexibility potential.

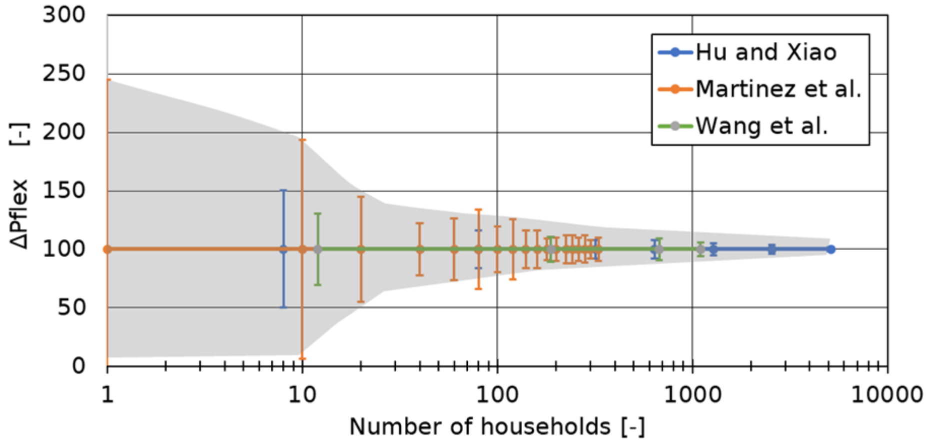


Figure 7: Variation in load during a flexibility event (dimensionless unit) with increasing number of households (shaded area corresponds to the estimated confidence interval).

A summary of mitigation strategies for key stakeholders can be summarized as follows:

- Occupants/customers can contribute towards uncertainty mitigation through increasing awareness and engagement of energy flexibility programs. Strategies include using smart appliances, real-time energy monitoring, and providing financial incentives such as cost savings or rewards. Empowering occupants with better information on energy use through user-friendly interfaces can improve participation. Additionally, wider use of automation tools can be used to help manage energy use in response to dynamic pricing signals at the single and multi-building scales – the seamless integration of this information from households to the aggregate level can help reduce unpredictability from human behavior in energy demand.
- Aggregators face challenges in managing diverse building energy demands and predicting their flexibility potential. Mitigation strategies include using real-time data for more accurate forecasting and clustering techniques to group similar energy profiles, improving the reliability of predictions, and diversity and redundancy of flexible resources. Aggregators can employ advanced optimization algorithms, such as stochastic modelling to handle uncertainties and reduce risks.
- Electricity utility companies need to have a reliable estimate of their energy need for the coming days/months/years. Reliable forecast, accounting for both technological and behavioral changes are thus necessary to optimize their investments. Encouraging wider participation of end-users through flexible pricing models and incentives might help balancing supply and demand and improve overall financial performance.
- Network operators need to manage aleatory uncertainties related to grid stability and capacity given by the external factors like weather and market fluctuations (mostly in the case of the transmission system). Mitigation strategies include improving communication infrastructure for real-time data sharing between demand and supply players. Regulatory reforms to incentivize energy flexibility and clearer market frameworks for demand response programs can reduce risks and enhance grid reliability and flexibility.

2.3 Flexibility and resilience

The information presented in this section is based on the review article "*Energy Resilience in the Built Environment: A Comprehensive Review of Concepts, Metrics, and Strategies*", published in Renewable & Sustainable Energy Reviews journal (Mingjun Wei et al., 2024). Flexibility and resilience are two different concepts, but both relate to the adaptation of the built environment to external factors. The work undertaken highlighted the similarities and differences between the two. Resilience and flexibility are deeply interconnected, particularly in the context of adapting to both anticipated and unforeseen extreme events, such as flooding, heat waves, winter storms, etc. While these two concepts are distinct, they often work side-by side to improve building performance, occupants' well-being and sustainability. For example, a building with PV and battery could be both resilient during a heat wave and flexible during a non-heat wave period through load shifting. The review work addresses four critical areas to systematically understand energy resilience: defining energy resilience, understanding relevant disruptions, quantifying resilience, and improving overall resilience. A significant contribution of this article is a comprehensive definition of energy resilience that includes attributes such as vulnerability, resistance, robustness, and recoverability. Under this context, four main research gaps were identified:

1. Lack of a universal definition: There is no consistent definition of energy resilience in the built environment, which hinders the ability to compare results across studies and develop comprehensive policies.
2. Understanding disruptions: A clear classification and understanding of the types of disruptions that affect energy resilience are lacking. The referred article categorizes disruptions into climatological, meteorological, and technology malfunction events, each requiring different resilience strategies.
3. Evaluation metrics: The metrics used to evaluate energy resilience are often context-specific and not universally applicable. The article emphasizes the need for metrics that consider the unique characteristics of different building types and energy systems.
4. Improvement strategies: Current building design and sizing methods may not provide sufficient resilience. The developed work identifies gaps in these methods and proposes strategies to enhance resilience, including innovative design approaches, technologies, and policies.

The developed work employs a systematic search methodology called Sub-keyword Synonym Subtopics Searching (SSSS) to identify relevant literature considering the Google Scholar database. This methodology ensured a comprehensive review of 357 papers, of which 104 were selected for detailed analysis based on expert domain knowledge. The keywords used are first "energy resilience," "thermal resilience," "energy availability", and "energy robustness", and then "disruptive event," "extreme climate event," and "power failure". The selected papers were summarized based on various factors, including disruptive event categories, resilience scope, resilience-related key performance indicators (KPIs), and strategies for enhancing resilience.

The remaining part of this section addresses the identified research gaps (Sections 2.3.1 – 2.3.4) and describes a framework proposed to evaluate and improve resilience (Section 2.3.5). It ends with the study conclusions and future research directions (Section 2.3.6).

2.3.1 What is resilience in the built environment?

The concept of resilience is first introduced by Holling in 1973 in the field of the ecological system, and this concept was used to describe the persistence of the ecosystem against fluctuations (Holling, 1973). Different studies and reports have used varying definitions and metrics to measure energy resilience. However, there is currently a gap in the literature regarding a universal definition of energy resilience in the built environment. This lack of a universal definition makes it difficult to compare and integrate findings across different research areas, thus making it a challenge to identify best practices and develop consistent policies and guidelines for improving energy resilience in the built environment. Thus, in this review work, energy resilience in the built environment is referred to as the inherent and adaptive capacity of buildings, infrastructure, and urban energy systems to anticipate, absorb, recover from, and adaptively respond to disruptions in energy supply and demand, while ensuring sustained functionality, efficiency, and equitability both in the short and long term.

Energy resilience features multiple phases which can be described as "resilience trapezoid" and "resilience triangle" (Charani Shandiz et al., 2020; Hosseini & Barker, 2016; Panteli et al., 2017). **Figure 8** outlines the stages a building or system undergoes after experiencing a disruptive event. Between t_0 and t_1, the system operates at its normal performance level. The disruptive event occurs at t_1, causing the system performance to degrade during "Disruption Stage 1" until it reaches its lowest performance level at time t_2. This point, referred to as the "degraded level" in the figure, signifies the worst performance experienced by the system. In the "resilience triangle" concept, the system immediately starts to recover at the end of "Disruption Stage 1", eliminating the need for a "Disruption Stage 2". The "Final State" stage follows directly after "Disruption Stage 1" as the system gradually returns to its pre-disturbance condition. However, in the "resilience trapezoid" concept, there is no immediate recovery action taken at the end of "Disruption Stage 1". This results in a constant degraded performance level during "Disruption Stage 2."

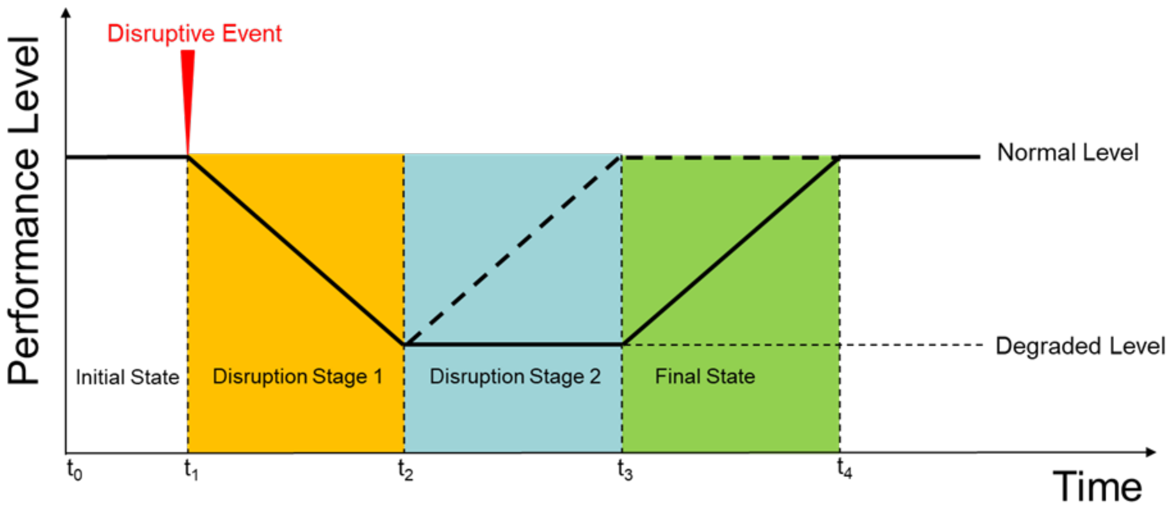


Figure 8: Resilience “trapezoid” and “triangle”.

2.3.2 Classification of disruptive events

The review categorizes disruptive events into three main types:

1. Climatological events: Long-term weather patterns and climate anomalies, such as global temperature rise.
2. Meteorological events: Short-term extreme weather events like heat waves and winter storms.
3. Technology malfunction events: Equipment failures due to aging or random malfunctions, often exacerbated by climatological or meteorological events.

The work reported in (Mingjun Wei et al., 2024) highlights that heat waves and component failures are the most frequently analyzed disruptions in the literature.

2.3.3 Strategies for improving resilience

The review identifies and categorizes resilience enhancement strategies into five main groups:

1. Design: Early stage building design strategies, such as optimizing building orientation, selecting materials with high thermal mass, and incorporating passive design features.
2. Retrofit: Strategies to improve existing buildings, including optimized building envelope retrofitting, combining mitigation and adaptation strategies, and employing passive cooling techniques.
3. Predictive control: Advanced heating, ventilation, and air-conditioning (HVAC) control systems and natural ventilation to enhance building resilience.
4. Micro-grid: Implementation of micro-grids with renewable energy sources and energy storage systems to maintain functionality during power outages.
5. Other approaches: Urban planning measures, policy and standard establishment, and training for building operators and end-users.

In summary, the commonly used resilience improvement strategies in terms of design, retrofit, predictive control, micro-grid and other solutions are provided. Among them, design strategy focuses on the early stage of buildings' life cycle, which can improve the building resilience more effectively and efficiently with lower capital investments. Retrofit aims at improving target building environment especially under extreme weather conditions. Predictive control strategy can increase the building resilience by optimized control algorithm which probably is the most cost-effective and viable approach. Micro-grid demonstrates great resilience potentials during an island mode under power outage. The other solutions stand from policy maker or urban planner point of view, which also provides possible resilience enhancement. For different stakeholders who are affected by various disruptive events, it is crucial to identify and implement tailored resilience improvement solutions.

2.3.4 Proposed framework to evaluate and improve resilience

In terms of evaluating resilience metrics, a workflow is proposed in choosing appropriate resilience metrics. Taking a case study from Flores-Larsen and Filippín (Flores-Larsen & Filippín, 2021) as an example, the following steps are considered:

1. Stakeholder identification: As in the previous scenario, the occupant is the stakeholder.
2. Scope/level determination: The focus remains on the building level.
3. Goal of energy resilience: The goal could be to ensure occupant comfort and safety during disruptive events.
4. Check subset key performance indexes (KPIs): The occupant should check if existing KPIs, such as heating index, indoor overheating degrees, and discomfort index during disruptive events provide enough information about their building's resilience.
5. Development of new KPIs: If existing KPIs are not sufficient, the occupant might need to develop new KPIs. For instance, they might need to consider how quickly their building can be operational after an extreme weather event, or how comfortable occupants are during such events, if such metrics have not been previously considered.

It is important to note that each scenario is unique and different, so the most important action is to understand the specific conditions, interests, and capabilities of the stakeholder in question, and apply the framework accordingly.

2.3.5 Concluding remarks

This work, reported in (Mingjun Wei et al., 2024), emphasizes the importance of enhancing energy resilience in the built environment to mitigate the impacts of disruptions. The authors propose a comprehensive definition of energy resilience and stress the need for context-specific metrics and strategies. The work also highlights the potential for future research to explore the minimum performance levels for resilience metrics and the relationship between energy resilience and flexibility in buildings. Buildings and building clusters that are energy flexible can improve the resilience of energy networks by reducing the load on the infrastructure. This also makes buildings more resilient to fluctuations in energy supply. When assessing resilience, it is often related to a building's ability to change its electricity use during a grid service or in the face of severe weather. Energy flexibility and resilience tend to have a close, sometimes inverse, relationship, as measures to improve resilience, such as reserving battery power for emergencies, can reduce the building's flexibility.

This review serves as a foundation for developing more effective strategies and policies to enhance energy resilience, ultimately contributing to more resilient and sustainable buildings. Future work could focus on the design and optimization of adaptive building and district energy systems and controls to uncover the relationship and balance between system flexibility and resilience.

3. Drivers for the use of flexibility and characterization methods

This section presents different methods and tools to help characterize and exploit energy flexibility. To activate this flexibility, external signals from the grid are needed and they can be of different types (section 3.1); short-term forecasting methods are also required to assess flexible assets and to estimate the potential load variation (section 3.2); finally, some characterization methods have been developed within Annex 82 to estimate the flexibility of buildings controlled according to a dynamic signal (section 3.3).

3.1 Activating buildings – Building-Grid Interaction signals

To activate a building's energy flexibility potential (whether thermal, electrical or other), an activation signal is required to allow a decision to be made when and how much flexibility is required. At the single building scale, decentralized control can be made by an occupant or facility manager through a building management system to optimize the building's/household's objectives. Whereas at the multi-building scale, a centralized controller may be used by the network operator or aggregator as this is more likely to provide a more significant flexibility response to the signal.
Due to further research on the acceptability of naming conventions and to enable the widespread adoption of flexibility (Langevin et al., 2024), we propose the use of a new term: Building-Grid Interaction Signal, or BGI signal. A key part of this research was to conduct a literature review to identify the most used flexibility BGI signals for building energy flexibility and DSM, how they were used and how they enable energy flexibility services. This review aimed to explore the impacts and use cases at the cluster level. However, it was noted that many of the example BGI signals were from studies which focused on single buildings - this does not, however, affect the applicability of the type of BGI signal used at the cluster scale. In the remaining of this section, we discuss several topics derived from the review which relates to energy flexibility and the use of different BGIs – specifically we discuss:

- Naming conventions for BGI signals;
- The different types of BGI signals used to trigger energy flexibility;
- How different BGI signals can achieve different service objectives;
- Use and relevance of BGI signals to stakeholders.

Where possible, we refer to the most relevant or state-of-the-art research and commercial applications to demonstrate both the theoretical and practical implementations of energy flexibility BGI signals.

3.1.1 Naming conventions for energy flexibility signals

In the scientific literature on energy flexibility in buildings, the term "penalty signal" refers to a mechanism that encourages reduced energy usage during specific periods by applying a penalty, typically to energy use. Junker et al. describe it as a system where a higher penalty prompts a faster or more significant demand reduction (Junker et al., 2018a). While their work used price as the driving factor, other signals like carbon intensity or renewable energy availability could also be applied.

More commonly, the term "price signal" is used in practice due to its alignment with market functions and energy costs. Such signals are produced with the aim of reducing expenses by lowering energy usage during “high signal” periods (Hall & Geissler, 2021; Junker et al., 2018a). The use of a “price signal” assumes the normalization of variables under a per cost basis; for example, applying a carbon price or carbon tax to the carbon intensity of the delivered electricity (CO_2e/kWh) transforms it into a price signal (cost/kWh).

The term "penalty signal", however, can carry negative connotations, making it less suitable for stakeholder engagement, especially in the context of business and commercialization of energy flexibility and DSM services (Langevin et al., 2024). The use of terms such as "penalty signal" and "flexibility control signal" are also too limited in their utility. Although "flexibility control signal" is neutral in its understanding, the use of such signals should highlight the interactions of buildings, grids and their external networks.

To offer a more neutral and specific alternative, we propose the term **"**building-grid interaction signal" (BGI signal). A BGI signal can be defined as “*a dynamic signal that can prompt adjustments in a building’s systems or processes to align with operational objectives and external factors such as grid service requirements*”. A BGI refers to any dynamic signal that prompts adjustments in a building's operations to meet both internal (customer/consumer) objectives and external grid requirements. Unlike penalty signals, a BGI signal is not cost-specific and can address multiple control factors, making it more appealing and engaging for stakeholders involved in DR programs. Building-grid interaction signals serve optimization goals such as reducing energy bills (Junker et al., 2018a), lowering carbon emissions (Hall & Geissler, 2021), or enhancing grid stability (Paterakis et al., 2015).

3.1.2 Different forms of Building-Grid Interaction signals

Different types of BGI signals can be applied to a building, each of which has its own reason for its use and is generally aligned with some sort of optimization or energy flexibility service goal. The application of various types of BGI signals can be categorized as follows:

1. <u>Single grid interaction signal</u>: A single signal can be used to minimize or maximize (a) particular objective(s), such as dynamic time-of-use pricing of energy to reduce operational costs, or dynamic CO_2 intensity to reduce carbon emissions as a result of energy use. An example of using one signal to produce contracted energy flexibility in a building cluster was described by Geneidy and Howard in (El Geneidy & Howard, 2020).
2. <u>Combined grid interaction signal:</u> Multiple parameters can be balanced and considered simultaneously, e.g., energy cost and CO_2 intensity can be evaluated simultaneously to produce one signal. This requires pre-processing either at the aggregator or single-building scale to produce a BGI value at each time step based on the pre-defined balancing of service objectives. The processing, therefore, depends on a decision tree and it is often described by a weighted balance for each signal according to importance, or the output of the objective function. An example of this application was explored by Mitic and Voss in (Kırant Mitić & Voss, 2023).
3. <u>Sequential grid interaction signals:</u> Different signals can be applied at the same time, but only one signal is “active”, due to a pre-defined priority sequence. For example, the baseline flexibility objective could employ a price-sensitive BGI signal to minimize cost, but when a more critical signal is triggered, then the system would operate to serve this new objective. A hypothetical example could be the use of a distribution transformer temperature signal, noting that when it exceeds a certain limit the buildings must react immediately otherwise the transformer may be damaged and further instability issues may ensue. During this time the transformer signal relating to the transformer temperature has priority and would override the baseline signal.

3.1.3 Aligning signals with service objectives

Within the context of increasing renewable generation to reduce the carbon intensity of supplied electricity, energy flexibility services play a crucial role in managing energy demand and integrating renewable energy sources. These services rely on signals that prompt adjustments in energy use to align with grid needs (Al-Mashhadani & Kurnaz, 2024). As described by (Gellings, 1985), there are several load-shape objectives which can be achieved by providing an energy flexibility service, as shown in **Figure 9**.

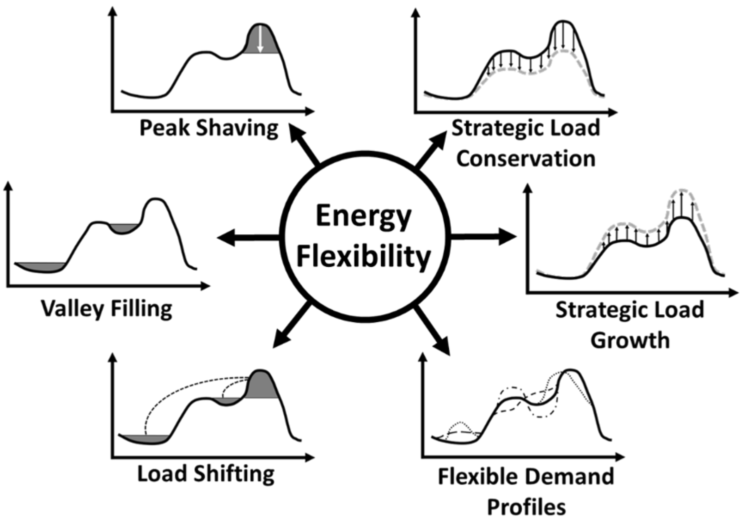


Figure 9: An adapted image from (Gellings, 1985) illustrating how different energy flexibility services can impact energy load profiles.

The use of single or multiple BGI signals can be one way of achieving such a variety of objectives. In addition to the service objectives, there are several types of signals which can be used to achieve different strategic objectives at the multi-building scale:

- Price signals are one of the most common forms in energy flexibility services provision. They use dynamic pricing models such as time-of-use or real-time pricing to encourage consumers to reduce or shift their energy use during peak periods. By aligning energy use with lower-priced periods, consumers can save money while helping balance grid demand (Junker et al., 2018a). This has the effect of allowing customers to save money whilst balancing the grid to reduce peaks by load shifting.
- Emissions-based signals aim to reduce greenhouse gas emissions by encouraging energy use when grid emissions are lower, or when renewable energy is more available (Clauß et al., 2019). Signals of this type can serve consumers by reducing their carbon footprint if they use grid-derived energy. However, it can also be a self-sufficiency objective, where the on-site generation can be maximized and therefore reducing reliance on more carbon intensive grid electricity.
- Grid-based signals can be used to maintain or achieve grid stability and resilience. The use of these signals is not widespread, mainly due to the lack of smart-grid infrastructure at various levels of the energy system (Ohanu et al., 2024; Powell et al., 2024) and the lack of regulatory basis for incentivizing action. Further, grid/network-based signals can only be applied to building clusters of a certain size, due to the required response rate being met from many tens, hundreds or thousands of buildings (depending on the grid scale) (Hall & Geissler, 2020). Under this context, the following should be taken into consideration:
    a. *Frequency regulation or voltage control signals* can help maintain grid stability by responding to real-time frequency deviations (Kim et al., 2017). When supply and demand become unbalanced, frequency regulation signals can prompt adjustments, ensuring that the grid operates within safe frequency limits. Voltage control signals ensure a consistent power supply

across the grid by monitoring and adjusting voltage levels, to help prevent fluctuations that could damage equipment or disrupt service (Ding et al., 2013).

b. *Load-based flexibility* focuses on shifting or modulating (through increase or decrease of) electrical loads to enhance grid stability, especially during peak demand periods (Darwazeh et al., 2022). Load shifting signals in energy flexibility programs can be used to reduce strain on the grid when either excess or lack of power has been forecast. This happens over a longer timeframe than frequency or voltage control signals and therefore can form part of a grid-focused load-shaping exercise.

