P.A.D. Turner, T.W. Rushby, M. Manfren, P.A.B. James, S. Gauthier, A.S. Bahaj, T. Sweetnam, S. Kim, Ellis Ridett, Embedding energy flexibility capability in air source heat pumps via third-party control: Insights from a field trial on residential buildings in England, *Applied Energy*, Volume 389, 2025, 125705, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2025.125705. (https://www.sciencedirect.com/science/article/pii/S0306261925004350)

Applied Energy 389 (2025) 125705

Contents lists available at ScienceDirect

Applied Energy

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

Embedding energy flexibility capability in air source heat pumps via third-party control: Insights from a field trial on residential buildings in England

P.A.D. Turner [a], T.W. Rushby [a], M. Manfren [a,*], P.A.B. James [a], S. Gauthier [a], A.S. Bahaj [a], T. Sweetnam [b], S. Kim [b], Ellis Ridett [a]

[a] Energy and Climate Change Division, Sustainable Energy Research Group, Faculty of Engineering and Physical Sciences, University of Southampton, Boldrewood Campus, SO16 7QF Southampton, United Kingdom
[b] Good Energy Ltd, Monkton Park Offices, Monkton Park, Chippenham, SN15 1GH, UK

HIGHLIGHTS

- Embed energy flexibility capability in ASHPs and clusters of buildings.
- Potential synergies between ESCOs, aggregators, installers and manufacturers.
- Field testing of automated third-party control for groups of customers.
- Average power reduction in percentage equal to 88.2% across different events.
- Possible development of physics-informed data-driven energy analytics.

ARTICLE INFO

Keywords:
Air Source Heat Pumps
Energy Flexibility
Demand Side Management
Demand Response
Decarbonisation
Energy Analytics
Energy behaviour
Data-driven methods

ABSTRACT

This research investigates energy flexibility in residential building clusters transitioning from gas boilers to air source heat pumps, within the broader context of rapid decarbonisation of both building stock and electric grid in the UK. The study field trialed a scalable control approach embedded in heat pumps, as part of the EPSRC funded project "LATENT: Residential heat as an energy system service". The project explores a flexibility paradigm where aggregators and Energy Service Companies (ESCOs) partner with installers and manufacturers to leverage small-scale flexibility sources, to enable swift flexibility deployment in clusters of buildings. Flexibility events were scheduled for ESCO customers in Southern England during typical UK electric grid peak hours, using an intervention and control approach across customer groups. Findings reveal insights into third-party control operation, events duration, override requests, achievable flexibility and user behaviour/comfort preferences. Peak shaving strategies implemented resulted in an average power reduction of 88.2% across events with a maximum demand reduction of 1.581 kW, averaged throughout the cluster of buildings. Override requests occurred in only 2.7% of potential cases, with events lasting from 30 to 120 minutes. The study also assessed temperature dependence in flexibility performance at the cluster level. Results indicate the feasibility of longer energy flexibility events, contingent on a more advanced analysis of technical and social constraints. In conclusion, the research emphasises the significance of conducting field trials to showcase potential for energy flexibility solutions in optimising the operation of electric infrastructure.

1. Introduction

The decarbonisation of the electricity grid and the concomitant electrification of transportation, heating of buildings, and industrial processes are the two primary strategies targeted at reducing greenhouse gas emissions. These effort support the international objective of limiting global warming, as established in the Paris Agreement and reaffirmed in subsequent Conference of Parties (COP) summits.

* Corresponding author.
E-mail address: m.manfren@soton.ac.uk (M. Manfren).

https://doi.org/10.1016/j.apenergy.2025.125705
Received 30 June 2024; Received in revised form 4 March 2025; Accepted 9 March 2025
Available online 20 March 2025

Embedding energy flexibility capability in air source heat pumps via third-party control: insights from a field trial on residential buildings in England

P. A. D. Turner[a], T.W Rushby[a], M. Manfren[*a], P. A. B. James[a], S. Gauthier[a], A.S. Bahaj[a],T Sweetnam[b], S Kim[b], Ellis Ridett[a]

[a] Energy and Climate Change Division, Sustainable Energy Research Group, Faculty of Engineering and Physical Sciences, University of Southampton, Boldrewood Campus – SO16 7QF Southampton, United Kingdom
[b] Good Energy Ltd, Monkton Park Offices, Monkton Park, Chippenham, SN15 1GH, UK

Abstract
This research investigates energy flexibility in residential building clusters transitioning from gas boilers to air source heat pumps, within the broader context of rapid decarbonisation of both building stock and electric grid in the UK. The study field trialed a scalable control approach embedded in heat pumps, as part of the EPSRC funded project "LATENT: Residential heat as an energy system service". The project explores a flexibility paradigm where aggregators and Energy Service Companies (ESCOs) partner with installers and manufacturers to leverage small-scale flexibility sources, to enable swift flexibility deployment in clusters of buildings. Flexibility events were scheduled for ESCO customers in Southern England during typical UK electric grid peak hours, using an intervention and control approach across customer groups. Findings reveal insights into third-party control operation, events duration, override requests, achievable flexibility and user behaviour/comfort preferences. Peak shaving strategies implemented resulted in an average power reduction of 88.2% across events with a maximum demand reduction of 1.581 kW, averaged throughout the cluster of buildings. Override requests occurred in only 2.7% of potential cases, with events lasting from 30 to 120 minutes. The study also assessed temperature dependence in flexibility performance at the cluster level. Results indicate the feasibility of longer energy flexibility events, contingent on a more advanced analysis of technical and social constraints. In conclusion, the research emphasises the significance of conducting field trials to showcase potential for energy flexibility solutions in optimising the operation of electric infrastructure.

Keywords: Air Source Heat Pumps; Energy Flexibility; Demand Side Management; Demand Response; Decarbonisation; Energy Analytics; Energy behaviour; Data-driven methods.

Highlights:

- Embed energy flexibility capability in ASHPs and clusters of buildings.
- Potential synergies between ESCOs, aggregators, installers and manufacturers.
- Field testing of automated third-party control for groups of customers.
- Average power reduction in percentage equal to 88.2% across different events.
- Possible development of physics-informed data-driven energy analytics.

[*] Corresponding author:
Tel: +44 02380592869
e-mail: M.Manfren@soton.ac.uk

Table of contents

1 Introduction

The decarbonisation of the electricity grid and the concomitant electrification of transportation, heating of buildings, and industrial processes are the two primary strategies targeted at reducing greenhouse gas emissions. These effort support the international objective of limiting global warming, as established in the Paris Agreement and reaffirmed in subsequent Conference of Parties (COP) summits.

