Energy & Buildings 316 (2024) 114310

Contents lists available at ScienceDirect

Energy & Buildings

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

Impact of practical challenges on the implementation of data-driven services for building operation: Insights from a real-life case study

John Clauß [a,*], Luis Caetano [a], Åsmund Bror Svinndal [b]

[a] SINTEF Community, Høgskoleringen 13, 7034 Trondheim, Norway
[b] Kiona AS

ARTICLE INFO

Keywords:
Predictive heating control
Data-driven applications
Building control
Office buildings
Energy flexibility
Data-driven services
Practical challenges

ABSTRACT

Data-driven applications in buildings using AI and machine learning have generated a lot of interest, but scaling these applications is challenging due to the uniqueness of each building. During the process of implementing a data-driven predictive heating control in a full-scale real-life office building in Norway, 24 practical challenges were encountered. In this work, those practical challenges are presented, discussed and attributed to four main categories: i) physical limitations, ii) data acquisition and communication, iii) data and model definition and iv) building occupants. Detailed examples for the challenges are provided and more than 15 lessons-learned with regards to developing and implementing data-driven services for building operation are presented. Furthermore, this work discusses how the practical challenges impact the choice of a data-driven approach to control the operation of an office building heating system in a predictive manner and how the practical challenges influence the creation of variation in the measurement data needed to identify a model during normal building operation. Finally, it is shown that a substantial number of practical challenges that were encountered during the operational phase are rooted in the design and construction phase of a building project or from rehabilitation during the operational phase. This highlights the fact that the possible use of data-driven services for building operation should be considered during the tendering and design phase to minimize the number of challenges regarding the widespread implementation of data-driven services for building operation, especially regarding predictive control.

1. Introduction

1.1. Energy flexibility and energy-related data-driven services

Energy use in buildings must be decreased drastically to meet the emission reduction targets of the European Union. The building sector currently accounts for about 40 % of the total energy use in Norway and the government aims to save 10 TWh in the existing building stock by 2030 [1]. Heating and cooling are central needs in commercial buildings, where large amounts of energy are required to maintain correct temperature levels and fresh air for the users. Future buildings will have a more proactive role in the future energy system, where demand side flexibility will be essential to make full use of electricity or heat generated from intermittent renewable energy sources [2]. Regarding heating systems in buildings, demand side flexibility can be understood as the margin by which a building can be operated while still fulfilling its functional requirements [3]. Numerous studies have been conducted concentrating on the building energy flexibility with special focus on the building heating system and its control, e.g. [4–9].

The emphasis on energy flexibility aligns with the shift towards a more sustainable energy use, where the digitalization of the building sector plays a pivotal role in improving energy efficiency and environmental impact. There is a need to facilitate a more targeted use of measurement data collected during building operation, and digital systems inside buildings must be seen in connection with digital technologies and solutions in the energy system. Data-driven applications have been widely adopted in various industries to improve efficiency, reduce costs, and enhance operational performance. Buildings are no exception, where data-driven solutions are implemented for building control [10] and maintenance purposes [11] to achieve a more energy-efficient and sustainable building operation [12] through leveraging the potential benefits of artificial intelligence and machine learning. However, the reality is also that scaling data-driven building services can be challenging due to the uniqueness of each building and due to the fact that the majority of buildings does not work as intended. Katipamula and

* Corresponding author.
E-mail address: john.clauss@sintef.no (J. Clauß).

https://doi.org/10.1016/j.enbuild.2024.114310
Received 1 December 2023; Received in revised form 5 February 2024; Accepted 16 May 2024
Available online 18 May 2024

Nomenclature

AI	Artificial intelligence
API	Application programming interface
BAS	Building automation system
BMS	Building management system
GDPR	General data protection regulation
GUI	Graphical user interface
HVAC	Heating, ventilation and air-conditioning
ID	Identification
KNX	Open standard communicationprotocol for building automation
MPC	Model-predictive control
PRBS	Pseudorandom binary sequence
TWh	Terrawatthours

Brambley [13] state in a review on fault detection in building systems that improper controls and faults in building systems can lead to 15 %-30 % higher energy use. Faults and faulty HVAC system operation can lead to excessive energy waste, undesirable performance of the energy system and poor indoor air quality or indoor thermal environments [14]. Tian et al. [15] conclude in a review on data-driven building performance analysis that building energy simulation techniques have been applied to building energy predictions and energy-efficient design projects, whereas data-driven models for continuous performance monitoring are rarely used in actual real-life projects. Continuous building performance monitoring [15] is a measure that relies heavily on building measurement data.

Predictive control cannot only be seen as a measure to reduce energy use in buildings, but also to facilitate the deployment of building energy flexibility once faults in a building energy system have been diagnosed and fixed. Model-predictive control (MPC) for building control has been a major research field for the last two decades [16]. MPC is a technique in which a mathematical model of a building is used to enable planning of the operation of the building heating system as a function of, for example, forecasted weather and building occupancy. In contrast, conventional building control systems react to a change in weather and occupancy situations when they occur [6]. The proactive "future-oriented" approach of MPC enables the operation of HVAC to be optimized, resulting in improvements in energy efficiency, comfort conditions, power management and building-grid interaction. Studies on MPC for building control usually focus on the development of a control-oriented model to predict the building thermal dynamics, with the building models being either white-box, grey-box or black-box models [17]. The models are most often a simplified replication of the real building, describing the whole building as a one-zone or few-zone model. Besides MPC, other predictive controls for building operation can be rule-based approaches [6], or lately, reinforcement learning has gained more attention due to the availability of measurement data from real-building operation. Most studies on predictive control of building energy systems are either simulation-based or single-case experimental tests [18].

Despite the proven potential to improve building operations, MPC has only rarely been implemented for the operation of real full-scale buildings [19]. When implementing such a control into a real (office) building with several hundred rooms, the approach of using single-zone models may not be accepted by the facility manager of the building or the building owner, as they have to make sure that each rooms keeps a desired thermal comfort level. Zhan and Chong [16] show in a review on the use of MPC in buildings that 70 % of previous research on MPC for building control use simulations and that only 17 % applied MPC to full-scale demonstration sites. As a step towards a more widespread implementation of MPC in real buildings, Zhan et al. [18] propose a "data-centric workflow" for predictive control which, in contrast to a traditional "model-centric workflow for MPC", starts with a control-oriented data curation to determine the data points that are required for the intended control purpose. The data-centric approach is supposed to avoid a cumbersome trial-and-error configuration procedure of the "model-centric workflow". Benndorf et al. [20] ascribe the scarcity of practical implementations to the significant demands of modelling, proficiency, data, hardware, ease of use, and expenses. Amato et al. [19] tested the load shifting potential of the space heating system for a Danish residential building, where a setpoint schedule that mimics the behaviour of an MPC is implemented into the real building. They point out several practical challenges related to the implementing robust solutions for remote control of hydronic radiators. Given the crucial part played by the building sector in reducing carbon emissions, enhancing the replicability, scalability and ease of implementation of MPC and other predictive controls in buildings is imperative.