3.1.4 Use and relevance of Building-Grid Interaction signals to stakeholders

BGI signals hold varying degrees of relevance for different stakeholders across the energy system. Understanding these differences is key to unlocking the full potential of stakeholders in dynamic energy flexibility. The key stakeholders that may find relevance in this work include:

- Grid/network operators;
- Energy service companies (ESCOs) or aggregators;
- Building owners and managers;
- Policymakers and Regulators.

For grid operators, the principal value of BGI signals lies in their ability to trigger load shaping exercises to smoothen the integration of variable renewable energy sources. When wind or solar generation spikes or dips unpredictably, the grid must maintain a balance between supply and demand to ensure grid stability and resilience. Traditionally, grid operators have relied on flexible generation assets such as gas turbines to provide this balance. However, the role of the built environment – at scale – can support this crucial task of demand-side flexibility. This can be done by allowing grid operators to treat clusters of buildings as virtual power plants, reducing or increasing demand to match the real-time supply (O׳Connell et al., 2014). To do this, BGI signals are used for clear, concise, and accurate communication between the network and buildings.

For building owners, particularly in the commercial and industrial sectors, BGI signals represent an opportunity to reduce operational costs, or generate revenue, through dynamic energy management. As originally hypothesized by (Gellings, 1985), buildings with a high-demand or building clusters can manage their operation to generate some form of income through participation in these services. Buildings equipped with flexible systems such as smart thermostats, hot water thermal storage, battery storage, and electric vehicles (EV) charging stations can respond to BGI signals by modulating or shifting loads away at peak times, thereby lowering energy bills or receiving compensation from flexibility programs (Afroz et al., 2023; Chen et al., 2021). This is particularly relevant in regions with time-of-use pricing, where electricity rates vary throughout the day based on grid demand. In such regions, there are growing incentives for buildings to participate in flexibility services (D'hulst et al., 2015).

Aggregators play a crucial role as intermediaries between customers (building owners and occupiers) and grid operators. These companies can manage contracted flexibility services and provide the technology infrastructure required for buildings to participate in energy flexibility programs (El Geneidy & Howard, 2020). ESCOs benefit from BGI signals by providing services both downstream (to customers) and upstream (to operators), such as automated energy management solutions that optimize a building's interaction with the grid. By aggregating the flexibility of multiple buildings, ESCOs can provide larger-scale services to grid operators, making them key players in the implementation of BGI signals (Stinner et al., 2016).

For policymakers, the use of BGI signals is central to meeting climate and energy policy goals. Regulatory bodies are beginning to establish standards and protocols for demand-side flexibility, with a particular focus on enhancing the role of buildings in supporting grid reliability (Albadi & El-Saadany, 2010). By aligning the use of BGI signals with the outcomes of cheaper energy for consumers, lower carbon intensity energy, and stability and reliability for grid operators, policy makers can tackle multiple problems simultaneously.

3.2 Characterizing buildings – Forecasting loads

Energy forecasting plays a central role in the operation and planning of electricity grids (Hong et al., 2020). On short time scales, generation and demand are forecasted to ensure stable system operation and mitigate adequacy risks. Likewise, these demand forecasts also often form the basis for establishing flexibility baselines and counterfactuals. At longer time scales, energy forecasts are also critical to inform investment decisions.

Energy forecasts can take different forms with respect to dimension and statistical properties, which affects how they should be evaluated. Traditionally, the use of (deterministic) point forecasts have been the most common, but there is an increasing demand for solutions that can account for statistical uncertainty. Such solutions are called probabilistic forecasts, and they can for example be full forecast distributions or consist of a set of quantiles (Sørensen et al., 2023). Furthermore, forecasts at different points in time are usually correlated, and thus multivariate, which is indeed the case for load. Taking this temporal correlation into account has been shown to yield significant improvements of electricity load forecasts (Nystrup et al., 2020) and heat load forecasts (Bergsteinsson et al., 2023).

Despite its widespread use, the systematic benchmarking and evaluation of competing methodologies represent a comparatively recent development. For instance, a recent study reveals wide-spread issues with national load and renewable generation forecasts in the European Union (Kazmi & Tao, 2022). Likewise, despite the existence of clear guidelines in the broader forecasting community (Hewamalage et al., 2023), energy forecasting and modelling communities continue to report predictive accuracy using different non-standardized metrics (Johra et al., 2023). This means that different studies often arrive at conflicting conclusions, which cannot be generalized to develop broader consensus on which methods work best for energy forecasting.

A popular approach to address these issues has been through forecasting competitions (Hyndman, 2020), which provide participants the same information in the hopes that the best forecasting methodology will emerge as the winner. The M competition series, initiated by Spyros Makridakis in 1982 and now in its sixth iteration, is arguably the most well-known in the field of forecasting (Makridakis et al., 2022), but is not focused on energy related problems in general. Unlike the M competitions, the ASHRAE Great Energy Predictor (GEP) and Global Energy Forecasting Competition (GEFCOM) have run through several iterations focusing on energy problems. The ASHRAE GEP competitions focused on data-driven building energy prediction using machine learning techniques, with GEP 1 focusing on energy load prediction, GEP 2 on energy data and retrofits, and GEP 3 on predicting metered building energy usage (Miller, Arjunan, et al., 2020). On the other hand, GEFCOM competitions targeted hierarchical energy demand forecasting among other use cases, with each iteration requiring point forecasts and specific quantiles for uncertainty quantification.

These competitions have successfully highlighted the importance of pre-processing noisy data and feature engineering. Often, the effectiveness of ensemble tree-based methods (e.g. those utilizing boosting methods such as LightGBM or XGBoost) compared to more complex neural networks has also been demonstrated, either in accuracy or in the effort required to build the winning models. This raises the obvious question: should researchers devote additional time, effort and (computational) cost into developing these more complex models, or should they instead focus on the end-to-end pipelines?

Despite their growing popularity, forecasting competitions face several challenges that limit their ability to replicate real-world scenarios. They often rely on static, limited datasets for evaluation, which may not generalize well to real-world situations, leading to overfitting and inflated accuracy scores. Variations in data quality and availability further complicate emulation of real-world conditions. Secondly, forecast competitions primarily fix both evaluation metrics and horizons upfront, which does not consider the complexity and diversity of relevant real-world situations. Lastly, many forecast competitions tend to emphasize accuracy over practical considerations such as scalability, applicability to downstream tasks, explainability, and computational complexity, which are equally (if not more) important in real-world applications.

Within Annex 82, the open-source benchmarking framework EnFoBench was developed to address these challenges by creating a community-driven tool for forecast creation and evaluation (Balint et al., 2024). This allows researchers and practitioners alike to quickly determine which forecast strategies work well for which kind of end use case (e.g. forecasting hourly energy demand in commercial buildings). The following sections describe the architectural choices behind this framework, the data utilized for benchmark purposes, and the results obtained to date. The section concludes with some reflections on future directions.

3.2.1 The EnFoBench framework

The Energy Forecasting Benchmark (EnFoBench) extends the idea behind the Monash Time Series Repository (MTSR) (Godahewa et al., 2021). MTSR moves beyond static forecast competitions by evaluating various forecasting models on 30 different datasets with different temporal resolutions. EnFoBench, on the other hand, is designed specifically for energy applications like electricity and gas demand forecasting as well as renewable generation forecasting. The framework comprises of three main components: open-source curated datasets[1], a Python package[2] with evaluation scripts and metrics, and a live dashboard[3] () displaying benchmark results. In addition to being a forecasting benchmark, it also provides researchers with access to well-curated data, relevant features, and examples across popular frameworks. Next, we describe the framework's most salient features, including the comprehensive evaluation framework, community aspects and evaluation metrics.

Comprehensive Evaluation of Forecasting Processes
Recognizing the complexity of evaluating forecasting methods, EnFoBench avoids relying on a single metric and instead reports multiple evaluation metrics for a comprehensive assessment of model performance. A core guiding principle behind EnFoBench is therefore to benchmark not just the forecast model, but the entire forecasting pipeline: i.e., encompassing pre-processing, feature engineering and selection, and pre-training or fine-tuning of global models. While this adds complexity, it addresses a common omission in competitions where non-model components of forecasting pipelines are often excluded or rendered irrelevant through the use of clean datasets. Recognizing that datasets are rarely clean in real-world scenarios, we hope that EnFoBench will contribute to research in robust and scalable forecasting for the energy sector.

The high-level architecture of the benchmarking service is shown in **Figure 10**. This service ensures that every model is evaluated on every building in the test set. To allow faster evaluation, the individual runs can be parallelized, through containerization. Utilizing separate containers also ensures full reproducibility and eliminates dependency clashes between different models. Each container is served by a FastAPI-based HTTP server, equipped with only two endpoints: one for accessing model information such as parameters and metadata, the other for inference. The containerization provides further advantages such as prevention of data leakage between buildings, whereas models could transfer knowledge from previous test buildings to later runs. To eliminate data leakage as much as possible, the test set is created from private, previously unpublished data.

[1] https://huggingface.co/datasets/EDS-lab/electricity-demand
[2] https://github.com/attila-balint-kul/energy-forecast-benchmark-toolkit
[3] https://huggingface.co/spaces/EDS-lab/EnFoBench-ElectricityDemand

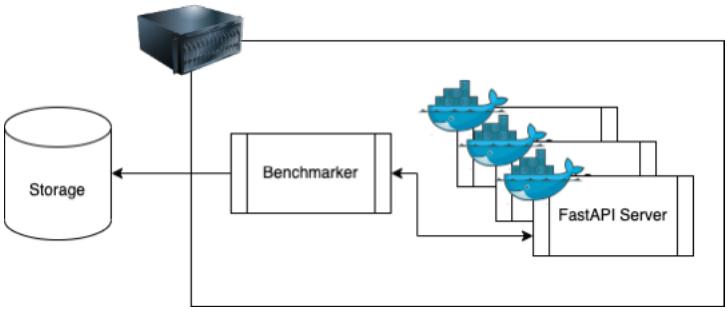


Figure 10: The EnFoBench evaluation architecture, which ensures each (newly added) building is evaluated on the full suite of forecast methods and evaluation metrics

The detailed process of the benchmarking service is illustrated in **Figure 11**. Initially, the relevant building data is loaded into the system's memory. Then, the service pauses execution until the model server is online and ready to handle requests. Once the model server is online, the evaluation commences through time-series cross validation. This is orchestrated by sequentially requesting forecasts from the model server. The selection of historical data and future covariates are managed by the benchmarking service. For each prediction request, only the relevant data for that prediction time is transferred to the forecasting service to prevent data leakage.

In the first use-case of load forecasting, the cross-validation emulates day-ahead market operations over a full year, comprising 365 folds, each representing a day. These evaluations occur daily at 10 AM and span a horizon of 38 hours (until the end of the subsequent day). The number of steps in this process varies based on the dataset's resolution, equating to 38 steps in hourly resolution and 72 steps in half-hourly resolution. Once the cross-validation step is completed, each prediction is stored in the system storage alongside its corresponding cutoff date (i.e., when the forecasts were requested). Subsequently, the selected metrics are computed for each cutoff date, followed by the calculation of aggregate statistics. Both daily and aggregated metrics are then saved onto the system's storage. Running the cross-validation and metrics calculation separately, while storing both, offers more detail and flexibility at the same time, as new metrics can be added to the results without necessitating the rerunning of the resource-intensive cross-validation step. Finally, the aggregated metrics are logged in the experiment tracking software (in our case Weights & Biases) and made available in real-time on a publicly accessible dashboard.

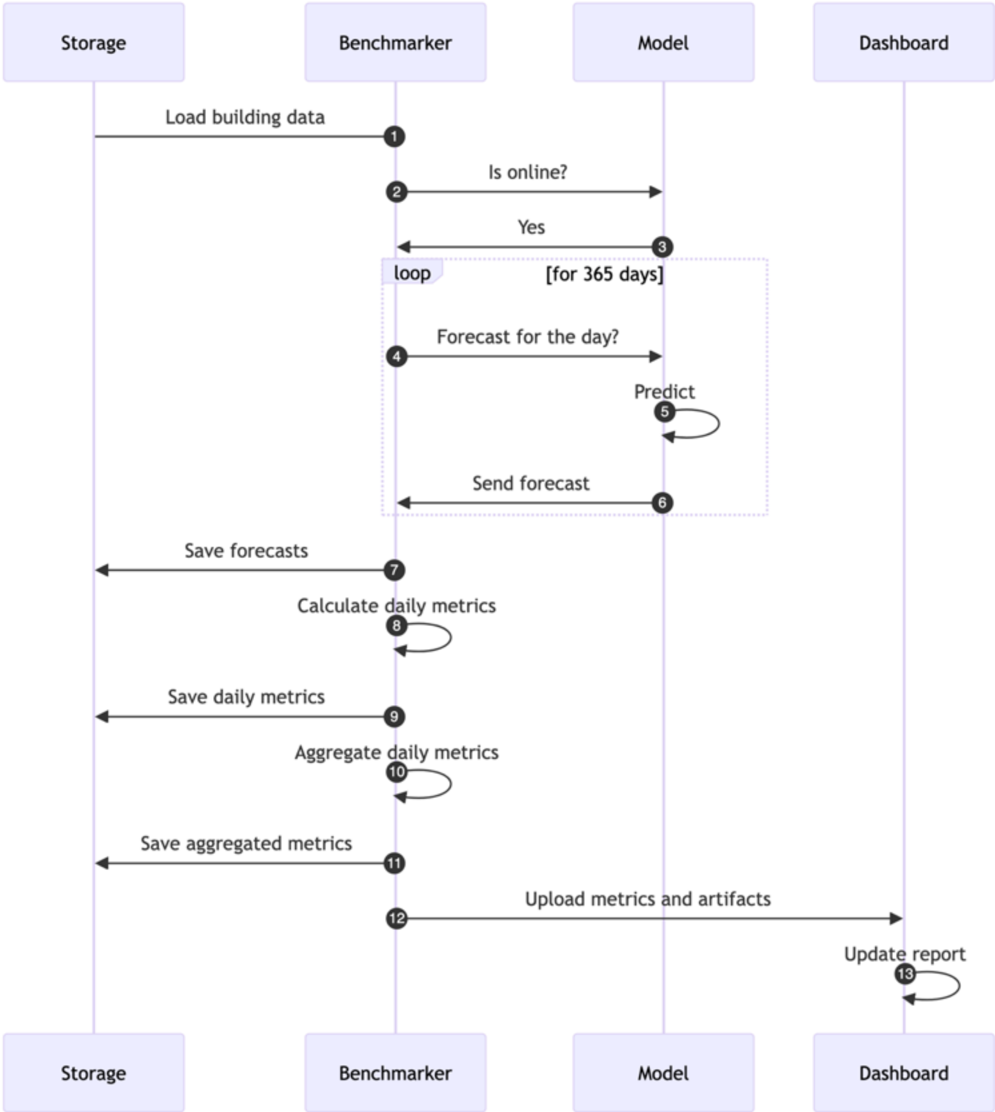


Figure 11: A detailed visualization of the EnFoBench workflow

Community driven benchmark

The community aspect of the framework lies at the heart of its mission. Serving as a foundational pillar, we intend for this benchmark to evolve into a living repository, continually expanding to contain a diverse array of end-to-end forecast pipelines applied across various energy use cases. The dashboard is then automatically updated based on the inputs from participants and the work can be continued in the IEA EBC Annex 96. To ensure a smooth on-boarding process for researchers of all levels, fully open-source example models based on popular forecasting frameworks are readily available, as well as a continuously expanding online documentation[4].

The collection, aggregation, and curation of open-source datasets, alongside the sourcing of relevant external features, significantly reduces the upfront time investment required by researchers entering this field. The EnFoBench python package serves as a dynamic codebase, accommodating common tasks in data analysis such as pre- and post-processing. Additionally, the defined standard interface of forecasting models allows the utilization of these models beyond the realm of benchmarks, as the published container images can be

[4] https://attila-balint-kul.github.io/energy-forecast-benchmark-toolkit

integrated into other processes such as building simulation, thus extending the benefits of EnFoBench to communities outside of traditional energy forecasting.

3.2.2 Metrics

EnFoBench offers a diverse array of metrics designed to provide comprehensive insights into forecasting performance. These include accuracy-related metrics and computational resource metrics, although more can be added easily.

Accuracy-related metrics
To evaluate model accuracy the following metrics are utilized in EnFoBench: Mean Absolute Error (MAE), Mean Bias Error (MBE), and Root Mean Square Error (RMSE). These metrics are evaluated separately for every prediction during the cross-validation and stored with their respective cutoff-points. These detailed (per-prediction) metrics are then aggregated using various statistical measures including mean, minimum, maximum, median, and standard deviation. These metrics are scale-dependent in the sense that they rely on the magnitude of the time series being forecast. As EnFoBench's objective is to benchmark models across a broad spectrum of series, we also incorporate scale-independent metrics into the benchmark in the form of relative error metrics. These metrics calculate errors per cross-validation and normalize them by a benchmark model's error for the same cross-validation period before aggregation.

EnFoBench utilizes two different naive seasonal models as benchmarks, a daily persistence and a weekly persistence model. Additionally, a third relative metric simply uses the (historically) better performing of these two naive seasonal benchmark model for each building, under the assumption that the appropriate naive model can be estimated based on the building's historical data. This adaptive benchmarking strategy provides a stronger baseline model for comparison and is particularly relevant given the diverse nature of building demand (i.e. different building types may show different predominant seasonal patterns). Traditional scale-independent metrics such as Mean Absolute Percentage Error (MAPE) produce asymmetric estimates and encounter difficulties when dealing with series containing zeros, which are prevalent in many energy use cases like net load with on-site photovoltaic (PV) production. Consequently, EnFoBench omits the reporting of MAPE.

For evaluation on the multivariate and probabilistic level, it is appropriate to choose metrics which are sensitive to misspecified correlation structures and poorly calibrated forecast distributions. State-of-the-art methods to address these include the Continuous Ranked Probability Score (CRPS) for probabilistic forecasts and the Variogram Score (VarS) for multivariate forecasts. Alternatively, the Logarithmic Score (LogS) can handle both the multivariate and probabilistic aspects simultaneously, but is not always feasible to implement, especially for high-dimensional problems (Bjerregård et al., 2021). An open-source R package which offers an implementation of these metrics is currently being developed under the IEA Wind Task 51 and is expected to be publicly available in 2025 (Möhrlen et al., 2023). It should be noted that these probabilistic forecasts are not yet implemented in EnFoBench.

Computational resource metrics
In addition to accuracy metrics, assessing the computational resources utilized by a forecasting model is crucial to understand its practical feasibility and scalability. Key computational metrics include CPU time in seconds, memory usage in megabytes, and Docker image size in gigabytes. These metrics are especially relevant with forecast models relying on deep neural networks etc.

3.2.3 Datasets

Any good forecasting benchmark must rely on a large training and test dataset for its conclusions to be statistically significant. Here, we describe the datasets that currently underpin EnFoBench.

Data, metadata and exogenous variables

Forecasting in the realm of energy presents unique challenges, particularly regarding data availability. While there are open-source datasets available for some use cases, a lack of standardization complicates their utilization. Each dataset typically adheres to different schema, employs different timestamps (UTC, local time with timezone info or without), utilizes different units of measurement, and adopts different approaches to handling missing data. EnFoBench therefore collates various open source datasets, and systematically gathers the same features such as weather data for every series in the dataset, ensuring consistency and comprehensiveness. This unified dataset is openly accessible for download, facilitating easy access and utilization. Moreover, EnFoBench integrates these datasets into its Python package, making it straightforward for researchers to access it. The public training set for electricity demand dataset currently aggregates the following datasets: *SmartMeter Energy Consumption Data in London Households* (UK Power Networks, 2024), *Building Data Genome Project 2* (Miller, Kathirgamanathan, et al., 2020) and Electricity Load Diagrams 2011-2014 (Trindade, 2015). The test set however, is sourced privately and is not publicly accessible to eliminate potential data leakage.

The datasets adhere to a consistent structure, consisting of three main tables: a metadata table, a data table, and a weather table. The metadata table lists every available metadata for each target series, including approximate location data of the building and the building's designated use type (which is also utilized as a dimension for model evaluation). The data table contains the raw measurement data for each target in the dataset. Finally, the weather table encompasses exogenous features sourced from Open-Meteo, an open-source weather API that offers free access to high resolution historical weather data, as well as weather forecasts. The list of weather features include temperature, relative humidity, dew point, rain, snowfall, surface pressure, cloud cover at different altitudes, wind speed and gusts, and many more. Each of these tables is described in detail below.

Metadata table

The metadata table lists every available metadata for each target series. Due to privacy concerns, the open-source datasets typically contain limited metadata, and the provided fields are rarely consistent across different datasets. However, EnFoBench mandates certain metadata to be included in the dataset for essential purposes. These include approximate location data of the building, crucial for sourcing weather data, and the building's designated use type, utilized as a dimension for model evaluation.

Data table

The data table contains the actual measurement data for each target in the dataset. Before inclusion, every target series undergoes a rigorous inspection and cleaning process. Series with large number of missing values or poor-quality data offer little added value and therefore are discarded. Clear anomalies are identified and replaced with NaNs, maintaining data integrity, similarly, to missing timestamps.

Weather table

The weather table encompasses exogenous features sourced from Open-Meteo, an open-source weather API that offers free access to high resolution historical weather data, as well as weather forecasts. The list of weather features includes temperature, relative humidity, dewpoint, rain, snowfall, surface pressure, cloud cover at different altitudes, wind speed and gusts, and many more.