The electrification of end-uses is widely acknowledged as a crucial element in attaining complete decarbonisation. However, it is important to understand its influence from a broader point of view, considering the entire energy system in a framework of accelerating changes [1], where complementaries needs to be addressed [2].

In turn, the transition to a decarbonized electricity system, driven by the rapid growth of renewable energy sources, like wind and solar in particular, has introduced new challenges related to price volatility in electricity markets [3]. The intermittent and variable nature of renewable power generation can lead to sudden swings in supply, causing significant fluctuations in wholesale electricity prices. For this reason, energy efficiency [4] and energy flexibility [5] have to be consider together for the evolution of energy infrastructures [6], acknowledging benefits at both the system and building level [7].

Flexibility at the building level can be defined as the building's ability to adjust over short periods of time, ranging from a few hours to a couple of days. This includes energy demand and/or generation in response to weather conditions, user requirements, and energy network conditions. This adjustment should not compromise the building's technical capabilities or occupant comfort significantly [8]. Examples of flexible loads commonly found in buildings are electric storage heaters, for both space and water heating, as well as heat pumps, air conditioners, and auxiliary devices like water circulation pumps.

Interest in energy flexibility in buildings has grown substantially in the last few years. This is in part due to the increasing penetration of renewable energy technologies potentially leading to higher volatility of prices, mentioned before, and the increasing adoption of heat pumps [9] and electric vehicles [10]. Additionally, rising energy prices have made customers more attentive to energy costs.

In an energy transition perspective, characterised by an increasing decarbonisation effort for the energy sector [11], the development of flexibility solutions is closely connected to the challenge of achieving an optimal balance between reducing energy demand through efficiency measures and decarbonising energy carriers [12]. This issue is especially pertinent in the context of heating in buildings [13] as determining the optimal balance is not a trivial problem [14]. Additionally, as highlighted by Le Dréau et al. [15], it is essential to consider the issue of creating flexibility solutions at the level of "clusters" of buildings. This is crucial due to the technical challenges associated with leveraging small sources of flexibility (e.g. individual residential buildings), which are influenced by the stochasticity of user behaviour [16]. Moreover, there are advantages for Transmission System Operators (TSOs), Distribution System Operators (DSOs), and utilities [17], where the aggregation of flexibility resources in a portfolio can reduce the uncertainty of single building behaviour.

Concurrently, heat pumps are essential for reducing carbon emissions in buildings [14] where the grid carbon intensity is low and represents, at least potentially, a flexible load. The widespread adoption of heat pumps in the UK poses considerable problems and opportunities for energy flexibility, especially in light of the distinctive characteristics of national electric load profiles [18]. The social acceptability of demand response is also crucial [19], and this represents a problem in the UK, where peak load conditions occur in winter around late afternoon and early evening, coinciding with the time when a high percentage of residential households are occupied [20]. The increased dependence on temperature of load profiles due to air source heat pumps [21] complicates grid management, as lower temperatures result in higher demand peaks [22], and the potential coincidence between high heating demand (during low temperature period) and reduced renewable power output (notably from wind) amplifies future challenges. Furthermore, the emerging problem of a "social license" to automate [23,24] and other socioeconomic factors [25] may hinder the effectiveness of flexibility measures and need to be considered.

The EPSRC project LATENT [26] seeks to tackle the challenge of harnessing the flexibility of individual residential buildings which are using heat pumps. Heat pumps are controlled by a third-party entity that acts as an aggregator of flexibility resources. The aggregator supervises a portfolio of energy efficiency interventions and energy flexibility measures and, therefore, needs a scalable solution that can be implemented for a broader set of clients in the near future. This paper presents and analyses the findings of a field trial carried out in 2023, offering insights for the characterisation of energy flexibility and suggesting areas for future research. The aim is to address the potential flexibility gap [27], inherent to the use of standard boundary conditions and a priori estimation, by means of seamless and socially acceptable solutions.

2 Background and literature review

The general framework of energy flexibility research at the state-of-the-art is presented in Section 2.1. It highlights the need for a multidisciplinary approach that addresses technical, social, normative, and economic barriers to the large-scale deployment of flexibility in building. Additionally, it considers its value chain to indicate potential market based solutions in this direction. Subsequently, Section 2.2 addresses the characterisation of energy flexibility, the definition of control strategies required to enable flexibility actions, and the implementation of energy flexibility events.

2.1 Energy flexibility research framework, value chain, regulations and social acceptability

Research on energy flexibility at the state of the art involves multiple components, such as technology, operational strategies, and economic, social, and normative issues, as underlined in recent literature reviews. Initially, the focus of research has been on the techniques employed to quantify the potential of energy flexibility, as documented by Lopes et al. [28]. Tang et al. discuss the categorization of different energy flexibility resources in buildings [29], mentioning the impact energy services related to heating and cooling. Energy flexibility has a broad definition [8], as mentioned in the introduction, and is frequently associated with the concept of demand response. Demand response refers to a mechanism that aims to shift peak electricity demand to periods of lower demand, usually through the use of financial incentives or a pricing scheme. Chen et al. reviewed measures to improve energy flexibility in buildings for demand response (DR) [30], while Yan et al. reviewed price-based mechanisms for DR [31].

The deployment of energy flexibility solutions can occur at multiple scales, ranging from individual buildings, to clusters of buildings and districts [32], as demonstrated in the analysis conducted by Luc et al. [33]. According to Vigna et al. [34], the indicators used to measure flexibility in clusters of buildings and districts can be fundamentally similar, but clearly they vary with respect to the data aggregation level. The issue of consolidating small flexibility resources will be addressed later in this section. Li et al. [8] indicated how a considerable amount of current research on energy flexibility has focused on the operational phase and on the development of control strategies.

Multiple types of control strategies for demand side flexibility are presented in the review by Clauß et al. [35], considering both rule-based and model predictive control (MPC). Péan et al. review control strategies more specifically focused on heat pump systems in buildings [36]. Control strategies frequently employ data-driven predictive techniques, as discussed by A. Kathirgamanathan et al. [37]. It is essential to critically evaluate the trade-off between the complexity involved in the formulation and implementation of control strategies and the actual results that can be attained [38], while also considering the user's freedom [39] of adjusting operations based on individual needs and preferences. Section 2.2 contains a more in depth analysis of the technical issues pertaining to the control of flexibility events.