1.2. Main contribution and scope of the study

Applying predictive controls or data-driven applications for continuous building performance monitoring is not a new phenomenon, but the literature review has shown that there is a big gap between current research activities and the widespread application of data-driven services such as predictive control strategies in full-scale real-life buildings [18]. This can be attributed to practical challenges that occur during the implementation of such services in existing buildings. The main contributions of this work answer to the following questions:

- **Q1: What are the most common *practical challenges for the implementation* of predictive controls and continuous building performance evaluation in office buildings?**
  Research on predictive controls for building energy systems has been ongoing for decades with special focus on types of control-oriented models and optimization algorithms to complement the control framework. Any data-driven application related to building operation relies on measurement data from sensors placed in the building or as part of the energy system as well as states from actuators. Based on a full-scale real-life case study, this work provides an overview of commonly experienced practical challenges, related to the data acquisition and communication, data/model definition and the impact of occupants on the building operation.
- **Q2: How do the practical challenges impact the choice of a *scalable data-driven approach* to control the operation of a building heating system in a predictive manner considering multiple zones in a building (300 +)?**
  Most of the literature focuses on the development of MPC algorithms with a control-oriented model that considers the building as one zone or very few zones [21], linking the energy use for heating to a representative indoor air temperature, which usually is an average of all room temperatures. Within the presented case study, an approach that considers each single zone in an office building with 300 + zones is developed.
- **Q3: How do the practical challenges influence the creation of variation in the measurement data needed to identify a model *during normal building operation*?**
  As pointed out by Drgona et al. [21], when developing a control-oriented model, dedicated experiments may be necessary to create variation in the dataset. However, the variation does not need to cover an entire frequency domain, but rather stay within control-relevant boundaries, e.g. for thermal comfort or radiator supply temperatures. The radiator supply temperature setpoint is most often determined by a *heating curve*, which provides a setpoint for the radiator supply temperature as a function of the outdoor air temperature. Such heating curves are an energy-saving measure where supply temperature setpoints decrease with higher outdoor air temperatures because less heat is needed to keep a sufficient indoor air temperature. However, these curves are usually dependent on the

experience of the person in charge of the operation of the system. The exact setpoints sufficient to keep a desired indoor thermal comfort are dependent on the specific system and building. Furthermore, heating curves only consider the outdoor air temperature for determining the radiator supply temperature, whereas they do not consider any other factors such as solar radiation or the actual heating demand in the building. Tests in the hydronic system were conducted in the case study building to generate data for the radiator supply temperature setpoint being independent of the outdoor air temperature.

This work aims at bridging the gap between the research on data-driven building operation and state-of-the-art building operation of existing real-life buildings. More specifically, this work mainly focuses on the practical challenges that must be overcome when implementing data-driven services for building operation, with only minor focus on the details of the actual data-driven services. Control algorithms will not work properly, if the building has not been prepared for it beforehand. By highlighting obstacles that typically occur during the implementation of data-driven services into real-life buildings, this work will contribute to a faster uptake of data-driven services for building operation. An increased awareness of the challenges is required in the research community because most of these challenges will occur in any building once a data-driven service such as predictive heating control is to be implemented into real-life buildings during their continuous operation.

This work highlights the practical challenges encountered during the implementation of data-driven predictive control in a full-scale real-life building and provides detailed examples for each of the challenges. The paper outlines considerations that must be taken and requirements that must be fulfilled to ease the way towards data-driven applications in buildings. Even though the challenges are mostly encountered with regards to data-driven predictive building control, most of the challenges are prevalent for other data-driven services for building operation. On a general note, it has to be distinguished between data-driven services that have different requirements with regards to reading and writing access to the building. For the implementation of predictive control, reading and writing on a (sub)hourly level is required to update and implement the latest setpoints which can consider external information such as the weather forecast. Other services, such as continuous building monitoring require foremost reading access to query the measurement data to be analysed.

The remainder of the paper is organized as follows: Section 2 introduces the methodology for implementing the exemplary predictive control into the full-scale real-life case study building. Section 3 outlines and elaborates on the practical challenges and lessons-learned regarding the implementation. Section 4 discusses the findings and Section 5 concludes on the lessons-learned and outlines future work.

2. Methodology and case study

The flowchart in Fig. 1 presents the different phases of the research process. The remainder of this section introduces the case study building and elaborates on the data-driven predictive heating control which was implemented during the study.

2.1. General information on the case study

This work presents lessons-learned from the development and implementation of data-driven predictive control into building operation. The work has a predominant practical approach and provides insights into current building operation as a starting point for the implementation of data-driven predictive control. The purpose of this work will be on the fundamental challenges that come with the implementation of data-driven services in real buildings rather than the description and evaluation of the actual data-driven services. Both, the building owner and the provider of the building management system (BMS), are part of the research project to draw from their invaluable experience on building operation. Naturally, it is of utmost importance for the building owner to always keep a satisfactory indoor thermal comfort, whereas the BMS provider also prioritizes the scalability of the solution for building control. These priorities set constraints for the solutions that can be implemented during normal building operation, especially regarding control on thermal comfort and user acceptance of predictive control.

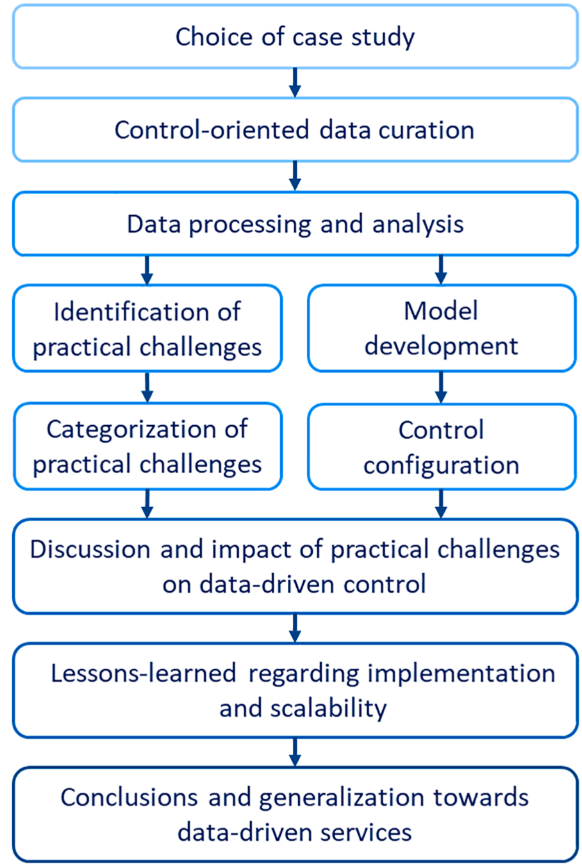

Fig. 1. Flowchart to illustrate the phases of the research process.

2.2. Description of the predictive heating control

This section provides high-level information on the predictive control algorithm. A detailed description of the developed solution for the data-driven model and control algorithm is out of scope of this paper.