3.2.4 Results

The results presented in this chapter represent a snapshot of the current state of the EnFoBench benchmark, which is a continuously evolving framework. Up-to-date results are always available on the dashboard. This section is divided into three parts: accuracy, computational resources, and relative performance. Figure 12 provides a brief overview of how the leaderboard looks like, with forecast models down one axis and individual buildings (#1 to #25) along the other axis. The values indicated in this figure corresponds to relative mean absolute errors, i.e. the error obtained by the model indicated in bold along each row, divided by the error obtained with using a simple persistence model (i.e. past predicts future). For energy demand, a weekly persistence is applied while for solar production, a daily persistence model works better.

model	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23	#24	#25
Darts.LinearRegression.DirectMultiModel.7D	1.06	1.01	0.99	0.95	0.94	1.12	1.08	1.03	0.81	0.81	0.85	0.93	0.82	0.91	0.89	0.76	0.88	0.83	0.83	0.85	0.89	0.81	0.81	0.82	0.76
Darts.LightGBM.Recursive.7D	0.90	0.92	0.92	0.79	0.88	1.08	0.98	0.91	0.79	0.82	0.86	0.82	0.84	0.96	0.91	0.76	0.94	0.94	0.83	1.08	1.27	0.95	0.81	1.18	0.79
Darts.LightGBM.DirectMultiOutput.7D	0.94	0.93	0.96	0.87	1.03	1.19	1.15	1.11	0.81	0.84	1.03	0.98	0.96	1.01	0.84	0.80	0.85	0.88	0.92	0.89	0.91	0.85	0.84	0.83	0.84
Darts.LinearRegression.Recursive.7D	1.27	1.21	1.18	1.08	0.98	1.13	1.08	1.02	0.92	0.88	0.85	1.01	0.82	0.93	0.90	0.82	0.89	0.87	0.85	0.85	0.88	0.82	0.84	0.82	0.79
Darts.LinearRegression.DirectMultiOutput.7D	1.19	1.17	1.13	1.06	1.04	1.55	1.24	1.27	0.90	0.91	0.99	1.16	0.93	1.08	0.93	0.86	0.94	0.91	0.92	0.91	0.94	0.86	0.89	0.84	0.84
Statsforecast.SeasonalNaive.7D	1.00	1.00	1.00	1.00	1.00	1.57	1.59	1.62	1.00	1.00	1.74	1.68	1.77	1.21	1.00	1.00	1.00	1.00	1.26	1.16	1.19	1.16	1.00	1.12	1.00
Amazon.Chronos.T5.Base	2.79	2.28	2.24	1.04	1.03	0.99	1.27	1.14	1.39	1.40	1.24	1.06	1.24	1.04	0.88	1.23	0.86	1.21	1.12	1.07	0.81	1.03	0.86	0.82	0.83
Darts.LightGBM.Recursive.1D	2.07	2.00	1.94	1.88	1.72	1.17	1.23	0.98	1.23	1.25	0.91	0.87	0.87	1.03	1.49	1.05	1.38	1.01	0.90	1.10	1.17	0.99	1.03	1.01	0.93
Amazon.Chronos.T5.Small	2.92	2.54	2.28	1.05	0.99	1.04	1.26	1.07	1.51	1.55	1.27	1.04	1.21	1.04	0.93	1.30	0.88	1.27	1.14	1.03	0.81	1.01	0.89	0.81	0.84
Amazon.Chronos.T5.Mini	3.09	2.71	2.43	1.04	1.06	1.07	1.27	1.15	1.61	1.65	1.34	1.11	1.24	1.03	0.91	1.35	0.90	1.26	1.14	1.04	0.80	1.01	0.94	0.83	0.87
Statsforecast.SeasonalWindowAverage.7D.W4	1.33	1.29	1.57	0.94	1.22	2.38	1.72	2.50	0.99	0.94	2.14	2.34	2.26	1.25	0.86	0.90	0.86	0.99	1.17	1.06	1.01	0.99	0.88	0.98	0.89
Darts.LinearRegression.Recursive.1D	2.74	2.59	2.39	2.71	2.73	1.52	1.44	1.01	1.29	1.17	0.90	0.93	0.88	1.17	1.36	1.02	1.34	0.92	0.92	0.85	0.81	0.83	1.08	0.78	0.96
Amazon.Chronos.T5.Tiny	3.29	3.12	2.82	1.25	1.20	1.08	1.24	1.13	1.75	1.62	1.21	1.13	1.26	1.09	0.97	1.37	0.94	1.27	1.11	1.08	0.82	1.03	0.95	0.82	0.91
Darts.LinearRegression.DirectMultiModel.1D	2.44	2.27	2.24	2.41	2.82	2.74	1.48	1.12	1.13	1.07	0.88	0.96	0.86	1.66	1.28	0.95	1.29	0.91	0.89	0.85	0.83	0.83	1.00	0.79	0.90
Darts.LightGBM.DirectMultiOutput.1D	2.66	2.71	2.67	3.68	3.04	1.30	1.50	1.17	1.54	1.52	1.07	1.04	1.00	1.21	1.42	1.22	1.42	1.00	0.98	0.89	0.93	0.88	1.14	0.82	1.04

Figure 12: An overview of EnFoBench results.

Accuracy

The three accuracy metrics (RMSE, MAE, MBE) averaged across all buildings for the best performing models are presented in Table 1. Based on these aggregated metrics, the recursive LightGBM model, which considers one week of history (7D) appears to be the best overall, with an RMSE score of 3.67. However, when examining performance at the individual building level, this model achieves the lowest RMSE for only 10 out of the 25 buildings, as shown in Table 2. In contrast, the direct linear model, which also considers the past week of historical data, outperforms it by achieving the lowest RMSE for 12 buildings. When considering MAE scores, the leaderboard changes in order again, with six different models achieving best performance on various buildings.

Table 1: Accuracy metrics of the best models, ranked by average RMSE across buildings.

Model	RMSE	MAE	MBE
Darts.LightGBM.Recursive.7D	3.67	2.85	0.44
Darts.LinearRegression.DirectMultiModel.7D	3.74	2.92	0.10
Darts.LightGBM.DirectMultiOutput.7D	3.93	3.05	-0.06
Darts.LinearRegression.Recursive.7D	4.00	3.15	0.12
Darts.LinearRegression.DirectMultiOutput.7D	4.31	3.42	0.14
Amazon.Chronos.T5.Base	5.02	3.83	-0.08
Darts.LightGBM.Recursive.1D	5.02	3.81	0.36
Statsforecast.SeasonalNaive.7D	5.04	3.93	0.06
Statsforecast.SeasonalWindowAverage.7D.W4	5.08	4.11	0.16
Amazon.Chronos.T5.Small	5.10	3.92	0.39
Darts.LinearRegression.DirectMultiModel.1D	5.28	4.16	0.62
Amazon.Chronos.T5.Mini	5.33	4.10	0.52
Darts.LinearRegression.Recursive.1D	5.37	4.22	0.54
Amazon.Chronos.T5.Tiny	5.67	4.36	-0.05
Darts.LightGBM.DirectMultiOutput.1D	6.09	4.81	-0.14

Notably, we find the performance of large pre-trained models (e.g. Amazon Chronos and Salesforce Moirai, which have generated a lot of excitement in the broader forecasting world) to be rather underwhelming, with Amazon's second-largest Chronos model achieving the lowest MAE on only one building. Given that these models typically boast millions of parameters and have often been trained on electricity demand profiles already (e.g. the London dataset forms part of the Monash repository, which is often used as the source dataset in large time series forecasting models), this is largely disappointing.

Table 2: Leaderboard by accuracy metrics across buildings (counting the number of buildings, for which the model performed the best).

Model	RMSE	MAE
Darts.LinearRegression.DirectMultiModel.7D	12	8
Darts.LightGBM.Recursive.7D	10	11
Darts.LightGBM.DirectMultiOutput.7D	2	2
Darts.LinearRegression.Recursive.7D	1	1
Darts.LinearRegression.Recursive.1D	0	2
Amazon.Chronos.T5.Base	0	1
Amazon.Chronos.T5.Small	0	0

Computational resources

Figure 13 illustrates the computational resources required for generating a year's worth of predictions per building, alongside their associated RMSE scores. The x-axis employs a logarithmic scale to accommodate the wide range of CPU usage times observed across different models. As expected, Seasonal Naive and other baseline models are the least resource-intensive, typically requiring less than a minute to complete their predictions. However, these models exhibit the highest variability in RMSE scores. Linear Regression models follow in terms of resource usage, requiring at least an order of magnitude more computational power than baseline models but generally offering much improved RMSE scores. LightGBM models are positioned next, demanding even greater computational resources while continuing the trend of enhanced accuracy. The most resource-intensive models are foundational models including Amazon Chronos and Salesforce Moirai (large language foundation models). These models require substantial computational resources (deep neural networks with millions of parameters that have been pre-trained on large corpora of time series datasets) but do not demonstrate a significant decrease in RMSE scores. Salesforce Moirai stands out as an outlier exhibiting a significant number of buildings with high RMSE scores.

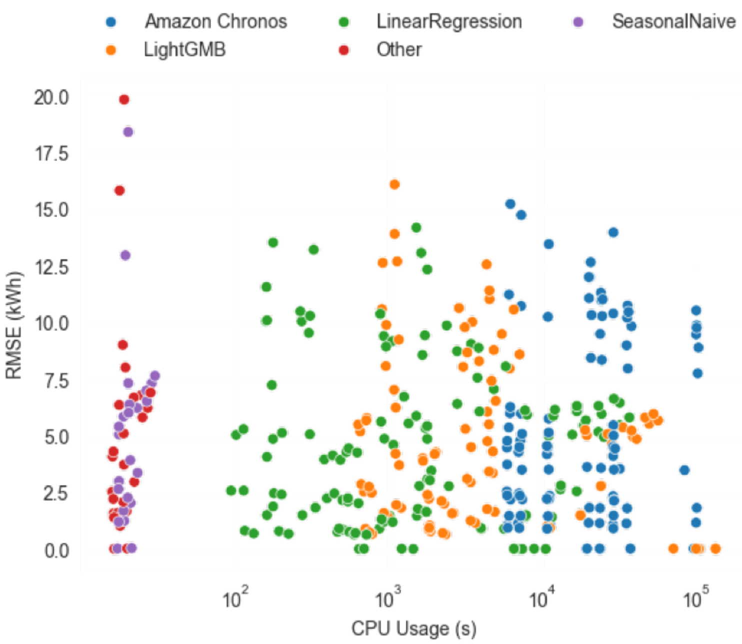


Figure 13: Forecast accuracy versus resource usage shows a general improvement in forecast accuracy with more complex models, even though notable exceptions exist.

3.2.5 Concluding remarks

Our analysis revealed that models such as recursive LightGBM and direct linear regression exhibit strong overall performance. However, their prediction consistency varies significantly across different buildings. By leveraging daily metrics, we gain additional critical insights into model reliability and variability, highlighting the importance of considering both accuracy and stability in model selection. This comprehensive approach enhances our understanding of model capabilities and informs better decision-making for deploying forecasting models in diverse energy use cases. Future work will continue benchmarking other state of the art models from previous competitions as well as diving deeper into identifying which forecast models perform well for which building energy demand time series. This should lead to more general insights for the broader research community. Additionally, we will also aim to couple the forecast models more closely with flexibility-related use cases (i.e. how can forecast-derived baselines be used for M&V use cases, for instance). Likewise, there is significant potential to further extend the benchmark to also consider more sophisticated forecasting use cases (e.g. hierarchical stochastic forecasts for energy demand in buildings and districts). While the datasets in question support this, the evaluation metrics will need to be updated to handle these new challenges.

3.3 Characterizing energy flexibility

Different methods can be used to characterize energy flexibility, ranging from static KPIs related to performance or electricity costs, to more dynamic representations such as the energy flexibility curves (Mugnini et al., 2021), the flexibility graphs (D'hulst et al., 2015) or the flexibility function FF (Junker et al., 2018b). This section focuses on the latter characterization methods, as additional development and testing has been carried out under Annex 82. This method can be used both *deductively* and *inductively*, i.e., based on physical prior knowledge available at the design phase of the building or using data-driven approaches to provide evidence about the flexibility in practice based on historical digital meter data. Moreover, this method can be used at all aggregation levels, including appliances, buildings, districts, and cities. Finally, this method is generalizable, such that the methods can be used for all types of flexible assets of the energy system like wastewater treatment plants, EV-charging, industry, etc. The FF represents a dynamic relationship that can serve as the foundation for developing rules or controllers to provide various grid services, like voltage control, frequency stability, and congestion management (De Zotti et al., 2020). The concept is also useful for the aggregation of flexibility, and hence for the trading of flexibility on conventional energy and power markets (Madsen et al., 2024).

A FF is simply a function describing the relationship between incentives (like prices) and responses (like the electricity load). The function can be estimated based on time series of input data (like prices) and output data (like the load). The function describes the dynamics of the relationship, and hence the function is optimal for the design of optimal control schemes for controlling the response, say the electricity load (Madsen et al., 2025). Depending on the characteristics of the flexible assets (e.g. a building) various parametric relationships should be used. For linear systems, the linear FF (section 3.3.1) is suitable, and this function can readily be estimated using classical linear models from time series analysis, like the ARMAX model identification (Madsen, 2007). However, since most flexible assets are nonlinear a nonlinear FF has been developed (section 3.3.2). This model considers a generalized battery thinking involving concepts like "state of charge". More recently, the FF concepts have been extended also to handle time-varying and slowly varying nonlinear behaviors by using adaptive modelling and estimation schemes (section 3.3.3). Two applications of the FF are presented in section 3.3.4 and 3.3.5.

3.3.1 Linear FF

Flexibility in energy systems refers to their ability to adapt efficiently to changing conditions. By leveraging dynamic penalty signals, energy systems can meet the challenges of a decarbonized future while supporting economic growth and energy security. Numerous studies have focused on methods for capturing demand behaviour and modelling energy flexibility. Accurate modelling requires dynamic, detailed representations of energy systems that incorporate technical constraints, occupancy patterns, and external conditions. A linear, time-invariant dynamic model representing the price-demand relationship and characterizing energy flexibility for buildings and districts is presented in (Junker et al., 2018a), where comparisons across buildings with varying flexibility characteristics are discussed. In the following, we introduce this function in more detail.

Assuming that the energy system and its response to the penalty (price) signal can be considered linear and time-invariant, the demand Y_t at time t can be described as:

$$Y_t = \sum_{k=0}^{\infty} h_k \, \lambda_{t-k} + R_t \tag{1}$$

where λ_t is the penalty signal, and R_t is the non-responsive consumption. The function h_k is called the impulse response function that characterizes the dynamic system's time-domain behaviour. Applying the z-transform, the impulse response can be written as

$$H(z) = \sum_{k=0}^{\infty} h_k z^{-k} \tag{2}$$

where z^{-1} is a unit-delay in the time domain. Input-output relationship of the developed flexibility function can be seen in the figure below.

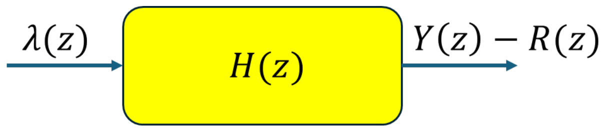


Figure 14: Flexibility dynamics as a system with input and output.

Then, by considering the price signal as step input, one can find the output of the energy system, which is the demand. The following figure demonstrates an energy system demand in response to a step penalty (price) change. As it is seen, the response reveals various information about the demand: i) the delay between the energy price adjustment and the observation of its effect on energy demand, ii) the maximum overshoot and undershoot after the price change, and iii) the duration of overshoot and undershoot after the price change. These characteristics are useful for determining the controller objective and energy source focus that would lead to optimal control implementation.

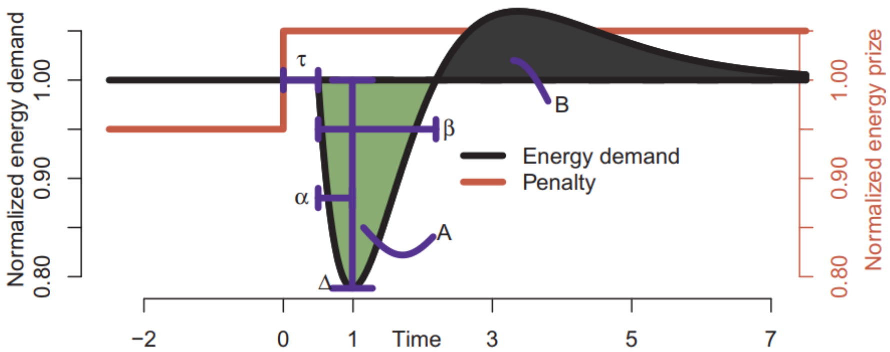


Figure 15: Flexibility characteristics and step price response (Junker, et al., 2018).

3.3.2 Non-linear FF

A nonlinear, and more realistic, model of energy flexibility is developed in (Junker et al., 2020),which represents the price-demand relationship using a time-invariant nonlinear stochastic differential equation. The stability of the dynamics of the flexibility function is investigated in (Tohidi, Ritschel, et al., 2024). This function is utilized in an optimization algorithm to generate optimal price signals for incentive-based control and demand-side energy management (Tohidi, Madsen, et al., 2024). In another study (Tsaousoglou et al., 2024), the flexibility function is integrated into a holistic electricity market including TSOs and DSOs. In the following, we introduce this function in more detail.

The dynamics of the nonlinear flexibility function is developed by considering the flexibility dynamics as a battery. This battery is being charged when the demand is more than the baseline demand, and is being discharged if the demand is less than the baseline demand. Therefore, the flexibility dynamics in state-space form can be written as:

$$\frac{dX}{dt} = (D_t - B_t)/C \tag{3}$$

where X is the state of charge, D is the expected demand (i.e. the predicted demand by the flexibility function), B is the baseline (i.e. heating demand in the absence of dynamic price), and C is the capacity of the battery.

In reality, uncertainties affect these dynamics. Hence, we add stochasticity and write the flexibility dynamics in the form of a stochastic differential equation as:

$$dX_t = (D_t - B_t)/C\ dt + X_t(1 - X_t)\sigma_X dW_t, \tag{4}$$

where W is a Wiener process, and σ_X is the process noise intensity. The factor $X_t(1 - X_t)$ is added to cancel the stochasticity in the boundaries. The expected demand is considered as a nonlinear function of the baseline, the flexibility state, and the price:

$$\delta_t = \ell(f(X_t; \alpha) + g(u_t; \beta); k), \tag{5}$$

$$D_t = B_t + \delta_t \Delta(\mathbf{1}(\delta_t > 0)(1 - B_t) + \mathbf{1}(\delta_t < 0)B_t), \tag{6}$$

where $\mathbf{1}$ is the indicator function, which is equal to 1 when the input is true and equal to 0 when the input is false. Additionally, Δ is the proportion of the flexible demand, $\ell()$ is a logistic function, and $f()$ and $g()$ functions are defined as:

$$g(u; \beta) = \beta_1\, Is_1(u) + \ \dots\ + \beta_7\, Is_7(u), \tag{7}$$

$$\mathrm{g(u; \beta)} = \left(1 - 2\mathrm{X} + \alpha_1(1 - (2\mathrm{X} - 1)^2)\right)(\alpha_2 + \alpha_3(2\mathrm{X} - 1)^2 + \alpha_4(2\mathrm{X} - 1)^6), \tag{8}$$

$$\ell(\mathrm{X}, \mathrm{u}) = -1 + \frac{2}{1 + e^{-k(f(X;\alpha)+g(u;\beta))}}, \tag{9}$$

where Is_1, ..., Is_7 are I-spline functions. The parameters $\beta_1, \dots, \beta_7$, and $\alpha_1, \dots, \alpha_4$ are the parameters and need to be identified. Finally, the output equation of the flexibility dynamics is modelled as

$$Y_t = D_t + \sigma_Y \epsilon_t, \tag{10}$$

where σ_Y is the measurement noise and $\epsilon \sim N(0,1)$.

The nonlinear functions f and g play a crucial role in defining a model that accurately captures the behavior of the energy system. The following figure illustrates examples of typical g and f functions as well as the ℓ function. While the function f projects the effect of the state of charge on the price-demand relationship, g reflects the impact of price. Together, they shape the dynamics of the flexibility function through their interaction within the logistic function, ℓ. For a price, U^*, the state of charge reaches a value, X^*, such that $\ell(X^*, U^*)$ converges to zero.

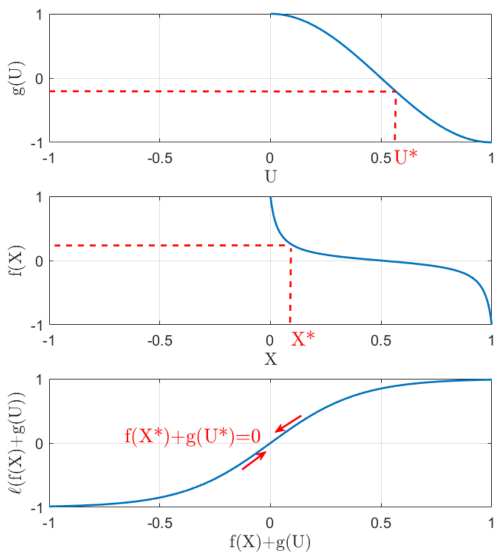


Figure 16: Typical g and f functions, reflecting the effect of the state of charge (X) and price (u), and the function ℓ that links these two functions

3.3.3 Adaptative FF

In a price-responsive energy system, the flexibility function method serves as a key link between price signals and consumer response. This approach quantifies how fluctuations in energy prices impact demand, creating a dynamic mapping of consumption behaviors in relation to pricing changes. At higher levels of smart energy operating systems, such as the aggregator level, the flexibility function plays an essential role in enabling aggregators to communicate effectively with DSOs. This facilitates more accurate demand forecasting and allows for the generation of penalty signals to be relayed to controllers at the physical level, promoting efficient and responsive energy management.

The linear and nonlinear flexibility function (Junker et al., 2018a, 2020) models the price-demand relationship as a time-invariant dynamic system. However, in practice, this relationship is inherently time-varying, influenced by factors such as seasonal changes, consumer behavior (e.g., during holidays), controller mode change, etc. To achieve a more accurate and responsive price-demand mapping, the flexibility function must be adaptable, adjusting dynamically in response to these changing conditions. This adaptability ensures more precise modeling of demand shifts and enhances the effectiveness of demand-side management in a price-responsive energy system.

As previously mentioned, the flexibility function can be utilized at higher levels of energy operating system to generate penalty signals (Tohidi, Madsen, et al., 2024). However, generating these signals typically requires solving an additional optimization algorithm, which can be computationally expensive. To overcome this limitation, it is essential to develop a more efficient approach that bypasses the need for such complex optimization, enabling faster and more scalable performance for demand management in price-responsive energy systems. Moreover, the linear and nonlinear flexibility function (Junker et al., 2018a, 2020) relies on historical data, including price, demand, and baseline consumption, for accurate parameter identification. This process requires both time and a comprehensive data set to ensure reliable results.

The above-mentioned problems motivate the design of a mechanism that accounts for dynamic variations. The adaptive flexibility function (Tohidi, Madsen, et al., 2024) addresses the aforementioned challenges effectively. On the one hand, it is an adaptive method capable of dynamically adjusting parameters based on real-time demand, baseline, and price data. On the other hand, it can generate penalty signals at the physical level without the need for computationally expensive optimization algorithms. Therefore, this time-varying approach enables more accurate and efficient penalty signal generation for demand management. Additionally, it eliminates the need for an offline parameter identification, thus removing the requirement for extensive historical data. Details about the implementation of the adaptive flexibility function are provided below:

The adaptive penalty signal is calculated by using the following formula

$$u_t = \hat{\alpha}_t X_t + \hat{\beta}_t r_t + \hat{\zeta}_t, \tag{11}$$

where the adaptive parameters $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\zeta}$ are calculated by solving the following differential equations

$$\dot{\hat{\alpha}} = \gamma_\alpha \operatorname{Proj}(\hat{\alpha}, Xe), \tag{12}$$
$$\dot{\hat{\beta}} = \gamma_\beta \operatorname{Proj}(\hat{\beta}, re), \tag{13}$$
$$\dot{\hat{\zeta}} = \gamma_\zeta \operatorname{Proj}(\hat{\zeta}, e), \tag{14}$$

where the projection operator Proj (Tohidi et al., 2020; Tohidi & Yildiz, 2022), is employed to keep the parameters of the adaptive system bounded, and X is calculated by using:

$$X_t = X_{t-1} + \int_{t-1}^{t} \frac{1}{C}(D_t - B_t)dt, \tag{15}$$

and C is the capacity of the flexible energy, D is the expected demand, B is the baseline demand and $r = D_{ref} - B$, where D_{ref} is the reference demand. Also, e is calculated as $e = X - Y$, and Y can be obtained from

$$\dot{Y} = \lambda Y + \frac{1}{C} r. \tag{16}$$

The block diagram of the adaptive FF is given in **Figure 17**, and the simulation results demonstrating the efficiency of the developed method are presented in **Figure 18**.