Despite the inherent advantages of adjusting energy demand in the short term and actively interacting with the electric grid, the widespread implementation of energy flexibility in buildings faces technological, regulatory, social, and economic obstacles. Li et al. [40] emphasize the need for advancements in technology, social acceptance, business models, and regulations to effectively harness this flexibility. D'Ettore et al. [5] examine the barriers and incentives affecting demand response programs from the perspectives of end-users and aggregators, acknowledging difficulties due to limited market products for small-scale consumers but not mentioning the potential benefits of aggregating end-users into clusters. In contrast, Le Dréau et al. [15] offer a comprehensive review of flexibility in clusters of buildings, considering policy, planning, design methods, and operational challenges. A crucial challenge identified is efficiently aggregating buildings into clusters to leverage small individual sources of energy flexibility on a large scale for system-level benefits.

To fully leverage the benefits of energy flexibility at the system level, understanding the importance of aggregation in the demand response value chain is essential. Sousa and Soares [41], suggest that benefits and barriers can be categorized into market, financial, social, technological, and environmental groups, relating them to various actors and stakeholders. A deep comprehension of both end-users' and aggregators' perspectives can lead to mutually beneficial frameworks that facilitate large-scale implementation of flexibility measures.

However, while the business model canvas presented by Hamwi et al. [42] offers a strong value proposition for utilities and grid operators, which can trade flexibility in the wholesale energy market [43] and gain additional revenues through demand-side services or arbitrage [44], it is less compelling for ESCOs and consumers, presenting challenges for large-scale deployment due to social acceptability issues. Social acceptability, often termed "social license" [23,24], is a key concern, measuring the level of public consent and approval. Simultaneously, entities like energy communities are emerging to support collective energy management and flexibility. For example, the European Union references communities in two directives: Directive 2018/2001 [45] defines Renewable Energy Communities (RECs), and Directive 2019/944 [46] outlines Citizen Energy Communities (CECs).

Implementing a Time of Use (TOU) tariff for retail customers, combined with an automated and scalable flexibility control mechanism, and allowing an aggregator to trade in the wholesale energy market, could offer a market-driven solution to harness flexibility benefits at the system level [47], even with small individual flexibility resources within an energy community. This approach would protect retail customers from the challenges and uncertainties of managing energy with dynamic pricing. Without a third-party entity managing flexible resources—such as aggregating customers, forecasting, and handling risks—retail customers would face significant price volatility, expected to increase for the reasons outlined in the introduction. Not all consumers are equally prepared to adjust or shift their demand, especially with devices like heat pumps that rely on thermostat settings. An information asymmetry between utilities and consumers makes it difficult for customers to adjust energy usage based on dynamic price signals, leading to equity and acceptability issues. Aggregators can collaborate with ESCOs to jointly provide energy efficiency and flexibility services [6], addressing these challenges and providing a better service to customers.

2.2 Energy flexibility characterisation, definition of control strategies and roll-out of events

Buildings can function as a flexibility resource because of the thermal inertia of their construction components, which enables them to store and release heat (heat storage is a "passive" characteristic of the building). They can also make use of thermal and electric storage technologies, operating as "active" storage solutions.

Building energy modelling techniques can be categorized into white-box, grey-box, or black-box models [48], going from the more detailed physical representation to a purely data-driven approach. Detailed physics-based models (white-box) can be used for an in-depth assessment of building-grid interaction [49], where buildings are active nodes [50] of a multi-commodity network [51,52], which can be represented using a graph-based formalism [53]. White-box and grey-box models are particularly important for establishing a connection between the potential for flexibility and the actual physical characteristics of building technologies. Data-driven models however are extremely useful in operation as they can be trained and validated using measured data, as will be illustrated later.

Short-term storage enabled by thermal inertia plays a critical role in relation to thermal load [54], which can be supplied by electric technologies such as heat pumps. Low-parameter grey-box models are particularly effective for simulating the impact of control strategies aimed at enhancing flexibility [55]. A part of this modelling approach is analogous to the Quick U-value Building (QUB) method used to estimate the heat loss coefficient in buildings [56] by means of temperature attenuation. At its very basics, the mechanism that can be exploited for flexibility consists in turning off the heating system for a short period of time and ensuring that, due to thermal inertia, the indoor temperature will not drop too much, creating uncomfortable conditions for users. Building "passive" properties, such as building thermal inertia, can offer grid assistance that is comparable in some cases to batteries, which are an "active" storage solution. An example of this may be seen in the study conducted by Papachristou et al. [57], where they examine the energy flexibility of Dutch office buildings both at the individual building level and as a cluster of buildings. The effect, at the aggregated level, can become comparable to a large battery park. For existing buildings, flexibility characterization often involves surveys and data-driven methods, previously mentioned; Measurement and Verification (M&V) principles and state-of-the-art open-source software can be leveraged to provide reliable baseline estimates for load profiles to be used for counterfactual analysis, i.e. to compare a typical load profile with one in which flexibility measures are applied, for example to understand heat pump flexible operation [58]. This approach can be implemented at scale, an example in this sense is the pay-for-performance programme delivered by PG&E [47].

Kazmi et al. [59] illustrate the importance of aggregating users' data to improve the accuracy of operational energy demand forecasts and load profiles. They also discuss the possibility of partitioning load profiles data into weekdays and weekends, and of taking into account the influence of outdoor temperature on heating and cooling related loads.