The case study building, located in Trondheim, Norway, is a six-story office building consisting of three triangular building blocks, with a capacity of in total 330 rooms/zones. The building was built in 2002, has a heated floor are of approximately 20600 m^2 and a calculated specific delivered energy of 125 kWh/m^2 per year, corresponding to energy class C. Heat is supplied by district heating and the building heat distribution system consists of seven radiator circuits, one floor heating circuit, two snow melting circuits and several circuits to supply heat to the heating batteries of the ventilation systems. Room heating is provided by six of the radiator circuits, with two circuits per building block. The radiator circuits are placed along the façades from the first to the sixth floor. The building is not a research facility, which obviously comes with obstacles when it comes to testing the developed control-oriented models together with the predictive control algorithms, but it also provides a great opportunity to showcase limitations regarding the implementation of data-driven solutions in the current building stock. To the best of the authors' knowledge, the setup of the heating system of the case study building is typical and thus representative for most of the office buildings in Norwegian cities.

Sufficient thermal comfort in all zones has to be ensured at all times. The zones can have different indoor air temperature set-points as well as thermal comfort is perceived differently by the occupants. Besides the measured indoor air temperature, the number of complaints from building occupants is used as an indication for "sufficient" indoor thermal comfort. The facility manager reports complaints, if any. An online survey was conducted during periods of normal operation as well as when the predictive control algorithm was running to investigate whether there is different feedback on perceived thermal comfort during the times that the developed predictive control algorithm is running. Detailed information on the actual survey and findings are not the scope of this paper.

Measurement data collected via the BMS from August 2020 until October 2023 is the basis for developing a data-driven predictive control-oriented model. No additional sensors dedicated to the project have been installed in the building as this would add a limitation for the scalability of the developed data-driven solution.

Regarding the developed predictive control algorithm, the main idea is to avoid using a heating curve to set the radiator supply temperature, but rather set the supply temperature independent of the outdoor air temperature. A data-driven (black-box) model is developed to predict the room temperatures for the next twelve hours. Input features to the data-driven model are weather forecast information, time of the day as well as parameters related to the operation of the HVAC system, such as status of the circulation pump, fan operation, and measured temperatures in the radiator circuits and ventilation system. The radiator supply temperature is also used as an input feature to predict the room air temperature, so that the control algorithm can decide which supply temperature is sufficient to satisfy the room temperature working setpoint. The initial model is trained on roughly two years of measurements that were performed before and during the project.

Within the predictive control framework, this model is used to predict the room temperature for each single room in the building. The predictive control algorithm checks every hour whether there is heating demand in any of the rooms over the control horizon of twelve hours. In case there is no heating demand, the radiator supply temperature is decreased. The algorithm then evaluates whether the measured room temperature is higher or lower than the working setpoints in a case-specific number of rooms (e.g. ten rooms). If the number of rooms that require heating is above the case-specific value, the supply temperature of the radiators is adjusted upwards towards the value of the heating curve. This approach is chosen to not let the coldest room determine the radiator supply temperature for the whole circuit. The case-specific number is chosen in cooperation with the building owner and the facility manager. The maximum possible value for the supply temperature setpoint is limited to the supply temperature which would be applied if the heating curve was used.

3. Practical challenges and lessons-learned

Data-driven predictive control of a building heating system requires continuous and near real-time communication between the server and the real building. There are numerous practical challenges that can occur during the preparation and implementation of a such data-driven service. This section outlines and elaborates on the practical challenges that occurred in the case study. All practical challenges occurred during normal building operation. The challenges are divided into four main categories (Fig. 2): a) physical limitations, b) data acquisition and communication, c) data and model definition and d) building occupants in the focus. The specific challenges are described in detail in the remainder of this section. The challenges are often related and are thus not delimited to one of the four categories. To describe the challenges in this paper, they are placed within the category that seems to be affected the most by each respective challenge. Detailed examples of the challenges are provided to make the reader aware of the issues related to the implementation of the data-driven applications into a real building.

A list of the experienced practical challenges is provided in Table 1, linking all challenges to the specific categories.

3.1. Physical limitations

In this work, physical limitations are linked to the normal building

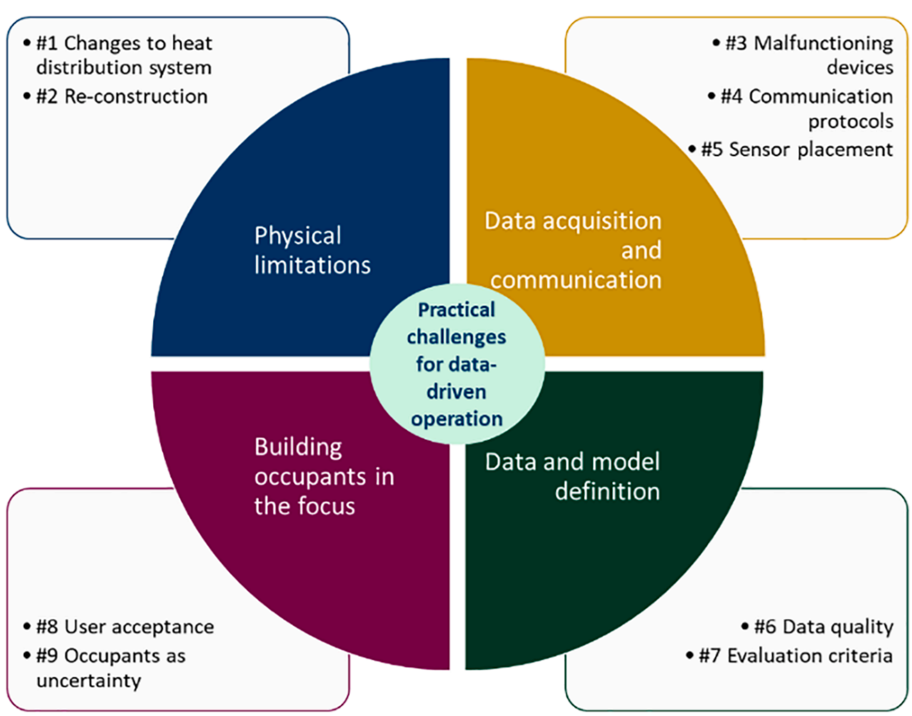

Fig. 2. Categorizing practical challenges for data-driven services for building operation.

Table 1
Practical challenges linked to four different categories.

Challenge	*Category Physical limitation*	*Data acquisition and comm.*	*Data and model definition*	*Building occupants in the focus*
1. Choice of automation communication protocol		x		
2. Level of instrumentation (one energy meter per heat supply component)	x	x		
3. Choice of room regulation and controllers	x	x		
4. Resolution of installed sensors	x	x	x	
5. Future-oriented considerations of the building owner regarding building operation	x	x		
6. Possibility of local room temperature setpoint regulation by the user	x		x	x
7. Possibility for manual solar shading			x	x
8. Improper sensor placement (room regulator, outdoor air temperature, in the hydronic system)		x		
9. Tagging/labelling of measurement points		x	x	
10. Modified settings in heat distribution system after commissioning	x		x	
11. Missing and inaccurate readings		x		
12. Unstable sensor communication		x		
13. Robustness of read/write API		x		
14. Mismatch between onsite-measured and forecasted weather data			x	
15. Stakeholder-specific evaluation criteria for the control objective			x	x
16. User resistance to change				x
17. GDPR-related issues / privacy concerns				x
18. Occupants' increased sensitivity to possible changes in thermal comfort			x	x
19. Use of autonomic local heating units			x	x
20. Manual window opening	x		x	x
21. Changes in the floor plan / re-construction	x	x	x	
22. Room heating and cooling work against each other	x			
23. Insufficient as-built documentation for the building	x			
24. Insufficient variation in historical measurement data for model identification			x	

operation and decisions taken by the building owner prior to the implementation of a data-driven predictive heating control. This can include decisions on choice of hardware (instrumentation), type of room regulation or capacity in the hydronic system.