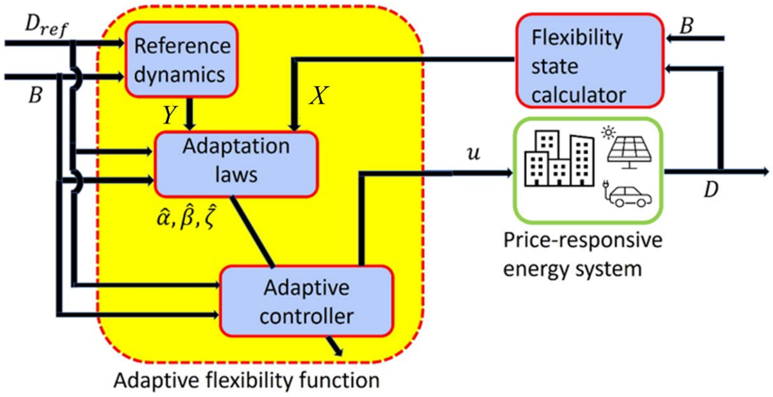


Figure 17: Block diagram of the adaptive FF

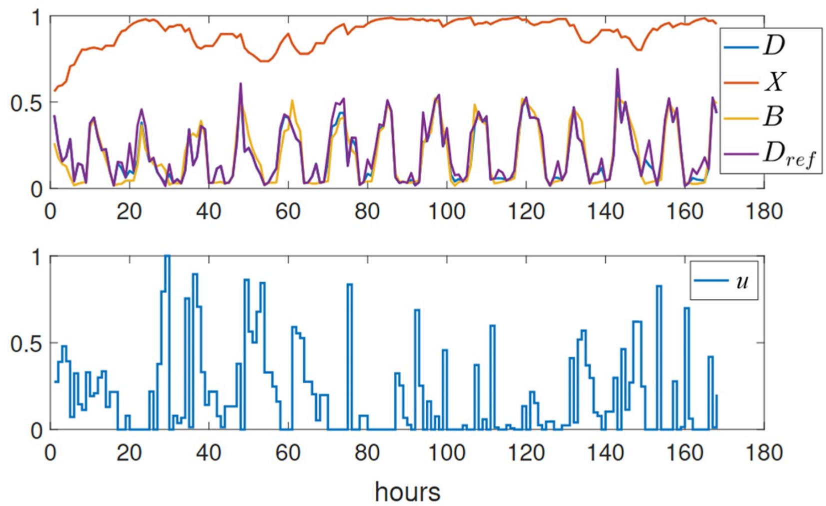


Figure 18: Expected demand D, Reference demand D_{ref}, State-of-charge X, Baseline B (top) and adaptive price generated u (down)

3.3.4 FF: application to heating

This section explores the application of the flexibility function to assess and enhance the energy flexibility of heating systems, and more detailed information is published in (Mokhtari et al., 2025). Using a stochastic nonlinear flexibility function (section 3.3.2), energy flexibility is characterized in response to dynamic pricing, allowing buildings to adjust their heating demand based on price signals. The flexibility function is then used to predict the response of building to price changes as follow:

$$D = FF(u, B, \theta)$$

Where D is the expected demand, u is the price signal, B is the baseline demand, and θ are the parameters of the FF. To find these parameters, the flexibility function should be trained using a long-enough database of baseline demand (B), price (u) and measured demand (Y), which is the actual demand given the dynamic price. The difference between the measured demand and baseline demand shows the response of the system to dynamic price, and the model can therefore use this dataset to learn how the system responds to different prices at different conditions. An iterative approach is used to fit the model to the data and find the flexibility function parameters. Accordingly, heat meter measurements of February 2022 are used as the baseline demand (B) in the training dataset. Since the heat price is currently flat in Denmark, no dynamic heat price data is available. Therefore, the dynamic electricity price from the Nordpool Day-ahead market for DK1 was taken as the dynamic heat price (u) (Nord Pool, 2024).To get the measured demand, a detailed simulation environment is used, where the effect of dynamic price on demand can be measured.

Aimed at designing the right dynamic price that can effectively trigger demand response, the dynamic pricing mechanism is developed through an inverse optimization problem (**Figure 19**). When the optimum price is achieved by the algorithm, it is then sent to the simulation model and the measured demand is calculated. In an ideal case, target demand, expected demand, and measured demand should be the same but there are differences. The difference between the target demand and the expected demand comes from the inverse optimization non-convergence, leading to suboptimal price profiles. The difference between the expected demand and the measured demand, however, comes from the stochastic behavior of the consumers. Using the flexibility function, the goal is to minimize the error between simulated and target demand profiles by adjusting price signals. This process relies on the flexibility function to predict system behavior under different price conditions and optimize heating costs for both consumers and district heating operators. Price profiles (u) generated by the optimization algorithm are evaluated by the trained flexibility function. Flexibility function requires a baseline demand (B) as an input which is the prediction of system behavior without dynamic price. The difference between the expected demand (D) and target demand (P_{target}) are taken as the cost function to be iteratively minimized.

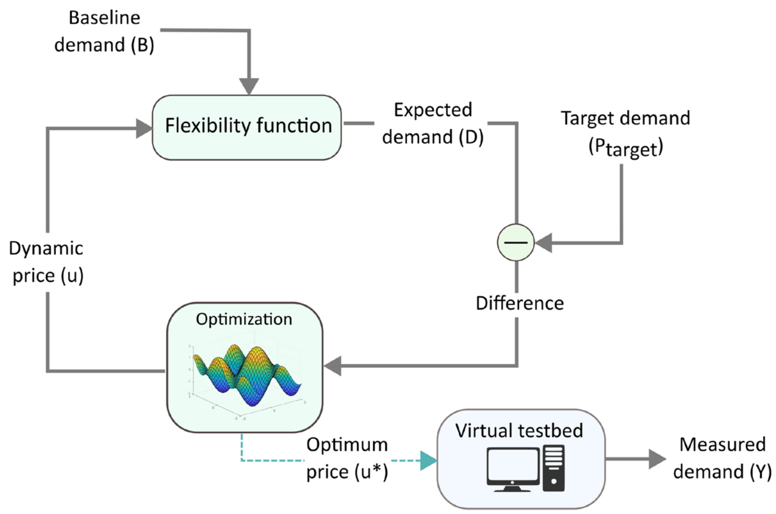


Figure 19: Workflow for finding the optimum price.

A case study involving a neighborhood of 19 apartment buildings was used to test the methodology. The buildings, located in Sønderborg, Denmark, are connected to a district heating network. Each building's heating system is managed through thermostatic radiator valves. Simulations were performed on a white-box model of the neighborhood, developed using Modelica within the Dymola environment, with detailed thermal and system behavior replicated (**Figure 20**). To be able to fit the FF, buildings should be price responsive. Therefore, the indoor temperature controls were replaced by a price-responsive fuzzy logic-based system responding to dynamic heat pricing in the model. The white-box model was validated against real heat consumption data, showing an average Normalized Mean Error of 4.8% to 8.9%.

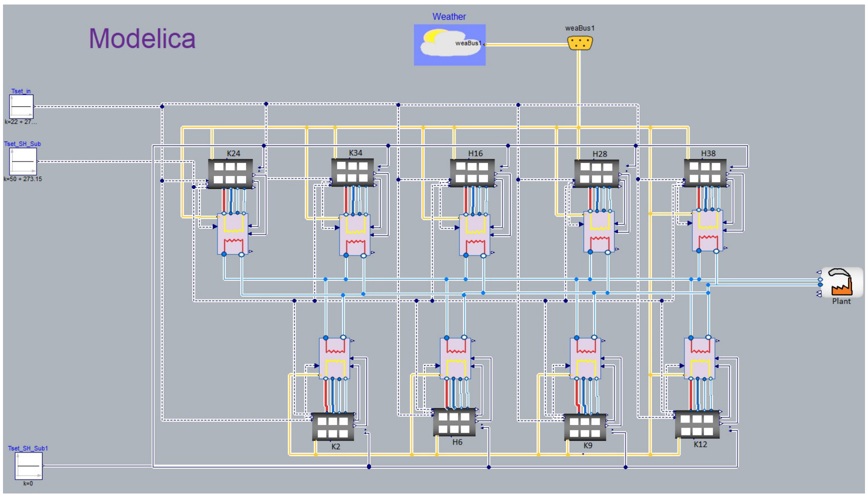


Figure 20: White-box model of the neighborhood created using Dymola.

The results of using the trained FF for designing the optimum price for the neighborhood are shown in Figure 21. The magnitude of the dynamic heat price depends on the flexibility of buildings and the target demand value. If buildings are not flexible enough, the resulting dynamic heat price is expected to have more extreme prices. The fluctuations in measured demand (Y), for example at 15:00, are due to the controllers in the model reacting to sudden changes in thermostat settings. Using the dynamic heat price, the neighborhood demand is matched to the target demand with a normalized mean error of 16.6%. This error reflects the effectiveness of the dynamic pricing approach and can serve as a comparative metric when evaluating alternative methods. However, unlike model validation in other domains, this metric does not have a predefined acceptable range, which limits direct interpretation but still allows for relative performance assessment. By implementing the optimized dynamic pricing, heating costs were reduced by 46.6% compared to a fixed price scenario. The dynamic system also reduced peak heating demand by incentivizing users to shift consumption to off-peak hours.

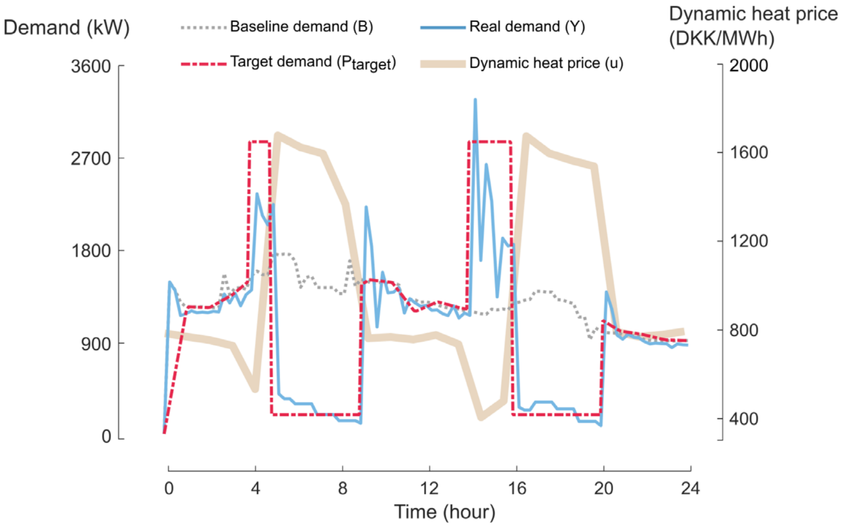


Figure 21: The optimum heat price u for the simulated load of the neighborhood.

The methodology demonstrated that a well-designed dynamic pricing system can activate energy flexibility in district heating networks, resulting in substantial cost savings and demand stabilization. However, challenges remain in real-world implementation, such as variability in consumer behavior and the need for real-time monitoring systems. Further studies are needed to explore the impacts of real-time dynamic pricing adjustments and to integrate renewable energy sources into the flexibility framework.

3.3.5 FF: application to cooling

Several HVAC-based DR strategies, such as global temperature adjustment (GTA), pre-cooling, and HVAC system adjustment techniques, have been used in different studies. Amongst these different DR strategies suitable for HVAC systems in large commercial buildings, GTA is the most widely used technique to modulate the temperature setpoint of air-conditioned spaces to deliver a flexible load.

Although the utilization of the FF to study the response of a heating dominated building has been previously considered, the suitability of this approach for a cooling dominated building has not been investigated before. A joint research team involving researchers from CSIRO (Australia), DTU (Denmark) and University of Colorado Boulder (USA) have carried out a study utilizing the nonlinear FF (section 3.3.2), to understand the response of a building to continually varying price signals. A prototype building model for a representative office building located in Brisbane, Australia, was used in this study. To study the response of the building to input price signals, this study used five-minute interval spot price data of National Electricity Market (NEM) in Australia. It is assumed that the building indoor temperature setpoints change automatically within the comfort band in response to spot prices (rule-based controller). Thus, variable price signal information triggers continuously the building demand via temperature setpoint adjustments. The baseline cooling demand is generated using a flat 21°C temperature setpoint. The observed demand will be higher or lower than the baseline demand depending on the cooling temperature setpoint defined based on the spot price. In order to convert changes in the spot price (u) to a cooling temperature setpoint ($T_{clg-set}$), a transfer function has been utilized – see (Afroz et al., 2024) for details. For this case study, the estimated maximum energy flexibility capacity is 7248 kWh over one month. The maximum and minimum demand was found to be 250 kW and 457 kW respectively.

Figure 22 shows the step-by-step approach used in this study to evaluate the energy flexibility using the FF.

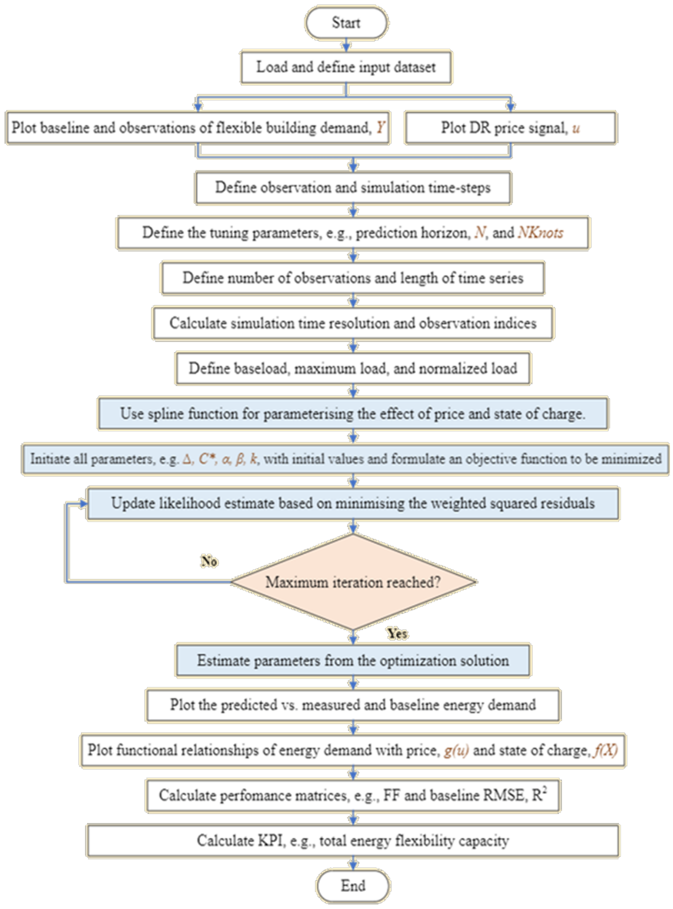


Figure 22: Flexibility Function based modelling and evaluation approach.

Figure 23 shows the predicted demand (D) of the office building obtained from the FF model compared to the observed (Y) and baseline demand (B). The predicted demand refers to estimated demand using FF. This figure shows that the predictions capture the variable demand profile quite well, except for the sharp spikes created mostly due to rapid fluctuation in temperature setpoint. While the FF attempts to model the nonlinear response of the building and its associated HVAC systems with a dynamic model of the building with a single capacitance, there are several time constants associated with zone air and structural building materials, which are lumped into a single value by the FF characterization. Therefore, during periods of rapid price changes, the model cannot capture the flexible behavior very well. Furthermore, the flexible behavior prompted by different operating states of the air handling units may require additional input parameters to capture properly the complexity. Hence, these aspects require further investigation.

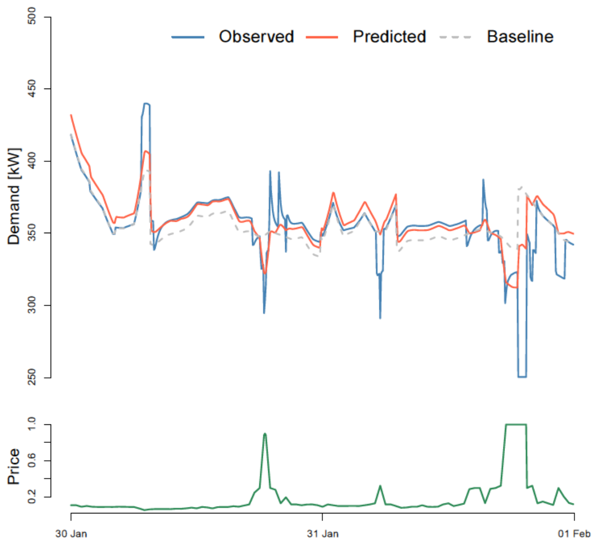
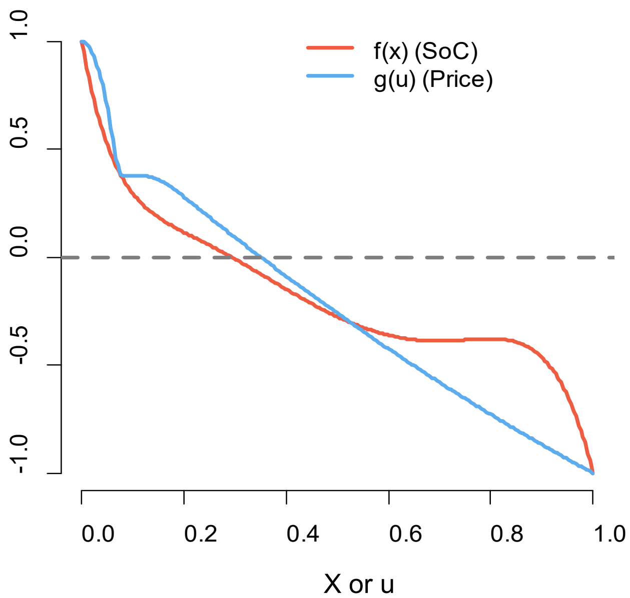


Figure 23: Comparison of FF (predicted) outputs with the observed demand. The corresponding spot price (u) for the studied use case is given below.

The functional relationship between the price (u) and state of charge as well as FF parameters are shown in **Figure 24**. The α_1 value lower than 0 indicates that the *f*-function intersects the x-axis for values of X smaller than 0.5, while a proportional flexible demand (Δ) equal to 1.0 suggests that the flexibility function estimates the entire cooling load to be flexible. X_0 =0.73 indicates that the state of charge was estimated to be at 73% during the start of the simulation. The rest of the parameters have no direct interpretation. The reader is referred to the paper for further details (Afroz et al., 2024).



Δ	1.0
C^* (hr)	35
α_1	-0.50
α_2	0.57
α_3	-0.56
α_4	1.0
k	0.53
X_0	0.73
$\log s_Y$	-3.83
β_1	0.00
β_2	0.31
β_3	1.12e-11
β_4	0.00
β_5	0.00
β_6	0.43
β_7	0.26
λ	500

Figure 24: Functional relationship between relative change in price and state of charge.

3.3.6 Concluding remarks

Benefits of a concept such as the Flexibility Function (FF) are many. They enable linking flexible responsive behavior of assets to external price signals. A Flexibility Function for a system (at asset level or aggregate level) once constructed can be used as a standardized way to communicate flexibility availability to various energy market actors (e.g., aggregator to DSO) and also dispatch of flexibility without detailed physics-based model of assets and aggregates.

Investigations carried out using Flexibility Function (FF) have demonstrated that implicit demand side flexibility can be achieved through continuous modulation of price signals for delivering heating and cooling system flexibility. The function that links the price signal and the system response can be learned (or constructed) based on the asset level or aggregate level response behavior of the system (say heating or cooling) to price signals. This response inherently considers response behavior of consumers. Considering the variability associated with individual consumer behavior, estimation of parameters of FF at aggregate level is expected to be relatively easier. Current studies have used calibrated models to estimate flexibility function parameters. Real world trials involving individual asset level and aggregate level flexible response will further advance the applicability of this promising approach for delivering implicit flexibility at scale.

4. Exploring energy flexibility

In the IEA Annex 82, a set of ten research teams from across the globe joined to address the topic of energy flexibility afforded by building portfolios by responding to a joint challenge, called the Common Exercise. The spirit of this Common Exercise was to enable the various research groups to come together in a joint challenge while still being allowed to use their respective toolchains, simulation environments, and datasets. While the spectrum of flexibility services that a building portfolio can provide is broad, we designed a simple scenario in which the building portfolio is expected to respond to a high price signal during "event periods", i.e., a small number of event hours every day of the month, with a base price prevailing for all other hours. Some teams were more concerned about electric system peaks associated with heating, others with cooling, and yet others with both. The Common Exercise was designed such that teams can choose which scenario or scenarios are of greatest interest to them.

4.1 Common Exercise instructions

The instructions for the Common Exercise provided to each participant team were as follows:

1. **Focus**: The CE targets electricity price signals and buildings with electrically-driven space heating, cooling, and water heating.
2. **Building Selection**: Participants choose a cluster of at least two buildings in a location of interest, with no constraints on building types or characteristics, except for electrically-driven HVAC and water heating systems.
3. **Peak Demand Identification**: Identify peak cooling and heating months for the location.
4. **Simulation**: Simulate each day in the peak months with specified price signals:
    - Peak Period Heating: 6:00-9:00 and 16:00-20:00.
    - Peak Period Cooling: 15:00-18:00.
    - Base price: $10/MWh, Peak price: $15/MWh (Low price ratio).
5. **Results Generation**:
    - **Flexibility Results**: Controller responds to price signals, activating load flexibility.
    - **Baseline Results**: Controller does not respond to price signals.
6. **Price Variations**: Repeat simulations with elevated peak prices ($25/MWh and $50/MWh), keeping the base price at $10/MWh.
7. **Comfort Maintenance**: Ensure acceptable comfort levels (e.g., indoor air temperature limits) during simulations.
8. **Additional Technologies**: Consider on-site solar PV and battery storage for self-consumption, not for grid remuneration.
9. **Economic Calculations**: Based on individual building utility meters, not portfolio level.