Along the same line, a modelling approach considering Time Of Week and Temperature (TOWT) dependence of load profiles was proposed initially by Price [60] to analyse electric load shape and its variability. It has been used then for the quantification of changes in electricity use due to demand response [61], in the context of utility scale efficiency programs [62] and it is currently implemented in open source software OpenEEmeter [63] and part of CalTRACK methods [64] for M&V, used in the pay-for-performance programme [47] mentioned earlier.
More recently, an updated version of the TOWT model has been used for short-term hourly predictions aimed at demand flexibility applications [65] and modified with the inclusion of solar radiation as input variable [66] to enhance its predictive ability; due to its formulation (based on piecewise linear regression) TOWT is also interpretable [67,68], i.e. intelligible in human terms, and does not require post-hoc techniques [69] for its interpretation, which represent in many cases a limit. TOWT model includes an ambient temperature dependent component, which can become very relevant when heat pumps are present. Both heating and cooling loads in buildings and the coefficient of performance (COP) of air source heat pumps (ASHPs) are temperature dependent, and these variables can be visualised both with scatterplots and regression lines [70] to ease the comparison between design and operational performance [71].
However, while reliable and interpretable data-driven modelling options are available [68] for load modelling, the definition of an appropriate baseline model requires at least a few months of hourly data (to characterize both time of week and temperature dependent energy usage). A baseline and comparison group approach is used by Glass et al. [72] for Demand Response (DR) performance evaluation at scale, where the underlying methodology is based on Difference in Differences (DID) technique; DID is used also with time series in the context of policy evaluation [73] and econometrics [74].
Alternatively, a simpler approach could involve examining influential variables considered by the aforementioned models (e.g., time of day, weekday/weekend, ambient temperature) in comparable conditions across a cohort of relatively similar customers, to evaluate the effect of flexibility through an intervention and control approach [75].
Further, beyond time of week and outdoor air temperature dependence, reported before, the control of load profiles requires the knowledge (even if approximated) of basic building characteristics and indoor temperatures.
Regarding building characteristics, assessing energy flexibility in residential buildings relies on understanding the key features of the building stock [76,77]. When dealing with the existing building stock, the combined use of open data (such as Energy Performance Certificates (EPC) [78]), statistics (such as building stock surveys and energy benchmarks), and audits to analyse and describe their characteristics can be considered a valid approach [37,79]. This is in light of the difficulties of defining control strategies for clusters of buildings with limited information.
Regarding building indoor air temperatures, one of the challenges with upscaling control for clusters of buildings is the need to consider indoor air temperatures, which are constrained by comfort requirements and depend on user behaviour. In principle, a large variety of control strategies may be implemented for energy flexibility [35] and more specifically for flexibility of heat pumps [36]. However, practical limitations in implementation may necessitate scheduled set-points [80], restrictions on the comfort band [81] or periodic temperature set-points [82]. These restrictions should not significantly disrupt the normal operation of the building, but rather aim to find a suitable balance between comfort and performance [83] and also act as an energy efficiency measure [84].

However, aggregating data from clusters of buildings simplifies the creation of load profile forecasting models, as it entails predicting an aggregate behaviour for the entire cluster, rather than concentrating on specific data for each individual building [59]. The possibility to control distributed resources in a privacy preserving way [85] is very interesting to make it acceptable from a social point of view. UK electric typical peak conditions happen in the late afternoon and early evening [18]. This time interval coincides with the time when households are typically occupied. This may create problems of social acceptance of events conducted within this time interval [19].

Therefore, especially in challenging operational conditions, field testing is essential to identify and address the potential gaps between estimated flexibility and the actual flexibility achievable in practice [27]. Further, the potential "flexibility gap" is influenced not only by technical factors but also by the characteristics and behaviours of customers [25], which can significantly impact the effectiveness of flexibility measures.

In the review on flexibility at the building cluster level by Le Dréau et al. [15] they report of no cases for the UK at the cluster level, but a number of ongoing projects in the UK context focuses on exploring and enhancing energy flexibility. For instance, four notable projects deal with the problem of flexibility in relation to heat load deferral, EQUINOX by National Grid [86]**,** HeatFlex UK [87], NEDO [88] discussed by Crawley et al. [89] and the research conducted by Gupta and Morey [90].

These projects include trials to reduce demand through directly messaging users to request a reduction of heating during a critical event or by direct control. Sending a text message to request turning off an ASHP is a low-cost approach and does not require any hardware change in the home but has clear limitations as it relies on the householder to take on "flexibility labour" [23]. Conversely, a direct control approach seems a more effective option, even though households acceptance, comfort preferences and the characteristics of the service provided [91] needs to be carefully evaluated. Indeed, the role and impact that heat pumps can have in providing flexibility to the UK electric grid is an open problem and Crawley et al. [92] highlight multiple areas to be explored further, among them the acceptability and effectiveness of third-party control.

3 Methods and tools

The previous section examined many interconnected aspects of the latest research on energy flexibility, emphasising the fast changing landscape within the field. The project “"LATENT: residential heat as an energy system service" [26] seeks to incorporate energy flexibility control capability into ASHPs to provide UK residential customers with a seamless experience of flexibility. This flexibility intervention is managed by a third-party entity, such as an ESCO (Energy Service Company) or aggregator.

Following the evidence from literature, the dimension of "clusters" of buildings seems the most interesting one for large scale deployment of building to grid (B2G) services. Small sources of flexibility, such as residential buildings, can be valorized by ESCOs and aggregators, partnering with installers and manufacturers to create innovative business models, within the complex (and rapidly evolving) value chain of energy flexibility, discussed in Section 2.

The magnitude of flexibility resources, both passive and active, at the individual building level, particularly in the residential sector, is relatively small. However, there is potential for exploiting these resources more effectively at the cluster level, encompassing multiple buildings and communities of users. Theoretically individual customers have the potential to take individual action and adjust their building operation according to a dynamic time-of-use (TOU) tariff. However, it is challenging for this to be enacted at scale without

automated solutions and proper data analytics. ESCOs and aggregators, on the other hand, are better equipped to address this problem as they already have energy analytics in place for their operations.
They may also partner with installers and manufacturers of technology, particularly heat pumps, to devise solutions suitable for large-scale implementation, leveraging the trend of natural gas boilers replacement with ASHPs.
LATENT project proposes a seamless third-party control flexibility experience to participants, where they do not have to take action individually (the flexibility event is managed by the third-party) but they can override the third-party control at any moment, with the ASHP returning to normal operation (e.g. if they feel uncomfortable, or for other contingent reasons). The field trial presented in this paper aims to answer the following research questions:

1. What is the effectiveness of third-party control mechanisms for energy flexibility of heat pumps for residential customers?
2. How much flexible load does this represent for a cluster of buildings?
3. What is the number of override requests received by customers during energy flexibility events?
4. What is the duration of a flexibility event that can be tolerable by households, before they opt to override?

The project methodology is structured to answer these research questions and in *Table 1* the project characteristics are summarised by presenting essential information such as location, building and participant characteristics, stakeholders and field trial duration.

Table 1: General characteristics of the research project and field trial

Location	Country	UK (Great Britain)
	Regions	Mostly Southern England
Building and customer characteristics	Type of customers/buildings	Residential
	Equipment controlled (technology for flexibility)	Electrical air source heat pumps (heating and domestic hot water)
	Control strategy	Rule Based Control (RBC)
	Other technologies involved	Smart thermostats, Data loggers, Weather data services
	Measured data	Details provided in *Table* 2
	Key performance indicator of energy flexibility	Peak power reduction (power reduction achievable during peak hours for the UK grid)
	Baseline (reference performance) and counterfactual analysis	Control and intervention approach (on similar set of customers)
	Communication	Gateways, Internet connection, custom API using Modbus standard
Stakeholders	Industrial partners	ESCO, aggregators, heat pump manufacturers and installers
	Customers	Households (residential)
	Other stakeholders	Distribution System Operator (DSO)

Field trial duration and participants	Duration	From 10/01/2023 to 29/03/2023
	Participants (buildings) and groups	Intervention group: 30 at the end of the trial. Control group: 30

Figure 1 summarises the key aspects of the approach proposed in the LATENT project, highlighting the type of industrial partners involved and the aim to achieve flexibility at the cluster level for buildings served by heat pumps. This happens by leveraging a third-party control solution that can operate seamlessly, but with the possibility for customer to override flexibility action at any time, depending on their needs.
In *Figure 2* a scheme of the third-party control infrastructure is reported. Each ASHP was installed with an internet connected controller managed by a MODBUS interface. A custom API application was developed based on ThingsBoard platform [93]. Data collected through the interface was transferred to an online platform via the householders' internet connection then stored on a cloud-based platform to enable further analysis.