3.1.1. Heat capacity in the hydronic system

Energy consultants or the facility manager adjusted settings of the building heating system after commissioning as part of an energy saving measure. Heat distribution systems are dimensioned based on a design outdoor temperature, which is −19 °C for Trondheim. The heating system is dimensioned to be able to keep a constant indoor air temperature of 21 °C for three consecutive days without any internal gains in the building. This implicates that the heating system is over-dimensioned for most of its operation. Hence, implemented energy saving measures contribute to running the system more efficiently and more cost-effectively. In our case, the volume flow in the waterborne heat distribution systems was set to approximately 30 % of the dimensioned volume flow which limited the possibility to excite the building to higher room temperatures. During the tests on the hydronic system during the winter season, in some cases, the heating system did not have the capacity to cover the heat demand in some of the zones and, it took several hours (up to seven hours) to increase room temperatures by one degree Celsius in the rooms at the end of the radiator circuit.

3.1.2. Resolution of sensors

The resolution of the installed sensors was too low for the intended purpose of the data-driven service, e.g., air temperature sensors with a resolution of 0.5 K or energy meters with a resolution of 100 kW. For example, regarding the temperature sensor, it is then not possible to differentiate between a temperature of 21.6 °C and 22.4 °C even though the perceived thermal comfort may differ. The measurement value would just be 22 °C.

3.1.3. Room regulation

The heating and cooling system in the room work against each other. The heating system tries to keep the room temperature close to the heating setpoint, which is usually around 21–22 °C. At the same time, the ventilation system supplies fresh air to the room via openings in the ceiling at a supply temperature 3–5 °C lower than the normal room air temperature. The setpoint for the cooling system is usually around 23–24 °C. This is an important consideration for predictive control that aims at varying room temperature setpoints for the heating system up to 23–24 °C, because in the worst case, the heating and cooling systems have the same setpoint and thus working against each other in full capacity.

There is the possibility of local room temperature setpoint regulation by *the occupant.* This can be an issue for a predictive control algorithm as it may happen that the occupants can locally overwrite the indoor air temperature setpoint, which would then interfere with the algorithm and lead to sub-optimal solutions.

3.1.4. Future-oriented considerations of the building owner

The initial building owner did not consider the application data-driven services during the design phase of the building. More and more building owners understand that the energy use in their building can be reduced by making use of measurement data and data-driven applications. However, depending on the services to be implemented, the required instrumentation may not be installed in the building and thus the implementation of a data-driven predictive control may only be possible if new sensors are installed, which again comes with investment costs which increases the time for return-of-investment.

3.1.5. Changes in the floor plan / re-construction

Reconstruction and/or changes in the floor plan have been done without notifying the BMS provider of the changes. The BMS usually comes with a graphical user interface (GUI) which illustrates which room temperature sensor is related to a certain room. In the analysis of room temperatures, low room temperatures were measured in some of the rooms that were surrounded by rooms with satisfying room temperatures even though they are connected to the same radiator circuit according to the GUI. During a visit to the building, it was found that additional internal walls were set up to create additional meeting rooms, but these meeting rooms did not have any radiator and hence no heating.

As-built documentation for the building is insufficient. In a building with several radiator circuits, this is a challenge regarding the development of a control-oriented model as it ideally is known which rooms are connected to a certain radiator circuit.

3.1.6. Lessons-learned

For the experienced practical challenges, the following lessons-learned can be concluded:

1. Check with the energy consultant or facility manager of the building early in the project to ask for potential changes that have been done after commissioning. Also, ask for a commissioning report.
2. Ask for an updated functional description of the room regulation to understand the working principles and setpoints applied for heating and cooling, so it is known to which extend setpoints can be changed.
3. A visit to the building is unavoidable to check any flaws mapped in a first analysis of the building measurement data prior to the visit. This will also help to understand better the measurement data.

3.2. Data acquisition and communication

Stable, robust, and continuous data acquisition and communication are essential for (real-time) data-driven services for building operation. Data quality and reliability of data-driven building services can be affected by technical problems that building systems and sensors may experience. These technical problems include malfunctioning devices, unstable communication, and limitations of communication protocols.

3.2.1. Malfunctioning devices

Unstable communication can result in data loss or delays in data collection and malfunctioning temperature sensors provide missing or inaccurate readings. This leads to incorrect conclusions about the performance of the building's heating system which usually relates energy use to room temperatures. Unstable communication can affect the accuracy of the data analysis. Buildings can experience loss of data and lack of logging which is time-consuming to fix. Three examples are:

i) *"Flatliners":* These are measurements that appear as a constant measurement value, but in fact the GUI of the BMS system is not receiving updated data. Here, the problem can be the Building Automation System (BAS) or the GUI which uses wrong ID tags rather than the actual hardware.
ii) *Related to the way measurement data is collected and transferred from the field level to the BMS, sensors may not respond to a request for measurement values from the BMS,* which continuously asks for a value until the sensor replies. This is a challenge for control-related tasks which require the latest state, for example a measured room temperature, to determine the next control action.
iii) *It may happen that the Application Programming Interface (API) does not return a response to requests,* or it provides different values for the same parameters for the same time slot when requested several times. Regarding online predictive control, it is recommended to implement a "fallback" strategy back to the business-as-usual control that is triggered if the API communication fails.

Thus, developing robust and reliable data collection pipelines is important to ensure that data-driven services are based on accurate and reliable data.

3.2.2. Communication protocols

The choice of communication protocols can limit the scalability of the data-driven application. In the case study on data-driven predictive control, the initial aim of the project was to alter room temperature setpoints in all zones to make use of the thermal mass of the building. However, some of the room regulators use a KNX system which is an event-based system in which a sensor must perceive a change/event before it acts. One of the biggest disadvantages of KNX room regulators is the limited number of changes that some manufacturer‘s devices may allow over their lifetime. It is therefore not desired to trigger an event, such as changing the values of room temperature setpoints within a mode, whereas it is preferred to switch between the "modes" of the regulators. The different modes are usually set up as a deviation (± 1 or $\pm$ 2 °C) from the basis setpoint. Frequent changes of the value within a mode will break the regulator in the long-term meaning that their internal "storage" is used up. For example, if a predictive control algorithm proposes three or four basis set-points per day, all changes will be stored and once the read/write limitation is reached, a new sensor must be purchased.