Importantly, participants must ensure that comfort and service levels are maintained while applying demand flexibility strategies. To aid in the comparability between research teams we suggest the following set of metrics at a minimum. Additional metrics could be chosen by each team. Calculations should include all end uses (e.g., HVAC, appliances, lighting, plug loads) necessary to calculate total site energy consumption and other related metrics.

1. **Cost Savings [%]**: Percentage savings between the baseline scenario (without energy flexibility) and the energy flexible operation (with demand response, precooling, preheating, load shedding, etc.), based on the price signal used.
2. **Change in Site Total Energy Consumption [%]**: Percentage difference in total site energy consumption between the baseline scenario and the energy flexible operation, indicating the energy efficiency of flexibility activation.
3. **Peak Demand [MW]**: Maximum electricity demand in megawatts, measured every 15 minutes. Report peak demand separately for baseline and energy flexible periods, including the time of occurrence.
4. **Peak to Valley Ratio [%]**: Ratio of daily peak demand to daily minimum demand, calculated every 15 minutes. Indicates the amount of generation needed to meet demand.
5. **Load Factor [%]**: Ratio of average daily demand to peak daily demand, calculated every 15 minutes. Indicates how well generation assets are utilized.
6. **System Ramping [MW]**: Total absolute daily change in portfolio electricity demand from one 15-minute interval to another. Indicates the total demand changes over a day.

These metrics help assess the impact of energy flexibility on cost savings, energy consumption, peak demand, load factor, and system ramping.

The ten research teams contributing to the common exercise focused on several key areas:

1. **Approaches to Flexibility**: Both reactive (rule-based setpoint adjustments) and proactive (model and forecast-based) approaches were used to unlock flexibility in building portfolios. Results showed how demand flexibility can reduce electricity demand and energy use during peak periods, leading to lower energy bills.
2. **Negative Side-Effects**: Potential negative effects of demand flexibility were identified, including rebound and prebound effects, increased total energy consumption, and impacts on occupant comfort. Mitigation strategies were explored to address these issues.
3. **Energy Consumption**: While flexibility can reduce peak period energy use, it may increase demand and energy consumption during non-peak periods, leading to a net increase in overall energy use. This is similar to inefficiencies in electric batteries. However, the net increase in energy use may be justified by cost and emissions reductions. The influence of building and occupant variability was also investigated.

The studies highlighted both the benefits and challenges of implementing demand flexibility in building clusters.

4.2 National Renewable Energy Laboratory (United States)

The National Renewable Energy Laboratory used the open-source software URBANopt™ SDK to model eight single-family detached, two-story residential buildings in Buffalo, New York. These buildings were modeled using 2019 Residential IECC templates and modified to use all-electric equipment. Occupancy and behavioral schedules were based on stochastic schedules derived from American Time Use Survey (ATUS) data. The study aimed to demonstrate demand flexibility in conjunction with on-site solar PV production within the Annex 82 Common Exercise framework. URBANopt-REopt workflows were used to analyze and optimize a stationary electric battery with solar PV production and residential building loads. The PV system was sized at 300 kW, and the battery systems were 300 kW and 300 kWh. The URBANopt-REopt workflows calculated the optimal battery dispatching strategy for different price signal scenarios. REopt optimized energy-related lifecycle costs by dispatching the battery to minimize operational costs. Results were generated for scenarios where grid charging of the battery was allowed and not allowed.

Key findings for an example winter day:

- **Flat Rate Scenario**: The battery was discharged to meet morning loads and recharged with excess PV during the day. No grid-to-battery charging occurred due to the flat rate.
- **High Price Scenario with Grid Charging**: The battery was charged from the grid during early morning and late evening, discharged during high-price periods, and recharged with excess solar generation and some grid charging.
- **High Price Scenario without Grid Charging**: Similar to the flat rate scenario, but with optimal grid electricity purchase during solar production to recharge the battery for evening peak periods.

The study highlights how battery demand flexibility can be optimized with solar PV production under different price signal scenarios.

Price Sensitivity:

- Changes in grid electricity prices impact the optimal energy flexibility strategy, causing the battery to be dispatched during peak periods.
- Even with a flat grid price, if grid-purchased electricity is more expensive than on-site PV-produced electricity, demand flexibility can still be valuable.
- The concept of demand flexibility with a battery can be extended to other technologies like HVAC setpoint control.

Unintended Impacts of PV and Demand Flexibility:

- A flat grid price signal can cause a sharp ramp in net load when PV production decreases at the end of the day, but demand flexibility can mitigate this.
- Demand flexibility reduces total energy consumption during peak periods but can increase net load after peak periods, especially if grid-charging of batteries is allowed.
- REopt optimizes energy-related costs without considering load peakiness, but controls could be used to balance recharging impacts and readiness for future flexibility events.

These concepts, demonstrated with a battery, can be applied to other demand flexibility technologies.

4.3 Polytechnic University of Marche (Italy)

Researchers from the Polytechnic University of Marche studied a cluster of ten residential buildings in Milan, built before 1970 and recently refurbished. The buildings were split between those with high thermal inertia heating systems (underfloor heating) and low thermal inertia systems (fan coils or low-temperature radiators). Each building used a modulating air-to-water heat pump and the following two scenarios were considered:

- **Baseline**: Indoor air temperature maintained at constant comfort setpoints.
- **Flexible**: Economic model predictive control (eMPC) adjusted indoor air temperature between a minimum setpoint and 22°C to minimize costs.

Key Findings:

- **Electricity Consumption**: Power usage decreased during high-price periods and increased during low-price periods.
- **Sensitivity to Price Signals**: Largest reductions in electric demand occurred when moving from low to medium price incentives, with smaller changes from medium to high prices.
- **Daily Variation**: Demand reduction was less significant on colder days with higher heating loads.
- **Cost Savings**: Significant savings were achieved, ranging from 10% (low price) to over 40% (high price), but energy use increased.
- **Peak Demand**: Higher in flexible scenarios compared to baseline.
- **Load Factors and System Ramping**: Load factors generally decreased with higher price signals, and system ramping showed less daily variation with high price signals.

The study highlights the impact of price signals on energy flexibility and the potential for cost savings and demand management in residential buildings.

4.4 Polytechnique Montreal (Canada)

Polytechnique Montréal studied a portfolio of 2400 all-electric residential buildings in Montreal, focusing on energy flexibility strategies during January. The buildings were heated by baseboard heaters and the study tested four rule-based flexibility strategies using the TRNSYS simulation program:

1. **Reactive**: HVAC setpoints reduced by 1°C during events.
2. **Predictive**: HVAC setpoints increased by 3°C three hours before events (except in sleeping zones), then reduced by 1°C during events.
3. **Reactive + DHW**: Same as reactive, plus the lower element of the water heater turned off during events.
4. **Predictive + DHW**: Same as predictive, plus water heater elements increased by 10°C for 1 hour before events, then the lower element turned off during events.

Key Findings:

- **Peak Reductions**: Strategies achieved 15-25% peak reductions during events, corresponding to 2.4-4.1 kW reduced on average.
- **Rebound Peaks**: Higher peaks were observed outside of events due to homes reverting to usual setpoints or preheating before events.
- **Energy Use**: Reactive strategies did not increase daily energy use compared to the baseline, while predictive strategies did due to preheating.

Participation Rates:

- Reducing participation rates (from 100% to 50% and 25%) mitigated rebound peaks while still reducing peaks during events.

Variability in Building Characteristics:

- Variability included envelope performance, HVAC, DHW, and non-HVAC electricity use schedules.
- Predictive strategies achieved higher energy reductions during events but increased overall energy use, while reactive methods achieved modest reductions without increasing overall energy use.
- Setpoint profiles and building envelopes influenced energy reductions, with high-performance homes offering the most flexibility.

The study highlights the impact of different flexibility strategies on energy use and peak demand in residential buildings.

4.5 Czech Technical University (Czech Republic)

The Czech research team analyzed demand flexibility for residential and commercial buildings using detailed HVAC system models. They conducted TRNSYS simulations for August and February using Prague climate data, scaling results to a district level. Demand response was activated at the HVAC system level, adjusting setpoints for air-handling fan speed and outlet water temperature.

Key Points:

- **HVAC Activation**: Less disruptive to indoor environments, but zone temperatures must be monitored.
- **Pricing Schemes**: Two levels of motivation based on rate structure incentives, with varying setpoints for low and high motivation.
- **Mitigation Strategies**: Three scenarios were tested:
  1. No mitigation.
  2. Ramping setpoints gradually after deactivation.
  3. Gradual release of activation signal over 15-minute intervals.

Results:

- **Winter and Summer Demand Response**: Rebound effects were most noticeable in winter without mitigation. Mitigation strategies reduced sharp increases in demand after flexibility periods.
- **Rebound Effects**: Demand response always accompanied by rebound effects, potentially exceeding baseline by up to 150%.
- **Indoor Conditions**: HVAC setpoint adjustments led to small reductions in winter indoor temperatures (0.5-1.0°C) and significant drops in DHW temperatures (as low as 30°C), affecting occupant comfort.

The study highlights the importance of mitigation strategies to manage rebound effects and maintain indoor comfort during demand flexibility events.

Cost savings from HVAC setpoint variations for different price and motivation levels during February (winter) and August (summer) were produced. Under Scenario 3 (gradual release of setpoints), the savings are:

- **Residential Buildings:** 17.8% savings at $15/MWh peak price, up to 34.9% at $50/MWh.
- **Blended Building Portfolio:** 14.6% savings at $15/MWh peak price, up to 47.2% at $50/MWh.
- **Commercial Buildings:** 17.5% savings at $15/MWh peak price, up to 57.3% at $50/MWh.

Savings were significantly lower during August due to the moderate climate.

Table 3 depicts the KPIs for the blended building portfolio, providing better insights into the demand response to different pricing signals and the mitigation strategy, as an example of the common exercise methodology.

Table 3: KPIs for the blended building portfolio of residential and commercial building for different prices and different mitigation strategies (ramp, mitigation with gradual release).

Residences/Administration district								
Base-Peak Price	Results		Scenario 1 - Base		Scenario 2 - Ramp		Scenario 3 - Mitigation	
		No Motivation	Low Motivation	High Motivation	Low Motivation	High Motivation	Low Motivation	High Motivation
10-15'	Total Electricity Cost [EUR]	3447	2971	2846	2924	2783	2943	2808
	Cost Savings [%]		-13.83%	-17.44%	-15.18%	-19.26%	-14.62%	-18.55%
10-25'	Total Electricity Cost [EUR]	4419	3324	3078	3278	3014	3298	3039
	Cost Savings [%]		-24.79%	-30.35%	-25.83%	-31.79%	-25.38%	-31.24%
10-50'	Total Electricity Cost [EUR]	6849	4206	3657	4162	3591	4184	3616
	Cost Savings [%]		-38.58%	-46.60%	-39.22%	-47.56%	-38.91%	-47.19%
	Electricity Consumption [MWh]	296	279	273	275	267	277	269
	Difference [%]		-5.65%	-7.80%	-7.24%	-9.91%	-6.59%	-9.09%
	Peak Demand [KW]	1597	1735	1731	1641	1653	1514	1568
	Ave. P-V Ratio [-]	9.76	47.83	52.32	41.67	48.29	34.72	40.28
	Ave. L-F Ratio [-]	0.49	0.30	0.29	0.34	0.32	0.41	0.37

4.6 University of Colorado (United States) and CSIRO (Australia)

The team from Colorado and Australia studied a portfolio of 2146 homes in Houston, Texas, using day-ahead economic model predictive control (eMPC) to manage air-conditioning loads. They compared this to baseline thermostatic control.

Key Findings:

- **Feeder Demand**: Price incentives led to proportional feeder demand impacts, with demand reductions lasting 1-2 hours before depletion. Significant prebound (precooling) and rebound effects were observed.
- **Demand Reductions**: Peak demand reductions ranged from 10% (low incentive) to 25% (high incentive), with cost savings of 7% to 21%.
- **Overall Demand**: Increased overall feeder demand and system ramping, highlighting the need for carefully designed eMPC controllers.
- **Mitigation Strategy**: Optimal load shaping was developed to create smoother demand profiles, reducing prebound and rebound peaks.

Conclusion:

- eMPC is effective for short-term demand variations but limited by thermal storage capacity.
- Transition to modulating heat pumps and AC equipment will improve demand flexibility.
- Optimal load shaping can enhance portfolio energy flexibility by reducing peak and rebound effects.

4.7 University of La Rochelle (France)

The University of La Rochelle team analyzed energy flexibility from space heating systems in a residential district of 10 buildings (300 dwellings) in La Rochelle, France, during January 2017. They used variable speed air-to-water heat pumps controlled by rule-based reactive control to adjust heating setpoints during event periods, with different incentives:

- **Low Incentive**: -1°C setpoint decrease with opt-out option.
- **Medium Incentive**: -2°C setpoint decrease with opt-out option.
- **High Incentive**: -2°C setpoint decrease without opt-out option.

Key Findings:

- **Occupant Behavior**: Modeled using an agent-based approach, significantly influencing flexibility potential and reliability.
- **Demand Reduction**: Achieved 6% savings for low incentive and 31% for high incentive. Flexibility limited cost increases from dynamic tariffs.
- **Occupant Influence**: Allowing occupants to override setpoints reduced average shifted power but also reduced rebound effects.
- **Building Envelope**: Older buildings with poor insulation showed stronger occupant influence on temperature changes.

Mitigation Strategies:

- **Technical**: Setpoint ramps and randomized end-times had limited impact on reducing rebound.
- **Occupant-Based**: Allowing opt-out was more effective in mitigating rebounds but raised questions about occupant acceptability.

The study highlights the importance of considering occupant behavior and building characteristics in evaluating energy flexibility potential.

4.8 Concordia University (Canada)

The research team from Concordia University modeled space heating systems for three households in Trois-Rivières, Quebec, using a thermal network representation. The homes, despite differences in layout and size, had similar annual energy consumption. They used an economic model predictive control (eMPC) for day-ahead optimization, considering temperature variations as a soft constraint to ensure thermal comfort and minimize costs.

Key Findings:

- **Temperature Variation**: Consistent deviations from setpoints were observed, with longer preheating and recovery stages under medium and high price structures.
- **System Ramping**: Increased system ramping was noted, with an average rise of 120% to 300% compared to the reference pricing structure.
- **Demand Flexibility**: Implementing demand flexibility led to cost savings but increased variability at the feeder level, posing risks to grid stability.
- **Power Demand**: Higher demand levels occurred more frequently with increased price incentives, forming new peaks up to 21.4 kW.
- **Peak-to-Valley Ratio**: Significant demand reductions during event periods increased the peak-to-valley ratio, suggesting that diversified pricing could reduce peak coincidences.

Energy Flexibility Potential:

- Characterized by the Building Energy Flexibility Index (BEFI) and Combined Building Energy Flexibility Index (CBEFI).
- CBEFI calculations showed that pricing incentives affected flexibility potential, with distinct preheating, event, and recovery stages.

The study highlights the impact of pricing incentives on demand flexibility and the importance of managing variability to maintain grid stability.

4.9 Wuppertal University (Germany)

The team from Wuppertal University studied a cluster of eight energy-efficient residential buildings in Wuppertal, Germany, part of the Solar Decathlon Europe 2021/22 competition. Each building has its own air-source heat pump and a 2.5 kWp PV system. The study focused on optimizing the setpoint temperature of the space heating circuit in response to price signals.

Scenarios:

- **BAU (Business as Usual)**: Setpoint temperature adjusted based on outdoor temperature.
- **No Mitigation**: Setpoint temperature reduced by 3°C during high-price periods.
- **Preheating Mitigation**: Supply temperature increased before high-price periods to reduce rebound effects.
- **Ramping Preheating**: Gradual increase in supply temperature before high-price periods.

Findings:

- **Low Price Case**: Preheating strategies reduced rebound effects, with supply temperature increased by 2°C or 1°C + 1°C before high-price periods.
- **Intermediate Price Case**: More effective rebound reduction with supply temperature increased by 3°C or 1.5°C + 1.5°C.
- **High Price Case**: Most effective cost reduction, with supply temperature increased by 4°C or 2°C + 2°C.

Overall Results:

- Preheating strategies led to higher peak demands but reduced grid power demand during high-price periods.
- Rebound effects were noticeable after high-price periods.
- Total energy consumption was higher than BAU due to preheating and rebound effects.
- Cost savings increased with higher price incentives, from 0.5% to 9%.
- Load factors decreased, and system ramping increased across all scenarios.

The study highlights the effectiveness of preheating strategies in managing energy costs and demand flexibility.

4.10 Syracuse University (United States)

The team at Syracuse University conducted economic model predictive control (eMPC) simulations on 580 residential buildings in Arizona during summer. The study focused on shifting HVAC loads during peak hours to achieve cost savings, which varied with price signals: 5.2% for low, 8.5% for medium, and 12.5% for high price signals. Load shifting remained consistent due to Arizona's hot climate, limited building thermal mass, and temperature setpoint constraints.

Key Findings:

- **Load Shifting**: Primarily occurred within the first hour of peak periods due to high ambient temperatures.
- **Occupant Variability**: Derived from measured load data using Bayesian change point detection, showing significant individual variability but smoother aggregated results.

The study highlights the potential for cost savings through eMPC-based HVAC control, despite challenges posed by the hot climate and adverse building characteristics.

4.11 Dubai Electric and Water Authority (United Arab Emirates)

The team from the United Arab Emirates studied summer energy flexibility strategies for a commercial building and a residential villa in Dubai. The commercial building is a LEED platinum office with high thermal mass and a chilled water cooling system, while the villa uses split coil systems for cooling.

Flexibility Strategies:

1. **Step Setpoint Reduction**: Lowering the setpoint from 22°C to 20°C from 12:00 to 15:00.
2. **Ramp Setpoint Reduction**: Gradually lowering the setpoint from 22°C to 20°C from 12:00 to 15:00.
3. **Step Down and Up**: Lowering the setpoint from 22°C to 20°C from 12:00 to 15:00, then increasing it to 24°C from 15:00 to 18:00.

Findings:

- **Load Factor**: Lower on days with large temperature swings, indicating more system ramping.
- **Cost Savings**: Increased with higher price signals, from 0.2% (low) to 19.5% (high).
- **System Ramping**: Highest with strategy 3, lowest with the ramp trajectory setpoint.

Conclusion:

- Pre-cooling strategies effectively shift load outside peak periods, reducing costs but increasing system ramping.
- Economic benefits depend on the magnitude of price signals, with higher savings for higher price signals.

4.12 Discussion

The results obtained by the different teams involved in the common exercise, as discussed in detail above, show interesting insights about the implementation of peak shaving strategies. The simulated cases range in a wide variety of number of buildings involved and climate conditions, the type of control implemented, and participation level of end users. Table 3 summarizes the cases from the ten teams.

Table 3: Summary of results from the Common Exercise.

Contributor	Type of buildings	Type of control	Location	Main findings
NREL	8 single family detached two story residential buildings, including BES + PV	Optimal cost dispatching of the battery system	Buffalo, NY	A change in grid electricity price causes the battery to be dispatched during peak time periods. If the energy flexibility provider is a BES, the careful control of the recharging phase is a key issue
Polytechnic University of Marche	10 single-family homes with underfloor heating or fan coils or low temperature radiators	eMPC (economic MPC)	Milan, IT	The largest reductions of electric demand are observed when going from low to medium incentives, while the change from medium to high prices during the event periods elicits little additional effect. Relevant impact of outside temperature: As the heating load increases, lower percentage demand reductions are achieved.
Polytechnique Montréal	2400 all-electric residential detached buildings	RBC	Montreal, QC	By reducing the participation in the DR program, the rebound effect can be limited. Clustering strategies are important. Variability in occupant behavior and thermal performance level has a large impact on flexibility achievable in each house.
CTU	2 building archetypes, a residential and a commercial building, including the detailed representation of the HVAC system	RBC	Prague	Winter case shows more flexibility than summer case. No impact on occupant comfort, even if the strategy is applied on temperature setpoints
University of Colorado & CSIRO	A residential building portfolio of 2146 homes with central cooling system	eMPC	Houston	a) The increase in price incentives does not lead to highly varied, i.e., proportionally increasing demand reductions, while b) participation levels lead to proportional feeder level demand impacts.
University of La Rochelle	A residential district with 10 buildings and roughly 300 dwellings	RBC	La Rochelle	When allowing occupants to override setpoints, a change in the shifted power can be observed.
Concordia University	3 households	day ahead eMPC	Trois-Rivières, Quebec	The implementation of time-of-use tariffs in thermal load management can lead to prebound and rebound effects in the investigated building portfolio that may pose a

				risk to grid stability at the substation and feeder levels
Wuppertal University	8 buildings	RBC	Wupper-tal, Ger-many	How pre-heating is done affects the re-bound effect
Syracuse Uni-versity	580 residential buildings	eMPC	Arizona	Due to consistently high ambient tempera-ture (above 30 °C during the daytime), space temperatures rise quickly once cool-ing is turned off, resulting in load shifting primarily occurring within the first hour of the peak period under different price sig-nals.
Dubai Electric and Water Au-thority	2 buildings including a commercial build-ing and a residential villa building	RBC	Dubai	Rebound effects can be reduced by means of a proper modulation of the indoor set-points, adjusted for flexible energy use

Regarding the three main points analyzed through the Common Exercise, i.e., sensitivity of demand flexibility to pricing signals, negative rebound effects and increase of overall demand, the ten teams involved reached similar conclusions:

- The electric energy demand from heating and cooling systems is fairly sensitive to control strategies based on pricing signals and its modification increases with the amplitude of price variation compared to the base scenario.
- The demand reduction during peak pricing is always accompanied by a rebound effect, before or after the peak clipping event with a severity dependent on the type of control implemented (RBC, proactive or reactive, MPC) if there are no mitigation strategies in place.
- The overall demand tends to increase in the presence of energy flexibility strategies.

Furthermore, thanks to the specific features of the different case studies, additional considerations can be retrieved from the analysis of the results:

- In the presence of PV and BES used as energy flexibility provider, the recharging phase of the electric batteries can be critical for the grid, especially if the BES recharging is also allowed from the grid. To control the rebound, typically in the evening when the PV production is not available, such recharging needs to be properly scheduled across different buildings.
- Increasing the pricing signal during the peak period leads to higher reduction of the peak demand, however such reduction is not proportional to the price increase. Therefore, the pricing signal needs to be properly set and it is reasonable to consider a maximum limit.
- Demand flexibility is closely linked to the heating and cooling demand during the peak shaving event. In the presence of more severe outdoor conditions, e.g., very cold or hot climate outside, the possibility of reducing the electric heating demand by using the storage capacity of the building is limited if the indoor comfort needs to be maintained.
- To limit the rebound effect caused by energy flexibility strategies, an effective strategy is to diversify the involvement of end users in the event, by for example clustering the demand and applying different pricing signals/control setpoints variations. In the latter case, the possibility for occupants to over-

ride the control action can have a considerable unexpected impact on the load shape. Another approach to reduce the rebound effect, when the energy flexibility comes from thermostatically controlled loads, is to modulate the ramping variation of indoor temperature setpoints, however this approach is case dependent and requires an active involvement of the local control.