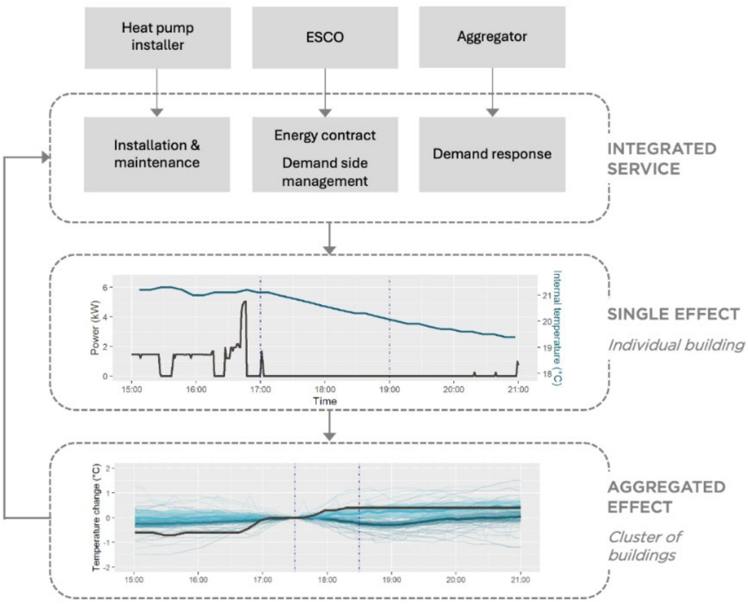

Figure 1: Key components of the LATENT project approach to energy flexibility

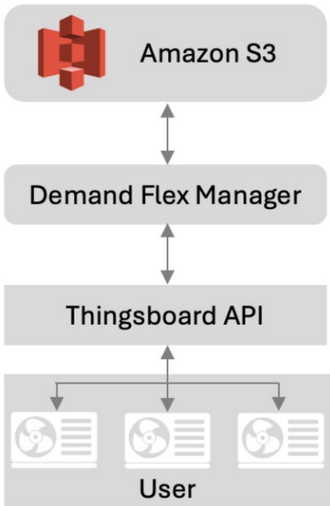

Figure 2: Third-party control and demand flexibility manager scheme

Table 2 details the key measured data collected for analysis during the field trial testing. Measured data are necessary in particular for the characterisation of flexibility potential in relation to ambient temperature (influencing both thermal load of building and COP of air source heat pumps) and for the analysis of indoor air temperature changes during events, influencing thermal comfort and, consequently, potential override requests. Internal air temperature sensors were located in the living spaces of the dwelling, typically a living room.

Table 2: Measured data in the field trial testing of energy flexibility

Category	Sensor installed	Measurement interval	Accuracy specification	Accuracy assumed
Power and energy metering	Modbus interface power data feeds: Electric power input to ASHP, Thermal power output space heating and domestic hot water (DHW)	1 minute	Actual (sensor specification)	0.001 kW (1 W)
Ambient temperature	Climate analytics (Visual Crossing web service [94].)	1 hour	Assumed based on web-service specifications	0.1 ºC
Internal air temperature	Indoor temperature sensors through either TADO [95], or installed Bluetooth monitor.	10 minutes	Actual (sensor specification)	0.01 ºC

Demand side management and energy flexibility strategies have been illustrated in Section 2 and they usually include efficiency measures, load modulation, load shedding (peak shaving), load shifting and on-site generation [8]. In relation to large scale impact of energy flexibility, Johra et al. examine the integration of energy efficiency measures, on-site generation, and flexibility actions such as peak shaving, load shifting, and valley filling to enable buildings in providing grid services on a nationwide scale [77]. In this research, the monitored buildings were exclusively subjected to the installation of the heat pump and did not undergo any additional efficiency measures, such as envelope renovations, nor did they incorporate on-site generation employing photovoltaics. For this reason, the focus of the trial was peak shaving but the mechanisms implemented and tested can be used to obtain other flexibility action like load shifting and valley filling using the thermal storage capabilities of the envelope [54,76] and of domestic hot water (DHW) storage [96] where present, as exemplified in *Table 3*.

Table 3: Flexibility actions, mechanisms implemented and examples of operation

N.	Flexibility action	Mechanisms implemented	Examples
1	Peak shaving (load shedding)	Modulation, switch-off HP	HP is switched off or power modulated to reduce power demand.
2	Load shifting	Modulation, switch-on HP	HP is switched on and/or power modulated, to pre-heat rooms and/or pre-heat DHW storage (increase of room/storage temperature).
3	Valley filling	Modulation, switch-on HP	HP is switched on and/or power increase, room and/or DHW storage temperatures increase.

Flexibility is quantified using two indicators [8] power reduction and flexibility factor (FF), following the considerations reported in Section 2. Heat pump power demand reduction during the flexibility event is expressed in kW and in percentage and calculated with the following formulas.

$$\Delta P_{el,HP,flex} = P_{el,HP,ref} - P_{el,HP,flex} \tag{1}$$

$$\Delta P_{el,HP,flex\%} = \left(1 - \frac{P_{el,HP,flex}}{P_{el,HP,ref}}\right) * 100 \tag{2}$$

Where:
$P_{el,HP,flex}$ is the heat pump power demand in a flexibility event.
$P_{el,HP,ref}$ is the heat pump power demand in reference operational conditions (baseline).
$\Delta P_{el,HP,flex}$ is the heat pump power reduction during the flexibility event.
$\Delta P_{el,HP,flex\%}$ is the heat pump power reduction during the flexibility event expressed in percentage.