3.2.3. Sensor placement

Improper physical placement of a sensor leads to measurement data that cannot be trusted. A correct placement is crucial for data-driven control purposes. Four examples of improper placement and their impact on the measurement data are:

i) *The room regulator is placed close to the ventilation outlet which supplies fresh air to the room.* This leads to rather low indoor air temperature measurements because ventilation is usually used to provide cooling to the rooms. The measured room temperature is therefore not representative of the temperature in the room.
ii) *The room regulator is mounted on an outer wall,* which can lead to a lower measured indoor air temperature, especially during shoulder and winter season. The measured temperature is the surface temperature instead of the indoor air temperature. In the case study, the sensor was mounted to a stone wall which was north facing, so that the measured room temperature was rather low.
iii) *The outdoor (sometimes also indoor) air temperature sensor is hit by the sun,* which will lead to higher measurement values that do not represent the actual conditions. As the outdoor temperature is usually used to determine the supply temperature of the radiator circuit or floor heating circuit via a heating curve (outdoor temperature compensation curve), a higher outdoor temperature will lead to a lower supply temperature, thus posing the risk of not

meeting the heat demand of the building. In the case study, several temperature sensors were placed on the outer facades of the building and the sensor on the west façade was connected to the heating curve, thus leading to a 15 °C lower radiator supply temperature compared to the "correct" heating curve in the afternoons and evenings.

iv) *Temperature sensors that measure the supply temperature of a radiator circuit are placed too close to the main heat exchanger towards district heating.* Heat conduction along steel pipes may lead to a high temperature measurement in times when the circulation pumps of the secondary circuit are stopped. In the case study, the measured supply temperature to the floor heating circuit was as high as 90 °C even though there was no heat demand, and the circulation pumps were stopped. This issue is critical for a safe operation. The facility manager moved the sensor further away from the district heating heat exchanger and also found that a bypass valve was missing in the heat distribution system, the day after these high temperatures were measured.

3.2.4. Level of instrumentation

Too few heating meters on hydronic loops or (electricity) sub-meters are installed. The non-existence of submetering is a known issue. It makes it difficult to separate processes on component or zone/room level. Regarding control purposes, the level of submetering determines what the most detailed level for a supervisory controller can be. Model-based predictive controls often express changes in the room air temperature as a function of energy supplied to the room. However, it is not common practice to have energy meters for each radiator. Submetering at radiator circuit level allows to investigate whether a representative area-weighted average temperature for the indoor air temperature of all rooms connected to the radiator circuit can be used in a control framework.

3.2.5. Lessons-learned

The following lessons-learned can be concluded (continued numbering):

4. The robustness of the API used to read from and write to a building must be monitored.
5. Regarding communication protocols, an analysis of the specific components used on-site should be done to evaluate whether the components meet the requirements regarding the data-driven service. Otherwise, a cost-benefit analysis of operational cost savings due to a price-based control vs. the investment costs for new room regulators would be required.
6. Regarding data collection, it is key to understand how data is being collected, which data is collected, and whether the data is reliable. This understanding is important because the collected data may not be relevant or useful for the intended purpose, e.g., for predictive control. Accurate and consistent data collection is necessary, along with monitoring and maintaining data quality. However, implementing these measures may increase the workload of facility managers, who often only take action when building occupants raise complaints.
7. The initial aim of the case study was to adjust room temperature setpoints for each zone in a model-predictive control framework to activate the building thermal mass as a thermal storage. However, it was found from initial tests that it was not possible to simply set room temperature setpoints, so that it therefore was decided to neglect the idea of adjusting room temperature setpoints, but rather focus on the deviation of the radiator supply temperature as an energy saving measure. Here, the focus on a solution that scales rather easily across buildings was the deciding factor.
8. Correct sensor place is crucial for developing reliable data-driven services. Also in this regard, a visit to the building is recommended.
9. Submetering is not crucial, but at least beneficial for the development and acceptance of model-based predictive control. The facility manager required a sufficient thermal comfort in all rooms at all times and thus a one-zone model of the building using one representative average indoor air temperature was not accepted.

3.3. Data and model definition

Poor data quality and availability are other major practical challenges for the development and implementation of data-driven applications for building control. Building control algorithms rely on accurate and timely data to make decisions regarding building operation.

3.3.1. Data quality

Data quality and availability can be compromised by a variety of factors. This challenge is closely related to Challenges #3 to #5 in Fig. 2, e.g., sensor malfunction, outdated and incompatible data formats as well as improper sensor placement. These challenges ultimately lead to poor quality in the raw data which in turn often requires cumbersome preprocessing of the data to be able to use it for developing data-driven / machine learning algorithms. Examples for challenges that require analysis and pre-processing:

i) *Sensor tags/labels were not standardized.* Erroneous labels, metric units and not using standardized naming conventions lead to misunderstandings, productivity loss for unknowing users and limit the scalability of a data-driven service as the screening of sensor labels is often cumbersome.

ii) *Control signals are typically subject to high-frequency variations and thus sensitive to averaging.* On top of that, instantaneous measurements may be available at too sparse resolution. In general, the sampling rate has a major impact on the data volume and the control response capability. For instance, averaging binary signals to an hourly-averaged value may not give physically useful information.

iii) *There can be a mismatch between the forecasted outdoor air temperature (from weather forecast) and the locally measured outdoor temperature.* This has direct implications in a predictive building operation framework. Even though most weather forecast providers correct their short-term forecasts (for the next hours) with observations interpolated from several nearby weather stations, the resolution and mean conditions of the grid that is used may not represent local climatic effects fully, which is due to unresolved features such as complex topography, an intricate coastline or urban heat islands. For example, in winter the outdoor air temperature measured locally at a single building within a one-kilometre scale may differ from the forecasted temperature due to thermal inversion effects which are not resolved. In the case study, temperature differences between forecasted and onsite-measured outdoor air temperatures of up to 5 °C were observed. This directly affects a predictive heating operation: e.g., if the weather forecast predicts a 5 °C higher outdoor air temperature, the radiator supply temperature will be lower compared to the heating curve and the energy demand of the building might not be met and the indoor thermal comfort requirements may be jeopardized.

iv) *BMSs in existing buildings are missing the integration for some of the technical equipment*, e.g. heat/cooling pumps, outdoor screen systems, or fan coils. The issue of missing integrations is a problem in many buildings and is a problem specifically for building operation because you don't have any data on systems that are not integrated into the BMS.