- When a large number of buildings is involved in the DR strategy, the rebound effect on the feeder, in terms of ramping of the demand variation, before or after the peak clipping event, can be so severe as to lead to a reduction in grid stability, introducing a relevant drawback effect that calls for mitigation actions.

5. Field tests exploiting energy flexibility

This section synthesizes the field tests based on the results presented in (Yin et al., 2024) and (Cai & Heer, 2024), conducted within the framework of this Annex. In this section, we focus on energy management at the building level. Active building energy management can advance low-carbon buildings and flexible smart city operations through digitalization. All the system configurations and data presented here are based on the NEST building at Empa in Switzerland.
This section is structured into two parts. Section 5.1 addresses a novel signal matrix model predictive control algorithm, which provides stochastic predictions with high-probability constraint satisfaction. The performance has been compared with existing methods through simulations and physical experiments. Section 5.2 explores how a fully equipped occupied building can function as an emission-aware prosumer with flexible energy use responding to the needs of the system. While the first part exploits the operational flexibility of individual components, the second part takes a step further to effectively harmonize the flexibility of all behind-the-meter assets. This section contributes to answering the key research questions including: 1) How can data-driven control and automation support scalable demand response? 2) How do we close the loop and exploit flexibility?

5.1 Testing data-driven control

This first field study concerns a novel signal matrix model predictive control algorithm designed to address the lack of scalable modeling and control procedures in practical implementations. Compared to existing data-driven methods, the algorithm explicitly provides stochastic predictions considering disturbance and measurement errors with few tuning parameters, ensuring reliability through high probability constraint satisfaction. This type of modelling lies in between the model-based and the model-free approaches. The performance is extensively compared with three state-of-the-art algorithms in a space heating case studies using a high-fidelity simulator *nestli*[5] (**Figure 25**). In its core, *nestli* is a calibrated EnergyPlus model of the NEST demonstrator. The results are further validated with physical experiments using the same system on which the simulator is based. To assess transferability, the algorithm is further implemented on diverse controlled systems, including a domestic hot water heating system and a stationary electric battery.

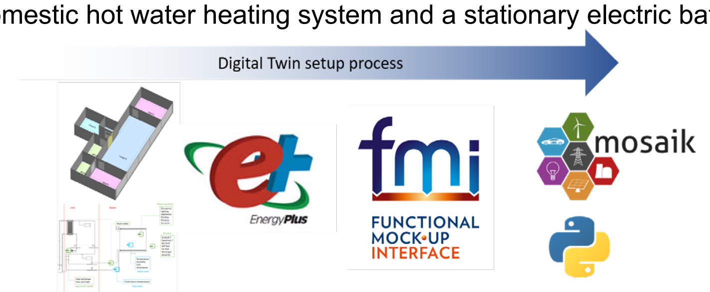


Figure 25: Overview of the setup process of *nestli*, based on the NEST demonstrator at Empa.

5.1.1 State-of-the-art and challenges

The proposed data-driven predictive control algorithm is based on behavioral system theory. Compared with techniques based on machine learning, such methods have excellent data efficiency with minimal training and tuning effort but are not inherently adapted to nonlinearities. The methods along this line include MPC based on subspace identification, indirect methods such as subspace predictive control (SPC), and direct

[5] https://github.com/hues-platform/nestli

methods such as data-enabled predictive control (DeePC). However, the subspace-identification-based approach requires additional model order selection and state observer design procedures. While DeePC provides a simple implementation pipeline, tuning the hyperparameters can be time intensive. DeePC is also not easily interpretable since the input–output mapping is not well-defined. More importantly, reliable frameworks to obtain uncertainty models for robust constraint satisfaction are not available yet, which is crucial for building-related applications.

5.1.2 Brief overview of the method

In conventional MPC algorithms, the output is predicted and controlled by iteratively propagating the state-space model, assuming the knowledge of the model parameters (A, B, C, D, E) is available. In data-driven predictive control, however, no prior knowledge about the parameters is required, but a matrix of input-disturbance-output trajectory data

$$Z = [z_0^d \quad \cdots \quad z_{M-1}^d] \tag{11}$$

has to be collected from historical data or high-fidelity simulation, where each column

$$z_i^d = \mathrm{col}\left(u_{t_i}^d, \dots, u_{t_i+L-1}^d, w_{t_i}^d, \dots, w_{t_i+L-1}^d, y_{t_i}^d, \dots, y_{t_i+L-1}^d\right) \tag{12}$$

is a length-L trajectory of the system. This matrix is referred to as the signal matrix. Using Willems' fundamental lemma in the behavioral system theory, a nonparametric input-output mapping can be directly obtained from the signal matrix Z when no noise is present, i.e., $v_t = 0$. The rest of this section focuses on a concise introduction of the method. With sufficiently persistently exciting inputs, the range space of Z contains all possible trajectories of the system. Let $L = L_0 + L'$, and L_0 be no smaller than the observability index of the system. Define a partition of Z as

$$Z = \mathrm{col}\{U, W, Y_p, Y_f\} = \mathrm{col}\{\Psi, Y_p, Y_f\},$$

where $U \in \mathbb{R}^{n_u L \times M}$, $W \in \mathbb{R}^{n_w L \times M}$, $Y_p \in \mathbb{R}^{n_y L_0 \times M}$, $Y_f \in \mathbb{R}^{n_y L' \times M}$, and $\Psi \in \mathbb{R}^{(n_u+n_w)L \times M}$. The noise-free signal matrix satisfies the rank condition

$$\mathrm{rank}(Z) = \mathrm{rank}\left(\mathrm{col}\{\Psi, Y_p\}\right) = (n_u + n_w)L + n_x.$$

Then, the L'-step-ahead output predictor $\mathbf{y}^t = (y_k)_{k=t}^{t+L'-1}$ at time t can be constructed from the past input sequence $\mathbf{u}_{\mathrm{ini}}^t = (u_k)_{k=t-L_0}^{t-1}$, the future input sequence $\mathbf{u}^t = (u_k)_{k=t}^{t+L'-1}$, the disturbance sequence $\mathbf{w}^t = (w_k)_{k=t-L_0}^{t+L'-1}$, and the past output sequence $\mathbf{y}_{\mathrm{ini}}^t = (y_k)_{k=t-L_0}^{t-1}$ by solving the linear equations

$$\begin{bmatrix} \Psi \\ Y_p \end{bmatrix} g = \mathrm{col}\{\mathbf{u}_{\mathrm{ini}}^t, \mathbf{u}^t, \mathbf{w}^t, \mathbf{y}_{\mathrm{ini}}^t\},$$

where $g \in \mathbb{R}^M$ is an intermediate vector. The output prediction is then given by

$$\mathbf{y}^t = Y_f g.$$

In this work, optimal controllers are designed to minimize given control objectives $J_u(\mathbf{u}^t)$, subject to input and output constraints $\mathbf{u}^t \in \mathcal{U}^t$, $\mathbf{y}^t \in \mathcal{Y}^t$ at time t. Similar to standard MPC, the scheme is applied in a receding horizon fashion.

In practice, however, when noise and uncertainties exist, the algorithm above cannot be directly applied. In this work, two sources of uncertainties are considered, namely: 1) Noise in output measurements and 2) Uncertainties in online disturbance measurements and predictions. The field implementation extends indirect data-driven predictive control to provide stochastic control using techniques similar to those in stochastic MPC. This leads to the following stochastic control problem:

$$
\begin{aligned}
\min_{\mathbf{u}^t} \quad & J_u(\mathbf{u}^t) \\
\text{s.t.} \quad & \mathbf{y}^t \sim \mathcal{F}(\mathbf{u}_{\text{ini}}^t, \mathbf{u}^t, \bar{\mathbf{w}}^t, \bar{\mathbf{y}}_{\text{ini}}^t; Z), \\
& \mathbf{u}^t \in \mathcal{U}^t, \quad \mathbb{P}(h_i^t \mathbf{y}^t \le q_i^t) \ge p, \qquad \forall\, i = 1, \dots, n_c,
\end{aligned}
$$

where $\mathcal{F}(\cdot)$ is the stochastic data-driven predictor. Due to the potentially unbounded noise, the output constraints cannot be satisfied robustly, but only with a high probability p as chance constraints for the ith constraint.

Several standalone tests have been carried out to ensure the robustness of the conclusion from the field implementation, covering multiple aspects encountered in real implementation. The studies, summarized in Figure 26, consider transferability across heterogeneous controlled systems, transferability from simulation to experiment, and the impacts of specific control tasks. Three controlled systems were investigated: a space heating system, a DHW heating system, and a stationary Lithium-ion electric battery, representing typical demand-side resources. We considered typical control tasks, including constrained resource planning and trajectory tracking. In the building sector, constrained resource planning is prevalent and aims to minimize total energy use while ensuring constraint satisfaction. Reference tracking tasks can also be formulated for stationary electric batteries, as they are promising candidates for ancillary service provision, as they track reference signals set by system operators. Please refer to the original paper for a detailed description of each case study.

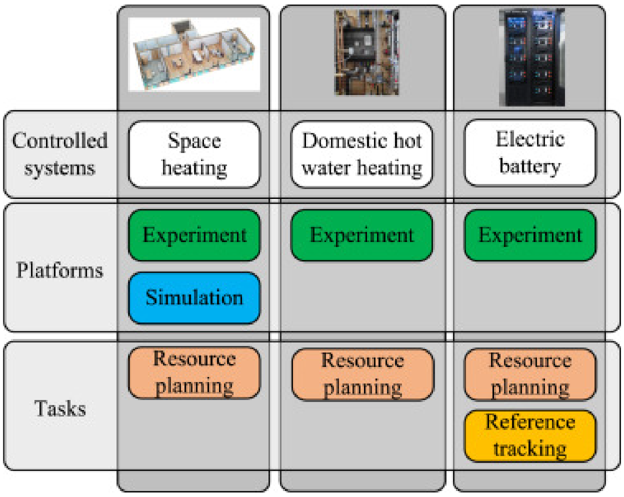


Figure 26: Summary of the investigated case studies.

5.1.3 Key results and discussion

To illustrate the effectiveness of the algorithm, we focus on explaining the results related to space heating, which is estimated to take the largest share of total energy use in cold climates such as in western Europe. Both simulation and experimental studies were conducted to evaluate the proposed algorithm. The simulation study was mainly used to benchmark multiple data-driven control algorithms under the same boundary conditions and the experimental study validated the performance and the transferability in real configurations. Specifically, the system considered in the field implementation is a three-room apartment, shown in Figure 27, hereafter referred to as the UMAR unit. Each room was treated as one thermal zone and heating power was dissipated into the zones through radiant ceiling panels. The continuous power setpoint, determined by the controller, was realized by regulating the valve opening, which was converted into discrete opening and closing sequences using power-width modulation logic. Additionally, the control of space heating was subject to constraints imposed by occupants' perception of comfort. Although other indoor conditions also influence occupants' comfort, temperature sensors are the most commonly available for assessment, so thermal comfort bounds were expressed as temperature limits in this study. As the unit is residential, the thermal comfort bounds were predefined, considering the unit to be unoccupied during the day and occupied during the night. The definition led to relaxed temperature constraints during unoccupied hours. Specifically, the constraints were set to be between 20 °C and 26 °C from 08:00 to 16:59 and between 22 °C and 24 °C from 17:00 to 07:59 of the second day. The comparison with other data-driven algorithms under various level of noises is summarized in **Figure 28**.

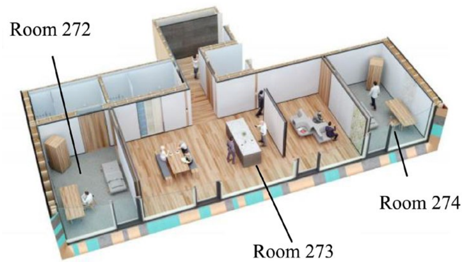
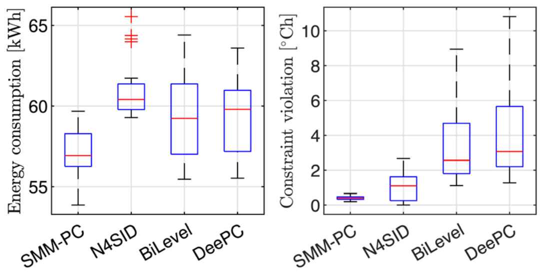
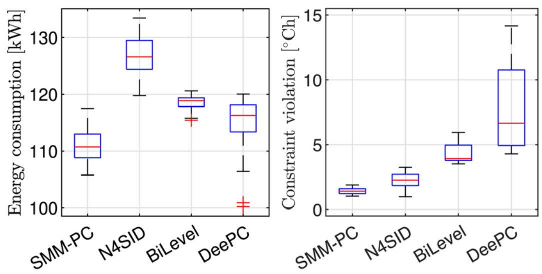


Figure 27: Layout of the UMAR unit in the NEST building with the controlled rooms marked. ©Werner Sobek.

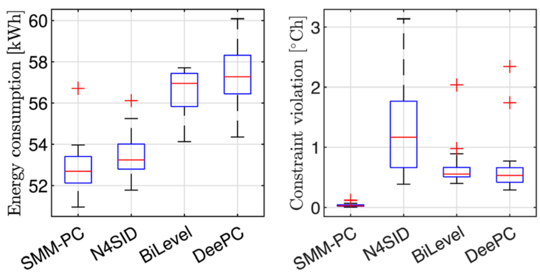
Monte Carlo simulations of 20 runs were conducted for each algorithm with different measurement noise and disturbance uncertainties realizations. All the predictive control algorithms considered performed significantly better than the baseline hysteresis control, with a 20%–27% reduction in energy use and a 30%–93% reduction in constraint violation. So, in what follows, we focus on benchmarking different data-driven predictive control algorithms. Results of energy use and temperature constraint violation for the three controlled rooms are shown in **Figure 28**. It can be seen that the SMM-PC algorithm outperformed the other algorithms in terms of both energy use and constraint satisfaction for all three rooms. Specifically, compared to N4SID (Van Overschee & De Moor, 1994), BiLevel (Lian et al., 2023), and DeePC (Coulson et al., 2019), SMM-PC reduced the constraint violation by 59%, 77%, and 90% with an average energy saving of 8%, 6%, and 4%, respectively.



(a) Room 272



(b) Room 273



(c) Room 274

Figure 28: Boxplots of energy consumption and constraint violation for different data-driven predictive control algorithms. SMM-PC refers to the algorithm in section 5.1.2.

The experiment was carried out between February 7 and February 11 of 2023. During the experiment, the prediction horizon and the confidence level of constraint satisfaction were set to be 15 (3.75 h) and 70%, respectively. They were the same as the values used in the simulation study. The experimental results summarized in **Figure 29** include temperature measurements, control decisions, and realized power inputs into each temperature zone (using the SMM-PC algorithm). Different from the numerical study in the previous section, the actuation of control decisions is subject to errors in practice, leading to mismatches between control decisions and realized power inputs in **Figure 29**. It can be observed that the temperature in each room was maintained within the predefined comfort zone most of the time. The accumulated constraint violations in Rooms 272, 273, and 274 are 0.025 °Ch, 2.159 °Ch, and 1.518 °Ch, respectively. Additionally, in all rooms, preheating decisions can be observed once the step increase of the lower temperature limits was within the horizon of the controller. The results validated the hypothesis that SMM-PC can ensure constraint satisfaction within the technical capability of the underlying physical system.

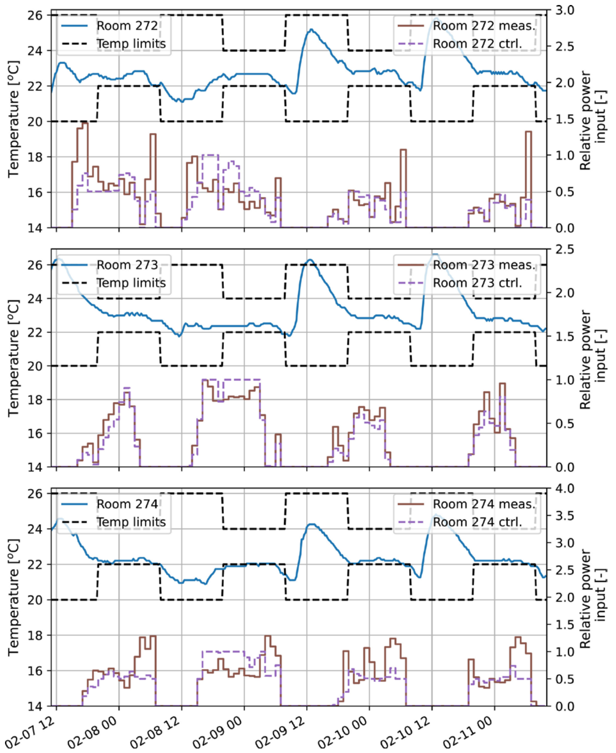


Figure 29: Experimental results of space heating control. Top plot: results for Room 272. Middle plot: results for Room 273. Bottom plot: results for Room 274. The dashed black curves show temperature limits. The solid blue curves show temperature trajectories. The dashed purple curves show controller decisions (scale at left). The solid olive curves show realized thermal power inputs into the rooms in relative terms by normalizing using thermal power capacities (scale at right).

5.1.4 Concluding remarks

Scalable modeling and control frameworks can facilitate demand side management, a critical tool for the reliable and efficient operation of future energy systems. To address current limitations, this work proposes a novel data-driven predictive control algorithm using the signal matrix model. By using a stochastic control framework, this algorithm is designed to be reliable against noise and disturbances. The performance of the methodology is comprehensively evaluated in high-fidelity simulations by comparing it against other data-driven predictive control approaches, and its transferability is validated in multiple experimental studies. The experimental results revealed that the SMM-PC algorithm performed well in terms of constraint satisfaction and its performance can be transferred to different types of systems with minimal tuning effort. Therefore, the SMM-PC algorithm has the potential to enable large-scale deployment of DSM across heterogeneous systems timely. The optimal selection of historical measurements as the input-disturbance-output trajectory data is a nuanced task and beyond the scope of this study. Nonetheless, a potential solution involves combining long datasets including relevant periods with preconditioning of data matrices. Expert knowledge can assist in identifying periods with sufficient system excitation. In the example of room heating, including periods with variable thermal power input and temperature fluctuations is crucial. Lastly, even though the methodology is only rigorously proven for linear systems, the numerical and experimental studies in this work show that it can be effectively applied to nonlinear systems by considering nonlinearity as an additional source of error, with better performance than conventional algorithms based on linear system identification.

Two main conclusions can be drawn from the quantitative results. Firstly, compared to the state-of-the-art data-driven predictive control algorithms, the proposed algorithm can improve occupants' thermal comfort and energy saving by up to 90% and 8%, respectively. Secondly, the algorithm effectively ensures constraint satisfaction, and its performance can be transferred to other systems with minimal tuning effort. Therefore, the proposed algorithm can facilitate large-scale deployment of demand side management to support energy system operation. However, several limitations must be noted. First, while data-driven approaches provide scalability, the interpretability of the decisions still needs to be assessed. Poor interpretability could potentially compromise the security of a system comprising many automated and distributed devices, which are aggregated to support a secure grid operation in the first place. Second, while the transferability across different controlled systems was evaluated, it remains to be assessed for the same type of controlled systems with different characteristics. For example, the proposed algorithm needs to be tested on a large group of heterogeneous buildings with different heating input systems.

5.2 Prosumer with flexibility quantification and provision

This study examined how a real occupied building, with all its energy assets, could function as an emission-aware prosumer with flexible energy use. An existing building energy management system was enhanced by integrating a model predictive control strategy. The setup reduced equivalent carbon emissions from electricity imports and provided flexibility to the energy system. The experimental results indicate an emission reduction of 12.5% compared to a rule-based controller that maximized PV self-consumption. In addition, a minimal flexibility provision experiment was demonstrated with a locally emulated distribution system operator. The results suggest that flexibility was provided without the risk of rebound effects, as flexibility was quantified and communicated to the system operator in advance. This study demonstrates the feasibility of low-carbon buildings and their support for flexible energy systems, while also identifying and discussing practical scalability challenges. Based on our literature review, there remains a gap in comprehensive experimental insights covering three aspects: (1) emission-aware operation, (2) flexibility quantification and provision to DSO, and (3) the impacts on and from occupants. A thorough investigation of these aspects is essential for the successful deployment of flexible buildings at a cluster scale.

5.2.1 Brief overview of the method

Without elaborating on the optimization control problem at the building level, we focus on illustrating the framework incorporating system operator into the close-loop implementation demonstrating a simple use case of energy flexibility combining all behind-the-meter resources. The flexibility envelope identification, as illustrated in **Figure 30**, starts with identifying the energy bounds by energizing flexible appliances to their extremes. The upper energy bound is identified by maximizing device consumption as early as possible, coupled with full PV output curtailment. To identify the lower energy bound, all the loads are set to consume as late (compared to expected service time) and as little as possible. Simultaneously, the stationary and EV batteries are set to discharge as early and as much as possible without curtailment of PV output. The upper and lower energy bounds are illustrated as the red and blue curves in **Figure 30**.

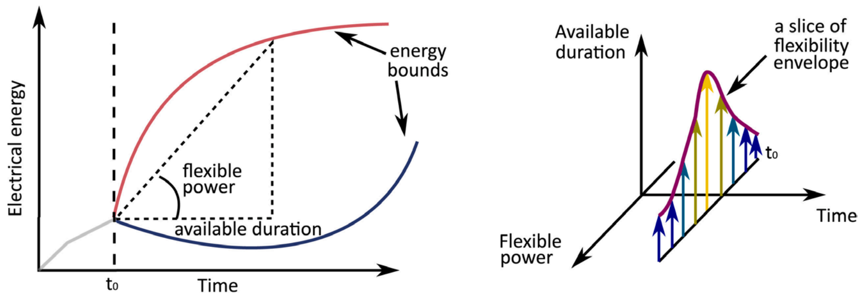


Figure 30: Illustration of the workflow to obtain one slice of a flexibility envelope at t0. The left figure depicts upper and lower energy bounds derived from extreme scenarios. The bounds indicate flexible power and the corresponding available duration. The right figure maps the power levels and duration onto a three-dimensional space and illustrates one slice of flexibility envelope. All slices at each time step within the horizon constitute a full flexibility envelope.