Flexibility factor (*FF*) indicator can be used to highlight the need and potential of flexibility. A broad definition of FF is provided in the review by Li et al. [8], showing how this indicator can be used with different quantities (e.g., energy, cost and emissions, which represent integrals in time), computed respectively during high load or high price hour intervals (typically aligning with peak demand hours, 16:00-20:00 in the UK electric system when demand is high and renewable generation low) and low load or low price hour intervals (e.g. overnight hours and midday periods). The high load/price intervals considered for the calculation in this research is 16:00-20:00 during weekdays, corresponding to peak hours for the electric grid system at national scale in the UK, as discussed earlier in Section 2. A general formula is provided hereafter.

$$FF = \frac{Quantity_{load,low} - Quantity_{load,high}}{Quantity_{load,low} + Quantity_{load,high}} \tag{3}$$

The indicator is defined in the range -1 to +1 and it can be interpreted as follow for electricity demand:

- ***FF* = -1**: Electricity demand only in time intervals with high load or price.
- **-1 < *FF* < 0**: Electricity demand predominantly in time intervals with high load or price.
- ***FF* = 0**: Electricity demand equally in time intervals with high load or price.
- **0 < *FF* < 1** Electricity demand predominantly in time intervals with low load or price.
- ***FF* = 1**: Electricity demand only in time intervals with low load or price.

More specifically, the quantity chosen for the calculation in this research is the electricity demand of the heat pump, computed as the integral of power in time, as shown in the formula hereafter.

$$FF = \frac{\int_{load,low} P_{el,HP} dt - \int_{load,high} P_{el,HP} dt}{\int_{load,low} P_{el,HP} dt + \int_{load,high} P_{el,HP}\, dt} \tag{4}$$

Where:
$P_{el,HP}$ is the heat pump power demand.

Flexibility factor is a type of indicator that doesn't require a baseline to be computed. Conversely, the calculation of power demand reduction requires a baseline reference load

profile. The issue of establishing a suitable baseline was previously discussed in Section 2, which underscored the importance of a sufficient quantity of data to train models for counterfactual analysis.
Considering the limited amount of data available at the initial stages of the trial, insufficient to establish a robust baseline, the intervention and control approach [75] was selected to analyse the energy flexibility achievable at the cluster level while testing the flexibility mechanism for different heat pumps and buildings in comparable conditions.
The treatment group represents a subset of the intervention group for the reasons explained hereafter. In order to implement the flexibility actions (the treatment applied to the intervention group), it was decided that they would only be applied when the indoor air temperature is above a certain limit threshold not to compromise comfort excessively. The choice during this field trial was to set a lower temperature boundary of 18 ºC.
However, in a number of cases the temperature logged prior to any flexibility interventions was 18 ºC or lower and thus the treatment, i.e. the flexibility mechanism tested, was not applied. The characteristic of intervention, treatment and control groups are summarised in *Table 4* hereafter.

Table 4: Intervention, treatment and control group characteristics

Group	**Description**
Intervention	Group of households (participants) selected to test the energy flexibility mechanisms during the field trial.
Treatment	Subset of the intervention group to which the flexibility mechanism is applied. Provided that the indoor air temperature is higher than a set threshold (18 ºC in this trial) not to compromise comfort excessively.
Control	Group of households (participants) used for the sake of providing a comparative performance, due to similar characteristics and conditions during the field trial.

Additional details regarding the subvision of customers between intervention and control groups and the mechanism for overriding third-party control are given later in Section 4. The proper execution of the flexibility actions and the underlying mechanisms during the event have been monitored at the individual building level during the trial to characterize their effectiveness at the individual building level. The following two criteria were employed to identify operational anomalies:

1. If mean power during the event > 0.35 kW, then anomaly.
2. Count the number of interval within 10 minutes of start and end time of event when power values are > 0.1 kW. If greater than 5, then anomaly.

The operation was deemed effective at the specific building only when neither of the two conditions were present, with the second condition being the most restrictive. Notably, the event could lead to a significant decrease in the mean power demand at the aggregate level, even in cases where anomalies are present in the operation of individual heat pumps. Alternative methods for assessing the event's effectiveness at the aggregate level are discussed earlier in Section 2, because tracking of individual heat pumps operational profiles may not be easily applicable at scale. The mean power demand reduction (treatment effect) for the group of heat pumps and buildings monitored was then calculated to determine the effect of flexibility at the aggregate level (cluster of buildings). Due to the limited sample size, the statistical estimates (mean, median and confidence intervals) were recomputed using bootstrapping method with 1000 samples.

4 Field trial characteristics

This section outlines the key characteristics of the field trial. Section 4.1 details the buildings and user characteristics, providing essential context for the trial by looking at both buildings' features and user's specific traits. Section 4.2 describes the characteristics of the flexibility mechanisms implemented and the schedules of the events, offering a an overview of the trial's operational framework.

4.1 Buildings and user characteristics and enrolment

The field trial started on 10th of January 2023 and ended on 29th of March 2023, the number of participants was not fixed at the beginning and growing during the trial; there were 30 buildings in the intervention group at the end and 30 in the control group. The dwellings of participants in both groups had space heating provided by air-source heat pumps, recently installed by the project's industrial partner. Participants in the intervention group were recruited starting in August 2022. Recruitment of the control group began later, during January 2023. The participants were recruited from the customer base of the industry project partner. They were either an existing customer who had an installed ASHP serviced or a customer who was in the process of installing an ASHP with an invitation to participate provided during the quotation process. All prospective trial participants were approached by the industrial project partner, where the aims, objectives, and incentives (free servicing in this case) of the project were described. Participants were initially provided with the option of being within the intervention group and those that turned down the opportunity were then asked whether they were willing to share building monitoring data for the purposes of research. The latter became the control group. Those that consented to actively participate in the field trial were required to complete a background survey (conducted post-installation for customers in the process of installing an heat pump), providing detailed demographic, household and thermal preference data. All the installed heat pump systems in the sample were designed, supplied and installed by the research project's industrial partner using heat pumps from a single manufacturer. The installations should therefore exhibit similar levels of quality and heat pump performance. The intervention group, composing 30 buildings by the end of the trial, was predominantly spread across Southern England as shown in *Figure 3*.