Insufficient variation in the historical measurement data. Developing control-oriented models for predictive controls requires available unbiased measurement data with sufficient variation in the data. Dedicated tests to excite the heating system or the rooms are often necessary to identify such a model. For example, aiming at room temperature setpoints changes to make use of the building thermal mass requires data that mimics the thermal dynamics of the building relating the indoor air temperature to the energy use for heating. It is very important to construct a test signal in such a way that there is no cross-correlation between the test signal and other input variables. For instance, it is important to avoid a 24-hour variation in the test signal (since this period is normally seen for solar radiation and outdoor air temperature). To that end, PRBS for building excitation is one possibility, but using a PRBS may lead to violations of the indoor thermal comfort [22], so that an approach must be developed that can be run during "normal" operation of the building, meaning that big variations in the indoor air temperature are not allowed as it may affect the indoor thermal climate negatively and thus the well-being and productivity of the occupants. Similarly, tests must be done to excite the heating system, if the intention is to set a radiator supply temperature that is independent of the outdoor air temperature. This will allow for possibly satisfying room heating demands with lower radiator supply temperatures. Regarding the case study, in the current and previous operation of the building, there may be room temperature setpoint changes, whereas the radiator supply temperature has always been set via the heating curve, which does not consider heat gains from solar radiator or building occupants. One part of the developed scalable solution for predictive heating control is a model that predicts the room air temperature and uses future weather conditions and a radiator supply temperature which is independent of the outdoor air temperature. As part of the model identification the hydronic system must be excited to populate the training data set with measurement data for a radiator supply temperature setpoint which is independent of the outdoor air temperature.

3.3.2. Evaluation criteria

Stakeholders have different prioritizations when it comes to the objectives of the predictive control. Stakeholders must understand the importance of different evaluation criteria that may counteract each other and agree on a prioritization. For example, the building owner may prioritize investment costs and return-on-investment, whereas improved occupant satisfaction, such as thermal comfort, may be difficult to quantify and may conflict with energy-saving goals, prompting the need for a trade-off decision.

3.3.3. Lessons-learned

The following lessons-learned can be concluded (continued numbering):

10. Regarding sensor labelling, it is very cumbersome to identify the relevant sensors manually from the database. Here, the development of semantic data models can be part of the solution.
11. Regarding evaluation criteria, a case-specific approach that considers the interests of all stakeholders is necessary to evaluate a data-driven service for building operation.
12. There is usually not enough variation in the historical measurement data to identify a robust control-oriented model. Dedicated experiments for exciting the building are necessary.
13. However, PRBS may lead to poor thermal comfort in the rooms during the experiment period (being either too warm or too cold) and can therefore not be used during normal building operation. Therefore, other approaches that can be applied during normal building operation should be used to excite the building and heating system.

3.4. Building occupants in the focus

In the remainder, the focus is on non-technical practical challenges regarding the development of data-driven services. Human factors such as behaviour, perception, and motivation can significantly impact the adoption, effectiveness, and acceptance of data-driven applications for building control, where the success of such an application relies deeply on the behaviour and actions of building occupants.

3.4.1. Occupants as uncertainty factors

Besides occupancy of a room, decisions taken by occupants individually can cause challenges for the predictability of usage of the building or certain rooms: building occupants or staff may actively change settings or override system controls, leading to suboptimal performance and increased energy consumption. The following challenges were experienced:

i) *Occupants installed additional movable electric heaters in the room.* It was noticed in a room temperature analysis, that the room temperature increased up to 25 °C from around 4p.m. to 8a.m. each day, even though the room temperature set-point given by the BMS was kept constant at 22 °C.
ii) *Windows are opened manually and left open for an extensive period.* In a room with an active HVAC system, it causes the system to work at increased capacity to maintain the desired temperature.
iii) *Internal shading devices are applied manually.* Knowledge about the possibility of manually drawing internal shading device is important for the use of data-driven predictive heating control, as the influence of solar radiation on the room temperature is different compared to what would be expected, and thus a control-oriented would consider a wrong predicted indoor air temperature.

3.4.2. User acceptance

The resistance to change as a user-related challenge can lead to slower adoption and limited benefits of data-driven services. Building occupants and staff may be resistant to new data-driven applications for building control, either due to a lack of understanding or a preference for traditional control methods. Examples related to user acceptance are:

i) *Privacy concerns and mistrust of the occupants slow down processes.* A survey among the occupants was performed during the project. GDPR-related concerns were raised even after it was communicated that the national centre for research data gave consent that no GDPR-related issues are expected for conducting the survey among the building occupants. When the research partner contacted the occupants only a very limited number of occupants participated, even though information about the project and the purpose of the survey was provided beforehand. Once the building owner asked the occupants to reply to the survey, a higher response rate was observed.
ii) *Increased awareness of possible thermal comfort changes can lead to increased negative feedback on perceived thermal comfort.* After notifying building occupants that tests of the heating system will be conducted in near future, the number of complaints increased even though no changes were implemented. Regarding user acceptance of predictive control, it is important to know there is a psychological effect of simply being aware that changes may happen because the occupants have an increased focus on thermal comfort.

3.4.3. Lessons-learned

The following lessons-learned can be concluded (continued numbering):

14. It is very beneficial to establish an ongoing open and transparent communication process between the building owner and occupants to remove any privacy concerns of the occupants.
15. Building owners and operators must prioritize user engagement and education. This may involve training staff and occupants on the benefits of a data-driven application for building operation and providing clear instructions on how to use the service effectively.

16. The occupants' awareness about possible changes in thermal comfort has implications on survey design which includes how and how much information should be communicated to the occupant at which time in the process.

4. Discussion

The presented practical challenges can be attributed to several categories outlined in Table 1. Similarly, some of the lessons-learned which are outlined in the previous section can be drawn from multiple challenges. The key lessons-learned for developing and implementing data-driven predictive control in real-life buildings are:

- The application of data-driven services is not trivial as each building is different. Therefore, more focus should be on the use of ontologies and interoperability.
- Think scalability of the solution to be developed. This is crucial for a widespread implementation of data-driven services and thus economic turnover for the provider. If you want to have a scalable solution, you should think about the application and what is required for it first. Installation of additional sensors dedicated to the purpose of the data-driven application can of course be beneficial but was out of scope of this study.
- Know what you measure! Collection of data is easy, but can you trust it? Is your measurement data labelled correctly? Placement of the sensors used in the data-driven predictive control setup, is a bottleneck for a successful implementation as unfortunate placement can lead to wrong assumptions for the control.
- After an initial analysis of the measurement data, a visit to the building is recommended. This will help to better understand the data. Furthermore, a checklist with the most important points should be prepared prior to the visit. Available documentation on functional descriptions as well as settings of the technical systems (e.g., heating, cooling, ventilation, snow melting), room regulation and communication protocols should be gathered.
- Malfunctioning devices, stable communication and limitations of communication protocols are major technical barriers for data-driven applications. Not all networking protocols used in building automation are designed for cloud-centric streaming.
- Data-quality is of outmost importance. Data processing and preparation is time-consuming. This task should be consistent and automated as much as possible. Furthermore, the question of "how much data is enough data" depends on the service to be developed and the variation in the dataset.
- Integration can be complex. BMS platforms may require programming and licensing fees to enable features. Also, the network infrastructure may need to be upgraded to support IoT connections.
- Identify user-friendly solutions that are easy to use and understand. Building owners and operators need to work closely with vendors and system integrators to achieve a user-friendly solution that satisfies all relevant user demands.
- Improved performance and energy savings can already be achieved by monitoring, analysing, and acting on spotted anomalies during operation. Data-driven predictive control including any machine learning models can come on top of this.
- Even though all the challenges are experienced during the operation of the building, the root of the practical challenge may not lay in the building operation. The experienced practical challenges are illustrated in Fig. 3 which also links the challenges to specific building project/lifecycle phases that have the major impact on the occurrence of each of the challenges. Even though, all challenges are experienced during building operation, it can be seen from Fig. 3 that almost half of the practical challenges can be attributed to the tendering, design and construction phase of the building lifecycle. When it comes to the tender, the technical requirements should be described in detail, e.g. on the instrumentation, data logging, sampling intervals or data structure.