The proposed flexibility envelope captures the thermal inertia of a building, the storage of a domestic hot water tank, the bidirectional charging of a stationary electric battery and/or an EV, and the curtailable PV power. When this envelope is communicated in advance, a DSO obtains a comprehensive overview of the available flexibility at a given building. Upon receiving the flexibility envelope, the DSO sends a flexibility provision request to the building, including the starting time and the ending time of flexibility provision, and the power level that needs to be tracked. At the building level, this is achieved by modifying the cost function of the optimal control problem upon receiving the request from the system operator. Moreover, the overall two-stage framework is illustrated in **Figure 31**. This framework can enable the DSO to address local network issues using local flexibility resources.

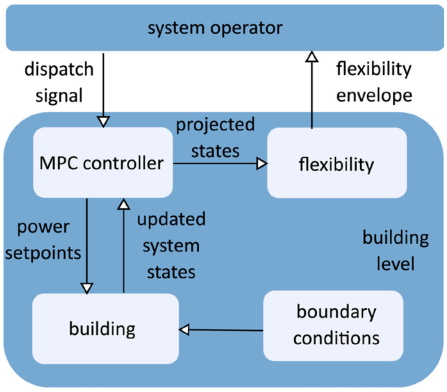


Figure 31: Illustration of the information flow among the building, the controller and the system operator. At each time step, updated building state measurements are retrieved by the controller to solve the emission-aware OCP. The resultant state trajectories are used to quantify the flexibility envelope, which is self-reported to the DSO. Upon receipt, the DSO may request the prosumer to provide flexibility depending on the condition of the network.

5.2.2 Key results and discussion

All assets, except the EV, were real components in the experiment. The EV with bidirectional charging was simulated with an identical model for both the emulator and the controller. This assumes perfect modeling of the EV battery and charging/discharging process. Occupants were living in the apartment during the experiments. Physical components are from the NEST, as Figure 32 shows. The hardware was distributed around the research infrastructure. The time-stamped measurements allowed emulating the actual operation of a prosumer equipped with all the assets. The total power of the emulated building combined all power measurements and the simulated EV power.

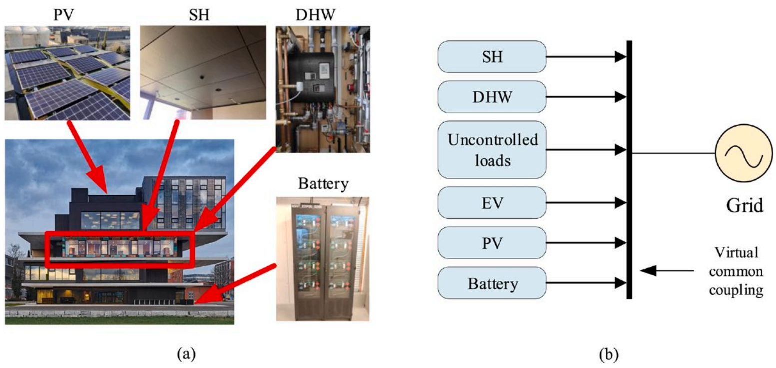


Figure 32: Physical layer of the experimental setup (a) and its single-line representation (b). The physical layer shown in (a) includes SH with ceiling heating panels, DHW with a buffer tank, fixed loads of an apartment unit marked in the red box, a PV installation and a battery. The facility's thermal networks for SH and DHW were treated as fully separate for simplicity, despite being cascaded in reality. The PV installation is placed on the roof and the battery is located in the basement, which are not directly visible in the figure. The bi-directional charging EV is a simulated entity and is not visible here. The spatially distributed hardware are virtually coupled via their timestamped measurements as shown in (b). Such a virtual coupling is seen as a billing point for the DSO.

Results of the flexibility provision with an emulated DSO are shown in Figure 33. We consider a scenario in which the DSO experiences network congestion due to low ambient temperature. More specifically, load peaks may be exacerbated due to simultaneous consumption from newly adopted HPs, which the distribution system may not be planned for. Thus, additional flexibility from buildings is needed to mitigate the issue. As per industry practice, ripple control has been used for decades for direct load control by broadcasting audio frequency signals to cease operation of devices such as HPs within a target group. However, ripple control represents unidirectional communication and addresses limited types of flexible devices. In the field test, we demonstrate the proposed framework with an emulated DSO. More importantly, we show similar performance, namely keeping total power consumption from the grid close to 0 kW, which can be achieved from the perspective of the DSO, while comfort levels and preferences of end users are shown to be respected.

In the experiment, a building self-reports its flexibility envelope to the DSO, who in turn remains idle until flexibility needs are foreseen according to the weather forecast. The DSO examines the reported flexibility envelope and notifies flexibility provision to the building at the time marked by the vertical line in Figure 33. Within the flexibility provision period, the building tracks the setpoint. The results show that the total power consumption from the grid is reduced to a marginal level, although not strictly zero. This can be attributed to the actuation errors. Besides, we can observe that the energy states of all devices are comfortably away from their lower limits. This indicates that there are no immediate needs for electricity imports from the grid. Therefore, there is no risk of rebound effects. While the building activates its flexibility, the DSO continuously monitors the building and remunerates the service provider afterwards.

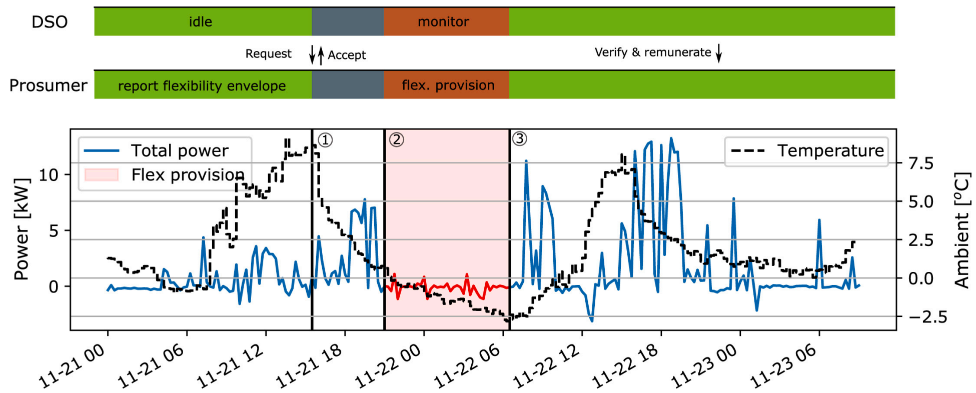


Figure 33: Results of flexibility provision example. The bars on top of the figure denotes the actions between the emulated DSO and the prosumer. The blue and red curves denote the aggregate power of all flexibility resources outside and inside flexibility provision period respectively.

Overall, the presented results demonstrate that the controller operates with emission-aware MPC as base strategy and can deviate from the optimal trajectory to provide flexibility upon request. In the experiments, control-oriented models, which were extracted from historical data, yielded satisfactory results both in terms of emission reduction and maintaining thermal comfort. Apart from the quantitative data, qualitative feedback on thermal comfort was also collected via an online feedback form during the experiments. For DHW, all feedback indicated “very satisfied”. As for the bedrooms, 37.5% of the time, the occupants indicated slightly cool indoor temperature in the 7-scale rating (cold, cool, slightly cool, neutral, slightly warm, warm, hot) matching low comfort violations. At other times, occupants expressed neutral opinions about the indoor temperature. We observed fatigue among users responding to survey requests, limiting the current survey. This suggests that the feedback strategy in the future needs to take a different form, especially for real-time control. Since the bi-directional EV was emulated, no feedback was gathered. Additionally, alternatives to the current definition of comfort zones and user preferences are available, such as continuous monitoring of occupancy coupled with real-time inference and forecasting. These alternatives have distinct cost and scalability implications compared to those discussed in our study.

Several challenges hinder scalability. The varied levels of digitalization across regions necessitate modifications to adapt to controllable resources and achieve a balanced trade-off between costs and benefits. Moreover, reliability might be a concern. This is because reliability decreases as the element count increases in any series system. The proposed prosumer is an example of such series systems. Additionally, ripple control is implemented in an open-loop fashion and the response can be expected within 7 s, whereas the proposed framework would take longer to quantify the flexibility envelope and establish flexibility provision agreements. Moreover, communicating the flexibility envelope requires significant bandwidth, necessitating further simplification. All in all, the existing ripple control scheme excels in simplicity and responsiveness. In contrast, the presented framework is favorable for DSOs that require automation and an optimization-based approach due to the complexity of handling numerous resources. Taking these into account, future research might benefit from exploring a hybrid strategy that achieves a trade-off between performance and robustness. Although such assessment carries substantial policy implications, it remains underexplored in the existing literature.

5.2.3 Key learnings

Despite buildings' promising role in energy transition, challenges arise due to the involvement of diverse operational objectives and stakeholders. This field study addresses gaps in experimental insights, focusing on emission-aware operations, flexibility quantification, and provision to DSO, alongside analyzing the impacts on and from occupants. A week-long experiment demonstrates a 12.5% reduction in equivalent emissions compared to a benchmark controller that focused on maximizing PV self-consumption, while also indicating improvements in end users' thermal comfort levels. The proposed flexibility provision framework, tested with an emulated distribution system operator, showcases a scenario where flexibility is utilized to mitigate network congestion, effectively coordinating behind-the-meter resources, maintaining user comfort and preferences under different conditions without rebound effects.

Nonetheless, limitations include not accounting for uncertainties in forecasts and model errors in the optimization problem. Additionally, challenges may arise if the maximum duration of flexibility provision is requested or when a DSO requires flexibility with a different lead time. It remains to be assessed whether flexibility can be reliably provided, and comfort guaranteed. Future research directions include assessing more scenarios for robust flexibility quantification and exploring probabilistic flexibility representations to better manage uncertainties. A full discussion of the remuneration scheme has not been examined, but this remains an important issue for future research.

5.3 Summary and outlook

Within the two reported field implementations, energy flexibility has been exploited thanks to the inertia of the system, such as thermal inertia and the operating regions of each system, such as the water tank temperature limits. Such flexibility provides the basis for modifying energy use patterns, necessary for energy efficient and secure operation of large-scale energy infrastructure. While these implementations have shown promising signs in the technical development, large scale demonstration covering heterogeneous types of system will further validate the robustness of the conclusions above. As Le Dréau et al. (2023) points out, there is still scarcity of the large-scale demonstration and it remains to be seen how the technological advancement can synergize the practical business development and social acceptance, to achieve massive rollout in practice, necessary for sufficient impacts on the energy infrastructure. Crucially, the implementation depends on availability of sensing, actuation and impacts from energy end users. Performance gaps exist and the authors suggest the implementation results are cautiously interpreted bearing in mind the dependency of performance on regional system regulation, climate conditions, occupant behaviors, energy consumption patterns and availability of supporting digitalization. When it comes to integrating advanced control schemes into critical infrastructure, reliability assessment should not be neglected. The learnings generated in section 5.1 and 5.2 are being tested in a decentralized district heating network of three row houses[6].

[6] https://www.aramis.admin.ch/ZugehoerigeProjekte/?ProjectID=53688&Sprache=en-US

6. Conclusions

The integration of energy conversion systems based on renewable energy resources, with limited dispatchable capabilities, the ongoing decentralization, and the increasing electrification of energy demand are contributing to the growing complexity of power systems' management. The building sector, a significant energy consumer and greenhouse gas emitter, can therefore support power systems' operation by offering energy flexibility-based solutions to implement demand response measures and tackle this challenge. This deliverable from IEA EBC Annex 82 (Energy Flexible Buildings Towards Resilient Low Carbon Energy Systems) contributes to the development of energy flexibility strategies on an aggregated scale. It reports the work carried out during this collaborative initiative derived from both existing literature and original research, focusing on the technical characterization and practical applications of flexibility. The content of this document is clustered by the type of work carried out by the involved IEA EBC Annex 82 participants.

The literature review-based work reported in Chapter 2 starts by examining energy flexibility at an aggregated level and by addressing the main barriers and research gaps for the development of this resource across three design and development phases (market and policy, early planning and design, and operation) and considering different perspectives (policy, economic, technical, professional, social). Participants engaged in this work also found a set of dominant factors impacting energy flexibility considering aleatory uncertainty (i.e., randomness of real-world actions data, such as weather or occupancy patterns) and epistemic uncertainty (i.e., a lack of complete knowledge, such as incomplete data on building characteristics or energy systems) and a group of mitigation measures to tackle identified factors. In terms of energy flexibility and resilience, the work reported in Chapter 2 concluded that both concepts are related to the adaptation of the built environment to external factors, highlighting similarities and differences between the two. The work addresses four critical areas to systematically understand energy resilience (i.e., defining energy resilience, understanding relevant disruptions, quantifying resilience, and improving overall resilience) and identifies four main research gaps (i.e., lack of a universal definition, understanding disruptions, evaluation metrics and improvement strategies).

Building on the work conducted during IEA EBC Annex 67 (Energy Flexible Buildings), Chapter 3 outlines the progression of the energy flexibility characterization method achieved during Annex 82. It traces the development of this concept from the linear Flexibility Function to a non-linear approach and ultimately to the adaptive method and presents two case studies where the Flexibility Function has been applied to explore energy flexibility provided by heating and cooling controllable loads. Moreover, this third chapter addresses different types of building-grid interaction signals and their importance in demand response measures exploring existing flexibility though, for instance, the concept of Flexibility Function. As forecasting energy use and generation profiles is often a requirement in energy flexibility developments, Chapter 3 also presents an open-source, community-driven tool for forecast creation and evaluation.

Examples of simulation and real-world case studies on energy flexibility characterization and use are presented in Chapter 4 and Chapter 5, respectively. Chapter 4 presents the main findings from a collaborative exercise involving 10 research teams worldwide. This joint effort characterizes energy flexibility and applies it in various contexts to reduce electricity costs, while considering grid-related metrics such as peak demand (MW) and peak-to-valley ratio (%). Two field implementations are then presented in Chapter 5, where existing flexibility results from the inertia, like thermal inertia, and the operating regions, like the water tank temperature limits, of existing controllable loads. Such flexibility provides the basis for modifying energy use patterns and achieving the goals of specific demand response measures. While these implementations have shown promising signs in the technical development, large scale demonstrations covering heterogeneous types of systems and application contexts will further validate the robustness of the reported findings.

References

Afroz, Z., Goldsworthy, M., & White, S. D. (2023). Energy flexibility of commercial buildings for demand response applications in Australia. *Energy and Buildings*, *300*, 113533. https://doi.org/10.1016/j.enbuild.2023.113533

Afroz, Z., Wu, H., Sethuvenkatraman, S., Henze, G., Grønborg Junker, R., & Shepit, M. (2024). A study on price responsive energy flexibility of an office building under cooling dominated climatic conditions. *Energy and Buildings*, *316*, 114359. https://doi.org/10.1016/j.enbuild.2024.114359

Albadi, M. H., & El-Saadany, E. F. (2010). Overview of wind power intermittency impacts on power systems. *Electric Power Systems Research*, *80*(6), 627–632. https://doi.org/10.1016/j.epsr.2009.10.035

Al-Mashhadani, S. W. T., & Kurnaz, S. (2024). A Novel Function of a Research Process Based on a Power Internet of Things Architecture Intended for Smart Grid Demand Schemes. *Applied Sciences*, *14*(13), 5799. https://doi.org/10.3390/app14135799

Antal, C., Cioara, T., Antal, M., Mihailescu, V., Mitrea, D., Anghel, I., Salomie, I., Raveduto, G., Bertoncini, M., Croce, V., Bragatto, T., Carere, F., & Bellesini, F. (2021). Blockchain based decentralized local energy flexibility market. *Energy Reports*, *7*, 5269–5288. https://doi.org/10.1016/j.egyr.2021.08.118

Asghar, M. R., Dan, G., Miorandi, D., & Chlamtac, I. (2017). Smart Meter Data Privacy: A Survey. *IEEE Communications Surveys & Tutorials*, *19*(4), 2820–2835. https://doi.org/10.1109/COMST.2017.2720195

Balint, A., Fu, C., & Kazmi, H. (2024). EnFoBench: A community-driven energy forecasting benchmark. *2024 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*.

Balint, A., Raja, H., Driesen, J., & Kazmi, H. (2023). Using Domain-Augmented Federated Learning to Model Thermostatically Controlled Loads. *IEEE Transactions on Smart Grid*, *14*(5), 4116–4124. https://doi.org/10.1109/TSG.2023.3243467

Bergsteinsson, H. G., Sørensen, M. L., Møller, J. K., & Madsen, H. (2023). Heat load forecasting using adaptive spatial hierarchies. *Applied Energy*, *350*, 121676. https://doi.org/10.1016/j.apenergy.2023.121676

Berthelon, M., Lemenager, F., & Sidat, P. (2020). *Systèmes électriques intelligents : le soutien de l'ADEME à l'innovation depuis 2010 - Résultats, enjeux, perspectives (2020)*.

Bjerregård, M. B., Møller, J. K., & Madsen, H. (2021). An introduction to multivariate probabilistic forecast evaluation. *Energy and AI*, *4*, 100058. https://doi.org/10.1016/j.egyai.2021.100058

Blaschke, M. J. (2022). Dynamic pricing of electricity: Enabling demand response in domestic households. *Energy Policy*, *164*, 112878. https://doi.org/10.1016/j.enpol.2022.112878

Brinsmead, T. S., White, S., Bransden, C., Stanley, C., Hasan, K., Alexander, D., Sprague, M., Northey, J., Walgenwitz, G., Nagrath, K., Briggs, C., Leak, J., Harkins-Small, L., Murray-Leach, R., & Jennings, K. (2021). *Flexible demand and demand control. Final report of opportunity assessment for research theme B4. RACE for 2030 CRC.* https://www.racefor2030.com.au/wp-content/uploads/2021/10/RACE-B4-OA-Final-report.pdf

Cai, H., & Heer, P. (2024). Experimental implementation of an emission-aware prosumer with online flexibility quantification and provision. *Sustainable Cities and Society*, *111*, 105531. https://doi.org/10.1016/j.scs.2024.105531

Cai, H., You, S., & Wu, J. (2020). Agent-based distributed demand response in district heating systems. *Applied Energy*, *262*, 114403. https://doi.org/10.1016/j.apenergy.2019.114403

Chantzis, G., Giama, E., Nižetić, S., & Papadopoulos, A. M. (2023). The potential of demand response as a tool for decarbonization in the energy transition. *Energy and Buildings*, *296*, 113255. https://doi.org/10.1016/j.enbuild.2023.113255

Charani Shandiz, S., Foliente, G., Rismanchi, B., Wachtel, A., & Jeffers, R. F. (2020). Resilience framework and metrics for energy master planning of communities. *Energy*, *203*, 117856. https://doi.org/10.1016/j.energy.2020.117856

Chen, Y., Zhang, L., Xu, P., & Di Gangi, A. (2021). Electricity demand response schemes in China: Pilot study and future outlook. *Energy*, *224*, 120042. https://doi.org/10.1016/j.energy.2021.120042

Coulson, J., Lygeros, J., & Dorfler, F. (2019). Data-Enabled Predictive Control: In the Shallows of the DeePC. *2019 18th European Control Conference (ECC)*, 307–312. https://doi.org/10.23919/ECC.2019.8795639

Crosbie, T., Broderick, J., Short, M., Charlesworth, R., & Dawood, M. (2018). Demand Response Technology Readiness Levels for Energy Management in Blocks of Buildings. *Buildings*, *8*(2), 13. https://doi.org/10.3390/buildings8020013

Darwazeh, D., Duquette, J., Gunay, B., Wilton, I., & Shillinglaw, S. (2022). Review of peak load management strategies in commercial buildings. *Sustainable Cities and Society*, *77*, 103493. https://doi.org/10.1016/j.scs.2021.103493

Dawes, G., Kirant-Mitić, T., Jiang, Z., Dréau, J. Le, Cai, H., Cui, J., Townsend, J., Bampoulas, A., Li, R., Lopes, R. A., & Dong, B. (2025). Energy flexibility at multi-building scales: A review of the dominant factors and their uncertainties. *Energy and Buildings*, 116157. https://doi.org/10.1016/j.enbuild.2025.116157

De Zotti, G., Binder, H., Hansen, A. B., Madsen, H., & Relan, R. (2020). Data-driven Nonlinear Prediction Model for Price Signals in Demand Response Programs. *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 1–6. https://doi.org/10.1109/PMAPS47429.2020.9183593

D'hulst, R., Labeeuw, W., Beusen, B., Claessens, S., Deconinck, G., & Vanthournout, K. (2015). Demand response flexibility and flexibility potential of residential smart appliances: Experiences from large pilot test in Belgium. *Applied Energy*, *155*, 79–90. https://doi.org/10.1016/j.apenergy.2015.05.101

El Geneidy, R., & Howard, B. (2020). Contracted energy flexibility characteristics of communities: Analysis of a control strategy for demand response. *Applied Energy*, *263*, 114600. https://doi.org/10.1016/j.apenergy.2020.114600

Ellabban, O., & Abu-Rub, H. (2016). Smart grid customers' acceptance and engagement: An overview. *Renewable and Sustainable Energy Reviews*, *65*, 1285–1298. https://doi.org/10.1016/j.rser.2016.06.021

Energy.Gov. (2024). *BENCHMARK DATASETS of Building Environmental Conditions and Occupancy Parameters*. https://bbd.labworks.org/dataset-search

Fierro, G., Prakash, A. K., Mosiman, C., Pritoni, M., Raftery, P., Wetter, M., & Culler, D. E. (2020). Shepherding Metadata Through the Building Lifecycle. *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, 70–79. https://doi.org/10.1145/3408308.3427627

Fleschutz, M., Bohlayer, M., Braun, M., Henze, G., & Murphy, M. D. (2021). The effect of price-based demand response on carbon emissions in European electricity markets: The importance of adequate carbon prices. *Applied Energy*, *295*, 117040. https://doi.org/10.1016/j.apenergy.2021.117040

Flores-Larsen, S., & Filippín, C. (2021). Energy efficiency, thermal resilience, and health during extreme heat events in low-income housing in Argentina. *Energy and Buildings*, *231*, 110576. https://doi.org/10.1016/j.enbuild.2020.110576

Forrester, S., & Cappers, P. (2021). *Opportunities and Challenges to Capturing Distributed Battery Value via Retail Utility Rates and Programs*. https://doi.org/10.2172/1835837

Foteinaki, K., Li, R., Péan, T., Rode, C., & Salom, J. (2020). Evaluation of energy flexibility of low-energy residential buildings connected to district heating. *Energy and Buildings*, *213*, 109804. https://doi.org/10.1016/j.enbuild.2020.109804

Gellings, C. W. (1985). The concept of demand-side management for electric utilities. *Proceedings of the IEEE*, *73*(10), 1468–1470. https://doi.org/10.1109/PROC.1985.13318

Gholinejad, H. R., Loni, A., Adabi, J., & Marzband, M. (2020). A hierarchical energy management system for multiple home energy hubs in neighborhood grids. *Journal of Building Engineering*, *28*, 101028. https://doi.org/10.1016/j.jobe.2019.101028