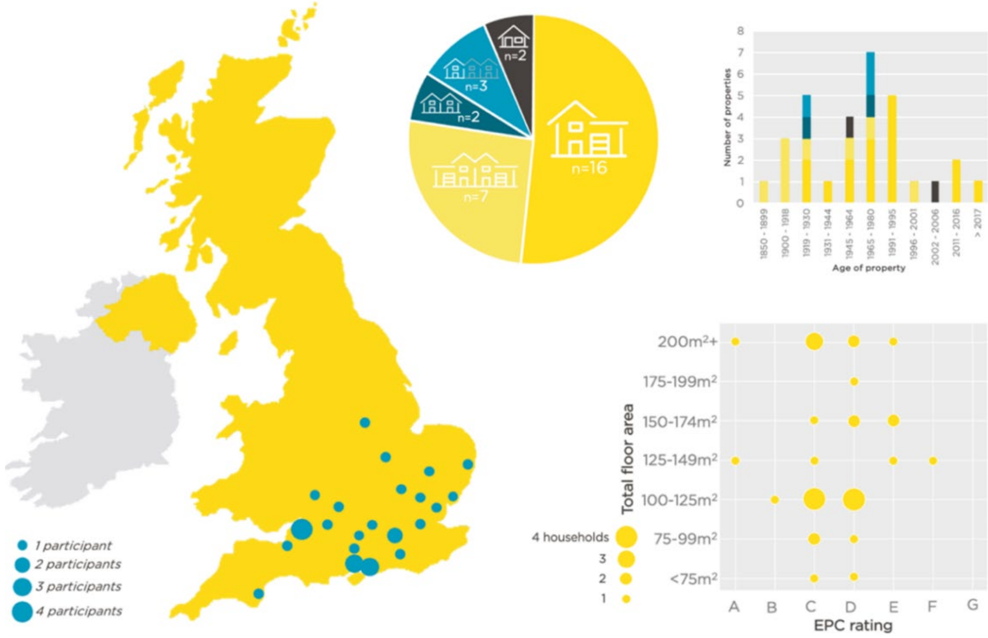

Figure 3: Geographical distribution of intervention group (left) and dwelling type distribution by typology, property age, floor area and EPC rating (right)

Across the pool of participants there was a spread of dwelling types, sizes, age and EPC ratings. While the mean floor area of participating dwelling was 141 m^2, considerably larger than the UK average of 97 m^2 or 111 m^2 for owner occupied properties [97], only 10% of households had an EPC of A/B with the majority rated C/D (75%). Added to this, the age distribution of the households was broadly representative of the English housing stock [97] with 33% built before 1945 and only 13% built post 2002. Detached properties were found to be the most prevalent however there were still a notable percentage of semi-detached and terraced properties (26% and 17% respectively). Participants are affluent and educated to a higher level than the general population and almost exclusively homeowners aged 30-64 as shown in *Table* 5 where intervention group characteristics are compared with statistics from the Office for National Statistics [98] and the Department for Work and Pensions [99].

Table 5: Intervention group characteristics compared to UK national statistics

Characteristics	Range	UK (%)	Intervention group (%)	Participants
Age	<18yrs	20.7	0	0
	18-29	14.3	0	0
	30-49	26.4	53	16
	50-64	19.5	40	12
	65-74	9.9	3	1
	>75	8.5	3	1
Household income	<£20,000	26	3	1
	£20,000-£39,999	33	7	2
	£40,000-£59,999	17	23	7
	>£60,000	24	67	20
Main occupation	Employed	47.7	70	21
	Self-employed	9.7	1	2
	Inactive	39.1	23	7
	Unemployed	3.5	0	0
Highest qualification	No qualification	18.2	0	0
	O levels/GCSEs[1] (any grade)	9.6	7	2
	5+ O levels/GCSEs[1] (A*-C)	13.4	0	0
	Apprenticeship	5.3	3	1
	2+ A levels/4+ As levels	16.9	3	1
	Degree or higher degree	33.8	80	24
	Other	2.8	7	2
Homeowner	Yes	62.5	97	29
	No	37.5	3	1
Type of household	One person	33.3	7	2
	Couple, no dependent children	34.8	30	9
	Couple with dependent children	21.8	43	13
	Lone parent with dependent children	6.2	7	2
	Other multi-person household	3.8	13	4
Accommodation type	Detached property (house/flat)	22.9	57	17
	Semi-detached property	31.5	26	8
	Terraced property	23.8	17	5
	Flat, maisonette or apartment	22.2	0	0
	Caravan or other mobile structure	0.4	0	0

[1] In the UK GCSEs and O levels are typically taken at age 16 with A levels the qualification that precedes University, typically taken at age 18.

4.2 Energy flexibility events' characteristics, communication and schedule

Two potential mechanism were considered during field testing, *"power limitation"* and *"call to heat",* whose essential characteristics are reported in *Table* 6 and are based on the considerations reported earlier in *Table 3*. In *"power limitation"* mechanism the heat pump power demand is constrained to around 60 % of full power. In *"call to heat"* the space heating operation of the heat pump was turned off. Another mechanism, named *"call to heat – DHW off"* has been included during the trial because of an unexpected behaviour encountered in some of the heat pumps, which activated DHW production when heating was turned off by the third-party control. The control strategies implemented are rule-based (RBC) and temporary operating parameters are communicated to each heat pump in the treatment group simultaneously via the cloud-based data platform.

Table 6: Energy flexibility events types and description

Type	Event mechanism	Control type	Description
1	*Power limitation*	RBC	Heat pump power reduced to 60% of full power
2	*Call to heat*	RBC	Space heating function of the heat pump system turned off
3	*Call to heat - DHW off*	RBC	Space heating and DHW functions of the heat pump system turned off

The research involved human participants (customers) and has received ethical approval (FEPS/70136). The project aimed to deliver a seamless unconscious experience of flexibility events from a customer point of view. Implementing events that can be tolerated by customers, in principle by not activating any flexibility measure for customers which are already at a temperature lower than 18 ºC as indicated earlier. The events were scheduled in the late afternoon/early evening, representing the peak conditions for the UK grid, as discussed in Section 2 and in the introduction. In terms of communication, people in the intervention group were told that heat flexibility events could take place between 4-9 pm on weekdays but didn't know how many or when events were taking place. By rule, there was a maximum of 1 event per day scheduled (typically 2 per week) and no flexibility events during weekends (Saturdays and Sundays). Participants in the intervention group were allowed to override the third-party control at any time using a dedicated webpage, accessible with a QR code, with a single field online form where they could enter their registered email address to request an override. This override could be requested preventively and during the specific day no event would be run for that customer. The override system was integrated within the heat pump control system and when an override request was submitted, the household would be removed from the treatment group and return to normal operation. Participants in the control group simply agreed to share building monitoring data (indoor temperature and ASHP power) for the purpose of research.

The testing schedule is designed as an iterative and incremental process, where data from the flexibility events are analysed quickly and the results are used to update the control strategies in subsequent events. The schedule of flexibility events deployed is shown in *Figure 4*. The "*power limitation"* mechanism is indicated by blue horizontal bars, while the "*call to heat"* and the *"call to heat – DHW off"* method are indicated by green bars and red bars respectively. Grey bars indicate where pre-heating of hot water was also implemented contextually to the application.