5. Conclusion

This paper presents practical challenges that must be overcome when

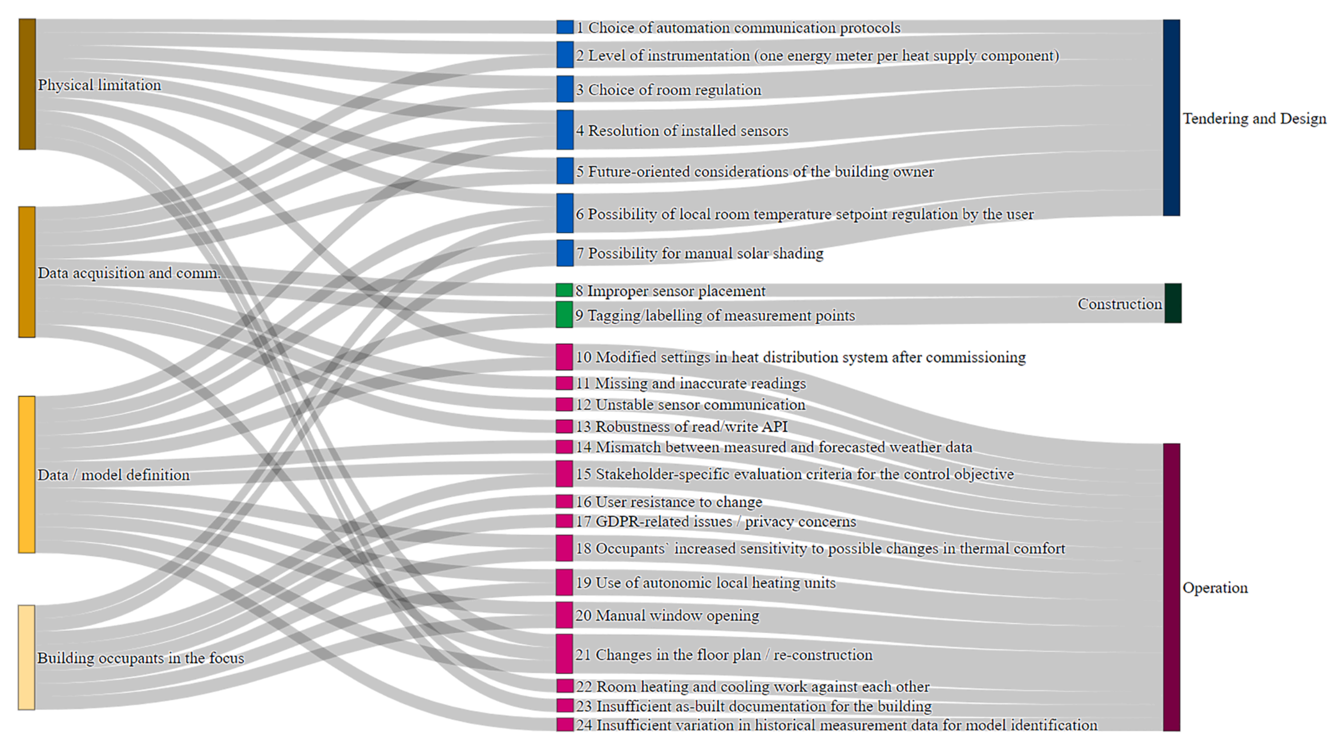

Fig. 3. Linking the practical challenges to the lifecycle phase that has the biggest impact on occurrence of the challenge.

developing and implementing data-driven predictive control in existing real-life buildings. 24 practical challenges are outlined, described, and attributed to four different categories. On top of this, more than 15 lessons-learned are presented aiming at providing guidance and easing the way for the successful implementation and development of data-driven services in buildings. The superior purpose of this paper is to help bridging the gap between research on data-driven building control and state-of-the-art building operation of existing real-life buildings. By increasing the awareness of the identified challenges this work will contribute to a faster uptake of data-driven services for building operation because most of these challenges will occur in any building once a data-driven service such as predictive heating control is to be implemented into real-life buildings during continuous operation. Hence, the outlined challenges can be used as a checklist to know what to focus on or to clarify before implementing a data-driven service. Even though the challenges are mostly encountered with regards to data-driven predictive building control, many of these challenges are prevalent for other data-driven services related to building operation.

The first research questions (Q1) focused on the *most common practical challenges* for the implementation of predictive controls and continuous building performance evaluation in office buildings. A total of 24 practical challenges are presented. These challenges were divided into four different categories: i) physical limitations, ii) data acquisition and communication, iii) data and model definition and iv) occupants in the focus. It can be concluded that a substantial number of practical challenges that were encountered during the operational phase are rooted in the tendering, design and construction phase of a building project. This highlights the fact that the possible use of data-driven services for building operation should be considered during the tendering and design phase. Furthermore, the commissioning process should have a special focus on whether requirements on what must be in place for implementing data-driven services for building operation are met.

The second research question (Q2) investigates how the practical challenges *impact the choice of a scalable data-driven approach* to control the operation of a building heating system in a predictive manner considering multiple zones in a building (300 +). This work concludes from the journey of getting towards a scalable solution that can be implemented in several buildings with minor efforts by the service provider. As part of the presented case study, an approach that considers each single zone in an office building with a total of 300 + zones is developed. Initially, the energy flexibility provided by the building thermal mass was to be deployed, but due to limitations in the communication protocols and very limited control over the room temperature setpoints resulting from the installed room regulators, it was found that this is a major obstacle towards a scalable solution. Considering the practical challenges that occurred during the development and implementation, a scalable solution that determines a radiator supply temperature independent of the outdoor air temperature was proposed and implemented in the building. This is a relevant contribution towards a more efficient operations of the building heating system. It can be concluded that an initial screening of how the building is operated is essential to decide on which components and technical systems to act on. After an initial analysis of the measurement data, a visit to the building is absolutely recommended. As part of the screening, available documentation on functional descriptions of the technical systems, room regulation and communication protocols should be gathered.