Godahewa, R., Bergmeir, C., Webb, G. I., Hyndman, R. J., & Montero-Manso, P. (2021). Monash Time Series Forecasting Archive. *ArXiv Preprint*. https://arxiv.org/abs/2105.06643

Guo, S., & Kontou, E. (2021). Disparities and equity issues in electric vehicles rebate allocation. *Energy Policy*, *154*, 112291. https://doi.org/10.1016/j.enpol.2021.112291

Hall, M., & Geissler, A. (2020). Load Control by Demand Side Management to Support Grid Stability in Building Clusters. *Energies*, *13*(19), 5112. https://doi.org/10.3390/en13195112

Hall, M., & Geissler, A. (2021). Comparison of flexibility factors for a residential building. *Journal of Physics: Conference Series*, *2042*(1), 012036. https://doi.org/10.1088/1742-6596/2042/1/012036

Hamwi, M., Lizarralde, I., & Legardeur, J. (2021). Demand response business model canvas: A tool for flexibility creation in the electricity markets. *Journal of Cleaner Production*, *282*, 124539. https://doi.org/10.1016/j.jclepro.2020.124539

Hanny, L., Wagner, J., Buhl, H. U., Heffron, R., Körner, M.-F., Schöpf, M., & Weibelzahl, M. (2022). On the progress in flexibility and grid charges in light of the energy transition: The case of Germany. *Energy Policy*, *165*, 112882. https://doi.org/10.1016/j.enpol.2022.112882

Hewamalage, H., Ackermann, K., & Bergmeir, C. (2023). Forecast evaluation for data scientists: common pitfalls and best practices. *Data Mining and Knowledge Discovery*, *37*(2), 788–832. https://doi.org/10.1007/s10618-022-00894-5

Hoicka, C. E., Lowitzsch, J., Brisbois, M. C., Kumar, A., & Ramirez Camargo, L. (2021). Implementing a just renewable energy transition: Policy advice for transposing the new European rules for renewable energy communities. *Energy Policy*, *156*, 112435. https://doi.org/10.1016/j.enpol.2021.112435

Holling, C. S. (1973). Resilience and Stability of Ecological Systems. *Annual Review of Ecology and Systematics*, *4*(1), 1–23. https://doi.org/10.1146/annurev.es.04.110173.000245

Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., & Zareipour, H. (2020). Energy Forecasting: A Review and Outlook. *IEEE Open Access Journal of Power and Energy*, *7*, 376–388. https://doi.org/10.1109/OAJPE.2020.3029979

Hosseini, S., & Barker, K. (2016). Modeling infrastructure resilience using Bayesian networks: A case study of inland waterway ports. *Computers & Industrial Engineering*, *93*, 252–266. https://doi.org/10.1016/j.cie.2016.01.007

Hyndman, R. J. (2020). A brief history of forecasting competitions. *International Journal of Forecasting*, *36*(1), 7–14. https://doi.org/10.1016/j.ijforecast.2019.03.015

IEA. (2017). *Digitalisation and Energy*. www.iea.org/reports/digitalisation-and-energy

IEA. (2021). *Electricity system flexibility by source, 2020 and 2030 in the Net Zero Scenario*. https://www.iea.org/data-and-statistics/charts/electricity-system-flexibility-by-source-2020-and-2030-in-the-net-zero-scenario

IEA. (2023). *Tracking Demand Response*. https://www.iea.org/energy-system/energy-efficiency-and-demand/demand-response

Johra, H., Schaffer, M., Chaudhary, G., Kazmi, H. S., Le Dréau, J., & Petersen, S. (2023, September 4). *What metrics does the building energy performance community use to compare dynamic models?* https://doi.org/10.26868/25222708.2023.1309

Junker, R. G., Azar, A. G., Lopes, R. A., Lindberg, K. B., Reynders, G., Relan, R., & Madsen, H. (2018a). Characterizing the energy flexibility of buildings and districts. *Applied Energy*, *225*, 175–182. https://doi.org/10.1016/j.apenergy.2018.05.037

Junker, R. G., Azar, A. G., Lopes, R. A., Lindberg, K. B., Reynders, G., Relan, R., & Madsen, H. (2018b). Characterizing the energy flexibility of buildings and districts. *Applied Energy*, *225*, 175–182. https://doi.org/10.1016/j.apenergy.2018.05.037

Junker, R. G., Kallesøe, C. S., Real, J. P., Howard, B., Lopes, R. A., & Madsen, H. (2020). Stochastic nonlinear modelling and application of price-based energy flexibility. *Applied Energy*, *275*, 115096. https://doi.org/10.1016/j.apenergy.2020.115096

Kaspar, K., Ouf, M., & Eicker, U. (2022). A critical review of control schemes for demand-side energy management of building clusters. *Energy and Buildings*, *257*, 111731. https://doi.org/10.1016/j.enbuild.2021.111731

Kazmi, H., & Tao, Z. (2022). How good are TSO load and renewable generation forecasts: Learning curves, challenges, and the road ahead. *Applied Energy*, *323*, 119565. https://doi.org/10.1016/j.apenergy.2022.119565

Kırant Mitić, T., & Voss, K. (2023). Development of a Joint Penalty Signal for Building Energy Flexibility in Operation with Power Grids: Analysis and Case Study. *Buildings*, *13*(5), 1338. https://doi.org/10.3390/buildings13051338

Langevin, J., Cetin, K., Willems, S., Kang, J., Mahmud, R., Christensen, T. H., Li, R., Knotzer, A., Olawale, O. W., Saelens, D., & O'Connell, S. (2024). Customer enrollment and participation in building demand management programs: A review of key factors. *Energy and Buildings*, *320*, 114618. https://doi.org/10.1016/j.enbuild.2024.114618

Lashmar, N., Wade, B., Molyneaux, L., & Ashworth, P. (2022). Motivations, barriers, and enablers for demand response programs: A commercial and industrial consumer perspective. *Energy Research & Social Science*, *90*, 102667. https://doi.org/10.1016/j.erss.2022.102667

Le Dréau, J., Lopes, R. A., O'Connell, S., Finn, D., Hu, M., Queiroz, H., Alexander, D., Satchwell, A., Österreicher, D., Polly, B., Arteconi, A., de Andrade Pereira, F., Hall, M., Kırant-Mitić, T., Cai, H., Johra, H., Kazmi, H., Li, R., Liu, A., … Saeed, M. H. (2023). Developing energy flexibility in clusters of buildings: A critical analysis of barriers from planning to operation. *Energy and Buildings*, *300*, 113608. https://doi.org/10.1016/j.enbuild.2023.113608

Li, H., Johra, H., de Andrade Pereira, F., Hong, T., Le Dréau, J., Maturo, A., Wei, M., Liu, Y., Saberi-Derakhtenjani, A., Nagy, Z., Marszal-Pomianowska, A., Finn, D., Miyata, S., Kaspar, K., Nweye, K., O'Neill, Z., Pallonetto, F., & Dong, B. (2023). Data-driven key performance indicators and datasets for building energy flexibility: A review and perspectives. *Applied Energy*, *343*, 121217. https://doi.org/10.1016/j.apenergy.2023.121217

Li, H., Wang, Z., Hong, T., & Piette, M. A. (2021). Energy flexibility of residential buildings: A systematic review of characterization and quantification methods and applications. *Advances in Applied Energy*, *3*, 100054. https://doi.org/10.1016/j.adapen.2021.100054

Li, R., Satchwell, A. J., Finn, D., Christensen, T. H., Kummert, M., Le Dréau, J., Lopes, R. A., Madsen, H., Salom, J., Henze, G., & Wittchen, K. (2022). Ten questions concerning energy flexibility in buildings. *Building and Environment*, *223*, 109461. https://doi.org/10.1016/j.buildenv.2022.109461

Li, W., Yigitcanlar, T., Erol, I., & Liu, A. (2021). Motivations, barriers and risks of smart home adoption: From systematic literature review to conceptual framework. *Energy Research & Social Science*, *80*, 102211. https://doi.org/10.1016/j.erss.2021.102211

Li, Y., Rezgui, Y., & Kubicki, S. (2020). An intelligent semantic system for real-time demand response management of a thermal grid. *Sustainable Cities and Society*, *52*, 101857. https://doi.org/10.1016/j.scs.2019.101857

Lian, Y., Shi, J., Koch, M., & Jones, C. N. (2023). Adaptive Robust Data-Driven Building Control via Bilevel Reformulation: An Experimental Result. *IEEE Transactions on Control Systems Technology*, *31*(6), 2420–2436. https://doi.org/10.1109/TCST.2023.3259641

M. Evans, M. Belusko, A. Taghipour, M. Liu, P.F. Keane, J. Nihill, R. Amin, A. Liu, J. Assaf, T. Lau, K. Xing, T. Rainey, C. Stanley, A. Date, B. Bonney, J. Leak, J. Hossain, W. Miller, G. Rosengarten, & F. Bruno. (2021). *Electrification & renewables to displace fossil fuel process heating*. https://racefor2030.com.au/wp-content/uploads/2023/03/B3-OA-Project-Final-Report-July-2021-20210721a.pdf

Madsen, H. (2007). *Time Series Analysis*. Chapman and Hall/CRC. https://doi.org/10.1201/9781420059687

Madsen, H., Tohidi, S. S., Ebrahimy, R., Banaei, M., Ritschel, T. K. S., & Mahdavi, N. (2025). *Recent Trends in Demand-Side Flexibility* (pp. 167–184). https://doi.org/10.1007/978-3-031-74741-0_12

Madsen, H., Tsaousoglou, G., Ritschel, T. K. S., Tohidi, S., Binder, H., Frandsen, H. L., & Junker, R. G. (2024). Incentivising and Activating Multi-Purpose Flexibility for the Future Power System. In *Incentives and Digitalizarion for Flexibility in the Green Transition (Vol. 2, pp. 28-42)*. Danish Utility Regulator.

Makridakis, S., Fry, C., Petropoulos, F., & Spiliotis, E. (2022). The Future of Forecasting Competitions: Design Attributes and Principles. *INFORMS Journal on Data Science*, *1*(1), 96–113. https://doi.org/10.1287/ijds.2021.0003

Martinez, S., Vellei, M., & Le Dréau, J. (2022). Demand-side flexibility in a residential district: What are the main sources of uncertainty? *Energy and Buildings*, *255*, 111595. https://doi.org/10.1016/j.enbuild.2021.111595

Mazzarino, P. R., De Vizia, C., Macii, E., Patti, E., & Bottaccioli, L. (2021). An agent-based framework for smart grid balancing exploiting thermal flexibility of residential buildings. *2021 IEEE International Conference on Environment and Electrical Engineering and 2021 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)*, 1–6. https://doi.org/10.1109/EEEIC/ICPSEurope51590.2021.9584738

Miller, C., Arjunan, P., Kathirgamanathan, A., Fu, C., Roth, J., Park, J. Y., Balbach, C., Gowri, K., Nagy, Z., Fontanini, A. D., & Haberl, J. (2020). The ASHRAE Great Energy Predictor III competition: Overview and results. *Science and Technology for the Built Environment*, *26*(10), 1427–1447. https://doi.org/10.1080/23744731.2020.1795514

Miller, C., Kathirgamanathan, A., Picchetti, B., Arjunan, P., Park, J. Y., Nagy, Z., Raftery, P., Hobson, B. W., Shi, Z., & Meggers, F. (2020). The Building Data Genome Project 2, energy meter data from the ASHRAE Great Energy Predictor III competition. *Scientific Data*, *7*(1), 368. https://doi.org/10.1038/s41597-020-00712-x

Mingjun Wei, Zixin Jiang, Pratik Pandey, Mingzhe Liu, Rongling Li, Zheng O'Neill, Bing Dong, & Mohamed Hamdy. (2024). Energy Resilience in the Built Environment: A Comprehensive Review of Concepts, Metrics, and Strategies. *Renewable & Sustainable Energy Reviews (Submitted)*.

Möhrlen, C., Zack, J. W., & Giebel, G. (2023). *IEA Wind Recommended Practice for the Implementation of Renewable Energy Forecasting Solutions*. Elsevier. https://doi.org/10.1016/C2021-0-03549-5

Mokhtari, R., Junker, R. G., Madsen, H., & Li, R. (2025). A methodological framework for designing dynamic heat price for demand response in district heating. *Energy*, *332*, 135937. https://doi.org/10.1016/j.energy.2025.135937

Mugnini, A., Polonara, F., & Arteconi, A. (2021). Energy flexibility curves to characterize the residential space cooling sector: The role of cooling technology and emission system. *Energy and Buildings*, *253*, 111335. https://doi.org/10.1016/j.enbuild.2021.111335

Naghiyev, E., Shipman, R., Goulden, M., Gillott, M., & Spence, A. (2022). Cost, context, or convenience? Exploring the social acceptance of demand response in the United Kingdom. *Energy Research & Social Science*, *87*, 102469. https://doi.org/10.1016/j.erss.2021.102469

Nord Pool. (2024). *Nord Pool Day-ahead prices*. https://data.nordpoolgroup.com/auction/day-ahead/prices?deliveryDate=latest&currency=EUR&aggregation=DeliveryPeriod&deliveryAreas=DK1

Nweye, K., Liu, B., Stone, P., & Nagy, Z. (2022). Real-world challenges for multi-agent reinforcement learning in grid-interactive buildings. *Energy and AI*, *10*, 100202. https://doi.org/10.1016/j.egyai.2022.100202

Nystrup, P., Lindström, E., Pinson, P., & Madsen, H. (2020). Temporal hierarchies with autocorrelation for load forecasting. *European Journal of Operational Research*, *280*(3), 876–888. https://doi.org/10.1016/j.ejor.2019.07.061

O′Connell, N., Pinson, P., Madsen, H., & O′Malley, M. (2014). Benefits and challenges of electrical demand response: A critical review. *Renewable and Sustainable Energy Reviews*, *39*, 686–699. https://doi.org/10.1016/j.rser.2014.07.098

Ohanu, C. P., Rufai, S. A., & Oluchi, U. C. (2024). A comprehensive review of recent developments in smart grid through renewable energy resources integration. *Heliyon*, *10*(3), e25705. https://doi.org/10.1016/j.heliyon.2024.e25705

Olgyay, V., Coan, S., Webster, B., & Livingood, W. (2020). *Connected Communities: A Multi-Building Energy Management Approach*. https://doi.org/10.2172/1659857

Pallonetto, F., De Rosa, M., D'Ettorre, F., & Finn, D. P. (2020). On the assessment and control optimisation of demand response programs in residential buildings. *Renewable and Sustainable Energy Reviews*, *127*, 109861. https://doi.org/10.1016/j.rser.2020.109861

Panteli, M., Mancarella, P., Trakas, D. N., Kyriakides, E., & Hatziargyriou, N. D. (2017). Metrics and Quantification of Operational and Infrastructure Resilience in Power Systems. *IEEE Transactions on Power Systems*, *32*(6), 4732–4742. https://doi.org/10.1109/TPWRS.2017.2664141

Paterakis, N. G., Erdinc, O., Bakirtzis, A. G., & Catalao, J. P. S. (2015). Optimal Household Appliances Scheduling Under Day-Ahead Pricing and Load-Shaping Demand Response Strategies. *IEEE Transactions on Industrial Informatics*, *11*(6), 1509–1519. https://doi.org/10.1109/TII.2015.2438534

Pauwels, P., Zhang, S., & Lee, Y.-C. (2017). Semantic web technologies in AEC industry: A literature overview. *Automation in Construction*, *73*, 145–165. https://doi.org/10.1016/j.autcon.2016.10.003

Pinto, G., Kathirgamanathan, A., Mangina, E., Finn, D. P., & Capozzoli, A. (2022). Enhancing energy management in grid-interactive buildings: A comparison among cooperative and coordinated architectures. *Applied Energy*, *310*, 118497. https://doi.org/10.1016/j.apenergy.2021.118497

Powell, J., McCafferty-Leroux, A., Hilal, W., & Gadsden, S. A. (2024). Smart grids: A comprehensive survey of challenges, industry applications, and future trends. *Energy Reports*, *11*, 5760–5785. https://doi.org/10.1016/j.egyr.2024.05.051

Pritoni, M., Paine, D., Fierro, G., Mosiman, C., Poplawski, M., Saha, A., Bender, J., & Granderson, J. (2021). Metadata Schemas and Ontologies for Building Energy Applications: A Critical Review and Use Case Analysis. *Energies*, *14*(7), 2024. https://doi.org/10.3390/en14072024

Sala, J., Li, R., & Christensen, M. H. (2021). Clustering and classification of energy meter data: A comparison analysis of data from individual homes and the aggregated data from multiple homes. *Building Simulation*, *14*(1), 103–117. https://doi.org/10.1007/s12273-019-0587-4

Sartori, I., Walnum, H. T., Skeie, K. S., Georges, L., Knudsen, M. D., Bacher, P., Candanedo, J., Sigounis, A.-M., Prakash, A. K., Pritoni, M., Granderson, J., Yang, S., & Wan, M. P. (2023). Sub-hourly measurement datasets from 6 real buildings: Energy use and indoor climate. *Data in Brief*, *48*, 109149. https://doi.org/10.1016/j.dib.2023.109149

Satchwell, A., Cappers, P., & Barbose, G. (2020). Retail Rate Designs: Current Developments in the Industry [In My View]. *IEEE Power and Energy Magazine*, *18*(3), 81–84. https://doi.org/10.1109/MPE.2020.2971844

Satchwell, A., Piette, M., Khandekar, A., Granderson, J., Frick, N., Hledik, R., Faruqui, A., Lam, L., Ross, S., Cohen, J., Wang, K., Urigwe, D., Delurey, D., Neukomm, M., & Nemtzow, D. (2021). *A National Roadmap for Grid-Interactive Efficient Buildings*. https://doi.org/10.2172/1784302

Sørensen, M. L., Nystrup, P., Bjerregård, M. B., Møller, J. K., Bacher, P., & Madsen, H. (2023). Recent developments in multivariate wind and solar power forecasting. *WIREs Energy and Environment*, *12*(2). https://doi.org/10.1002/wene.465

Srithapon, C., & Månsson, D. (2023). Predictive control and coordination for energy community flexibility with electric vehicles, heat pumps and thermal energy storage. *Applied Energy*, *347*, 121500. https://doi.org/10.1016/j.apenergy.2023.121500

Stinner, S., Huchtemann, K., & Müller, D. (2016). Quantifying the operational flexibility of building energy systems with thermal energy storages. *Applied Energy*, *181*, 140–154. https://doi.org/10.1016/j.apenergy.2016.08.055

Tang, R., Li, H., & Wang, S. (2019). A game theory-based decentralized control strategy for power demand management of building cluster using thermal mass and energy storage. *Applied Energy*, *242*, 809–820. https://doi.org/10.1016/j.apenergy.2019.03.152

Tchouanguem Djuedja, J. F., Karray, M. H., Foguem, B. K., Magniont, C., & Abanda, F. H. (2019). *Interoperability Challenges in Building Information Modelling (BIM)* (pp. 275–282). https://doi.org/10.1007/978-3-030-13693-2_23

Thorvaldsen, K. E., Korpås, M., & Farahmand, H. (2022). Long-term Value of Flexibility from Flexible Assets in Building Operation. *International Journal of Electrical Power & Energy Systems*, *138*, 107811. https://doi.org/10.1016/j.ijepes.2021.107811

Tohidi, S. S., Madsen, H., Calì, D., & Ritschel, T. K. S. (2024). *Optimal price signal generation for demand-side energy management*.

Tohidi, S. S., Ritschel, T. K. S., Tsaousoglou, G., Thygesen, U. H., & Madsen, H. (2024). *Stability analysis of nonlinear stochastic flexibility function in smart energy systems*.

Tohidi, S. S., & Yildiz, Y. (2022). Handling actuator magnitude and rate saturation in uncertain over-actuated systems: a modified projection algorithm approach. *International Journal of Control*, *95*(3), 790–803. https://doi.org/10.1080/00207179.2020.1823019

Tohidi, S. S., Yildiz, Y., & Kolmanovsky, I. (2020). Adaptive control allocation for constrained systems. *Automatica*, *121*, 109161. https://doi.org/10.1016/j.automatica.2020.109161

Trindade, A. (2015). ElectricityLoadDiagrams20112014 [Dataset]. In *UCI Machine Learning Repository*. https://doi.org/10.24432/C58C86

Tsaousoglou, G., Junker, R., Banaei, M., Tohidi, S. S., & Madsen, H. (2024). Integrating Distributed Flexibility Into TSO-DSO Coordinated Electricity Markets. *IEEE Transactions on Energy Markets, Policy and Regulation*, *2*(2), 214–225. https://doi.org/10.1109/TEMPR.2023.3319673

UK Power Networks. (2024). *SmartMeter Energy Consumption Data in London Households*. https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households

United Nations Environment Programme; Global Alliance for Buildings and Construction. (2020). *2020 Global Status Report For Buildings And Construction: Towards A Zero-Emissions, Efficient And Resilient Buildings And Construction Sector - Executive Summary*. https://wedocs.unep.org/20.500.11822/34572

United Nations Framework Convention on Climate Change. (2015). *Paris Agreement*.

United Nations Framework Convention on Climate Change. (2022). *COP26 Goals*. https://ukcop26.org/Cop26-Goals

U.S. Department of Energy. (2024). *Connected Communities Funding Program*. https://www.energy.gov/eere/solar/connected-communities-funding-program

Van Overschee, P., & De Moor, B. (1994). N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, *30*(1), 75–93. https://doi.org/10.1016/0005-1098(94)90230-5

Yin, M., Cai, H., Gattiglio, A., Khayatian, F., Smith, R. S., & Heer, P. (2024). Data-driven predictive control for demand side management: Theoretical and experimental results. *Applied Energy*, *353*, 122101. https://doi.org/10.1016/j.apenergy.2023.122101

Yu, H., Zhang, J., Ma, J., Chen, C., Geng, G., & Jiang, Q. (2023). Privacy-preserving demand response of aggregated residential load. *Applied Energy*, *339*, 121018. https://doi.org/10.1016/j.apenergy.2023.121018

Zhang, H., Xiao, F., Zhang, C., & Li, R. (2023). A multi-agent system based coordinated multi-objective optimal load scheduling strategy using marginal emission factors for building cluster demand response. *Energy and Buildings*, *281*, 112765. https://doi.org/10.1016/j.enbuild.2022.112765

Zheng, L., Zhou, B., Cao, Y., Wing Or, S., Li, Y., & Wing Chan, K. (2022). Hierarchical distributed multi-energy demand response for coordinated operation of building clusters. *Applied Energy*, *308*, 118362. https://doi.org/10.1016/j.apenergy.2021.118362

ANNEX 82

EBC

Energy in Building and Communities Programme

www.iea-ebc.org

Technology Collaboration Programme

by iea