The third research question (Q3) addresses how the practical challenges influence the creation of variation in the measurement data needed to identify a model *during normal building operation*. To identify a model that can be used in the predictive control framework, dedicated tests to excite the heat distribution system need to be performed. In the historical measurement data, the radiator supply temperature is always directly correlated to the outdoor air temperature via the heating curve. The initial dedicated tests were performed over a three-month period during heating season aiming to get a greater variation in the measurement data. As part of the tests, a continuous room temperature analysis is developed to determine whether there is heating demand in the rooms. This analysis is done on an hourly basis. As the general approach, if there is no heating demand, meaning that the measured indoor air temperature is above the room temperature setpoint, the radiator supply temperature is decreased. The maximum possible radiator supply temperature is the one determined by the heating curve. The algorithm evaluates whether the measured room air temperature is higher or lower than the room temperature setpoint. If the number of rooms that require heating is above a case-specific value (e.g. 10 rooms), the supply temperature of the radiators is adjusted upwards towards the value of the heating curve. The case-specific value is chosen in cooperation with the building owner and the facility manager. It was found that the proposed approach can be applied during normal building operation, so that it is possible to increase the variation in the measurement data in a continuous way without risking strong violations in the indoor thermal comfort. It furthermore allows to identify models considering data from various operating conditions, such as weekends and weekdays as well as ventilation “on” or ventilation “off”.

Further research should focus on solutions that help streamlining the implementation of data-driven services for building operation. Focus will be on the use and implementation of semantic data modelling as part of the solution.

CRediT authorship contribution statement

John Clauß: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Luis Caetano:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Åsmund Bror Svinndal:** Writing – review & editing, Project administration, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

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

Data availability

The authors do not have permission to share data.

Acknowledgements

The authors would like to acknowledge the Norwegian Research Council for the financial support under the grant number 317442 – Databygg project. Furthermore, the authors would like to acknowledge IEA EBC Annex 81 “Data-driven smart buildings”.

References

[1] Regjeringen, “Granavolden Plattformen,” 2019. [Online]. Available: https://www.regjeringen.no/no/dokumenter/politisk-plattform/id2585544/#k13.

[2] S.Ø. Jensen, et al., IEA EBC Annex 67 Energy Flexible Buildings, Energy Build. 155 (2017) 25–34, https://doi.org/10.1016/j.enbuild.2017.08.044.

[3] J. Clauß, C. Finck, P. Vogler-finck, P. Beagon, “Control strategies for building energy systems to unlock demand side flexibility – A review Norwegian University of Science and Technology , Trondheim , Norway Eindhoven University of Technology , Eindhoven , Netherlands Neogrid Technologies ApS / Aalborg,” pp. 611–620, 2017.

[4] R.E. Hedegaard, T.H. Pedersen, S. Petersen, Multi-market demand response using economic model predictive control of space heating in residential buildings, Energy Build. 150 (2017) 253–261, https://doi.org/10.1016/j.enbuild.2017.05.059.

[5] D. Fischer, H. Madani, On heat pumps in smart grids: A review, Renew. Sustain. Energy Rev. 70 (2017) 342–357, https://doi.org/10.1016/j.rser.2016.11.182.

[6] J. Clauß, S. Stinner, I. Sartori, L. Georges, Predictive rule-based control to activate the energy flexibility of Norwegian residential buildings: Case of an air-source heat

pump and direct electric heating, Appl. Energy 237 (Mar. 2019) 500–518, https://doi.org/10.1016/J.APENERGY.2018.12.074.
[7] Y. Ruan, J. Ma, H. Meng, F. Qian, T. Xu, J. Yao, Potential quantification and impact factors analysis of energy flexibility in residential buildings with preheating control strategies, J. Build. Eng. vol. 78, no. June (2023) 107657, https://doi.org/10.1016/j.jobe.2023.107657.
[8] F. Lu, Z. Yu, Y. Zou, X. Yang, Energy flexibility assessment of a zero-energy office building with building thermal mass in short-term demand-side management, J. Build. Eng. vol. 50, no. February (2022) 104214, https://doi.org/10.1016/j.jobe.2022.104214.
[9] Y. Li, et al., Energy flexibility analysis and model predictive control performances of space heating in Japanese zero energy house, J. Build. Eng. vol. 76, no. June (2023) 107365, https://doi.org/10.1016/j.jobe.2023.107365.
[10] E.T. Maddalena, Y. Lian, C.N. Jones, Data-driven methods for building control — A review and promising future directions, Control Eng. Pract. 95 (2020), https://doi.org/10.1016/j.conengprac.2019.104211.
[11] Q. Wen, J.P. Zhang, Z.Z. Hu, X.S. Xiang, T. Shi, A Data-Driven Approach to Improve the Operation and Maintenance Management of Large Public Buildings, IEEE Access 7 (2019) 176127–176140, https://doi.org/10.1109/ACCESS.2019.2958140.
[12] C. Fan, D. Yan, F. Xiao, A. Li, J. An, X. Kang, Advanced data analytics for enhancing building performances: From data-driven to big data-driven approaches, Build. Simul. 14 (1) (2021) 3–24, https://doi.org/10.1007/s12273-020-0723-1.
[13] S. Katipamula, M.R. Brambley, Review article: Methods for fault detection, diagnostics, and prognostics for building systems—A review, part I, HVAC R Res. 11 (1) (2005) 3–25, https://doi.org/10.1080/10789669.2005.10391123.
[14] Z. Chen, et al., A review of data-driven fault detection and diagnostics for building HVAC systems, Appl. Energy 339 (2023), https://doi.org/10.1016/j.apenergy.2023.121030.
[15] Z. Tian, X. Zhang, S. Wei, S. Du, X. Shi, A review of data-driven building performance analysis and design on big on-site building performance data, J. Build. Eng. 41 (May) (2021), https://doi.org/10.1016/j.jobe.2021.102706.
[16] S. Zhan, A. Chong, Data requirements and performance evaluation of model predictive control in buildings: A modeling perspective, Renew. Sustain. Energy Rev. 142 (February) (2021), https://doi.org/10.1016/j.rser.2021.110835.
[17] M.D. Knudsen, L. Georges, K.S. Skeie, S. Petersen, Experimental test of a black-box economic model predictive control for residential space heating, Appl. Energy 298 (2021) 117227, https://doi.org/10.1016/j.apenergy.2021.117227.
[18] S. Zhan, M. Quintana, C. Miller, A. Chong, From Model-Centric to Data-Centric: A Practical MPC Implementation Framework for Buildings, in: BuildSys 2022 - Proc. 2022 9th ACM Int. Conf. Syst, Energy-Efficient Build. Cities, Transp, 2022, pp. 270–273, https://doi.org/10.1145/3563357.3564077.
[19] V. Amato, R.E. Hedegaard, M.D. Knudsen, S. Petersen, Room-level load shifting of space heating in a single-family house – A field experiment, Energy Build. 281 (2023) 112750, https://doi.org/10.1016/j.enbuild.2022.112750.
[20] G.A. Benndorf, D. Wystrcil, N. Réhault, Energy performance optimization in buildings: A review on semantic interoperability, fault detection, and predictive control, Appl. Phys. Rev. 5 (4) (2018) pp, https://doi.org/10.1063/1.5053110.
[21] J. Drgoňa, et al., All you need to know about model predictive control for buildings, Annu. Rev. Control 50 (May) (2020) 190–232, https://doi.org/10.1016/j.arcontrol.2020.09.001.
[22] M.D. Knudsen, R.E. Hedegaard, T.H. Pedersen, S. Petersen, System identification of thermal building models for demand response - A practical approach, Energy Procedia 122 (2017) 937–942, https://doi.org/10.1016/j.egypro.2017.07.426.