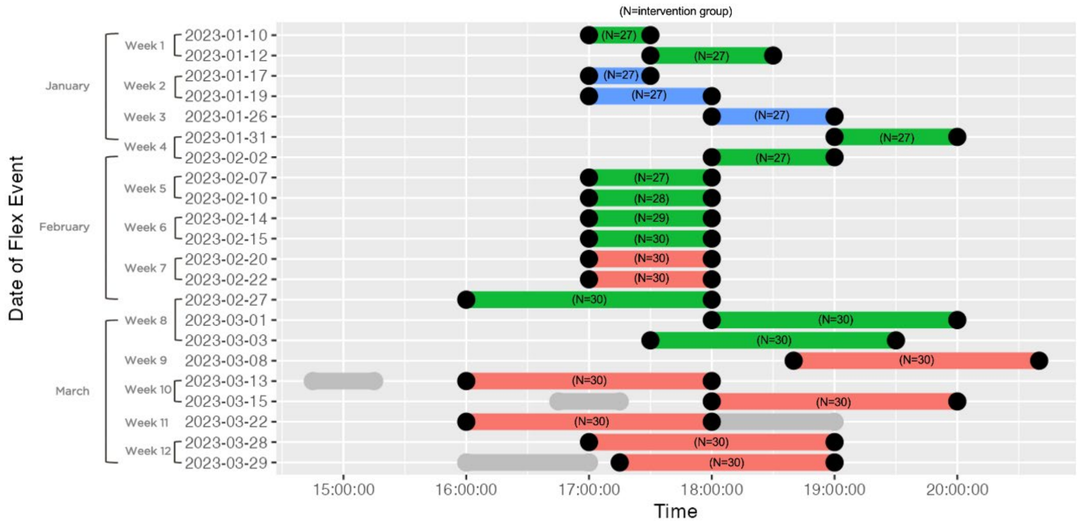

Figure 4: Energy flexibility events schedule showing flexibility mechanisms deployed. Blue bars indicate "power limitation" mechanism, green bars indicate "call to heat" mechanism, red bars indicate "call to heat – DHW off" mechanism.

5 Results and discussion

In this section experimental results collected during the field trial are presented and analysed. Section 5.1 discusses the testing of the mechanisms listed in *Table* 6 during events scheduled in *Figure 4*, highlighting the power reduction achieved by the treatment group compared to the control group. Section 5.2 examines the indoor air temperature drop, comparing again treatment and control groups during the events. Section 5.3 explores the temperature dependence or independence of the energy flexibility mechanisms reported in *Table* 6. Section 5.4 reports and analyses the number of override requests of the third-party control strategy made by users during event days. Finally, Section 5.5 summarizes the trial findings and suggests potential areas for further research, emphasizing the elements that will be considered in the future phases of the project.

5.1 Energy flexibility events mechanisms testing

The effectiveness of heat demand deferral to temporarily deliver reduced power demand at peak network times is measured as the difference between the mean power demand in each trial group. *Table 7* shows the mean heating power across all heat pumps in each trial group for each energy flexibility event run during the field trial (n=22 flexibility events, as reported in *Figure 4*). *Table 7* also reports the difference between the mean power of the treatment and control groups, which is the result of the flexibility event. The quantity reported in table is $\Delta P_{el,HP,flex}$, indicating the power reduction achieved during the flexibility event. The average outdoor air temperature for the day of the energy flexibility event and throughout the event is shown as well as the difference between the treatment and control group. The reporting of temperature is necessary due to its significant influence on performance, particularly regarding building thermal demand and coefficient of performance (COP) of ASHPs, as discussed in Section 2. Additionally, temperature reporting allows for a meaningful comparison between the treatment and control groups by highlighting their proximity in terms of average conditions (respectively "Group temp. daily diff." and "Group temp. event diff." in *Table 7*). On average, the heating power difference among groups, depending on the flexibility mechanism put in place, ranges from maximum of 1.581 kW (07/02/2023, daily average ambient temperature 1.6 ºC, event average ambient temperature 4.2 ºC) to a minimum of

0.239 kW (22/03/2023, daily average ambient temperature 11.0 °C, event average ambient temperature 11.7 °C). In percentage terms, the power reduction achieved with the *"power limitation"* mechanism is the range 21.5-42.9% (average 30.6%), while the "*call to heat*" and "*call to heat – DHW off*" mechanisms achieved respectively percentages in the range 68.6-97.3% (average 88.6%) and in the range 67.8-96.6% (average 88.0%). Due to the nature of the last two mechanisms, a power reduction of 100% (heat pump turn-off) would be theoretically possible; what was found empirically was that even if the events presented anomalies at the single heat pump level (detected according to the criteria reported in Section 3), it was possible to achieve in both cases a significant reduction in percentage terms.

Table 7: Mean heating power by group and mean ambient temperature, showing difference between control and treatment groups,for each heat flex event. Negative group differences indicate values for treatment group are lower than control.

N.	Date	Type	Control group power $P_{el,hp,ref}$	Treatment group power $P_{el,hp,flex}$	Group power diff. $\Delta P_{el,hp,flex}$	Group power diff. $\Delta P_{el,hp,flex\%}$	Mean temp. daily	Group temp. daily diff.	Mean temp. event	Group temp. event diff.
			kW	kW	kW		°C	°C	°C	°C
1	2023-01-10	2	0.686	0.067	0.619	90.2	8.8	-0.3	12.1	0.2
2	2023-01-12	2	0.880	0.283	0.597	67.8	10.1	0.0	9.0	0.0
3	2023-01-17	1	1.780	1.398	0.382	21.5	-2.0	0.2	-0.8	-0.1
4	2023-01-19	1	2.002	1.144	0.859	42.9	0.5	-0.1	1.2	-0.4
5	2023-01-26	1	1.550	1.123	0.427	27.5	4.5	0.0	5.4	-0.1
6	2023-01-31	2	1.233	0.042	1.192	96.6	6.8	0.0	5.6	0.0
7	2023-02-02	2	0.815	0.167	0.648	79.5	8.6	0.0	9.4	0.0
8	2023-02-07	2	1.704	0.124	1.581	92.7	1.6	0.0	4.2	-0.1
9	2023-02-10	2	1.252	0.074	1.178	94.1	4.0	-0.2	7.5	0.0
10	2023-02-14	2	0.888	0.069	0.819	92.2	5.3	0.0	8.2	0.2
11	2023-02-15	2	0.729	0.081	0.648	88.9	7.1	0.0	10.0	0.4
12	2023-02-20	3	0.432	0.027	0.405	93.8	8.9	0.4	10.3	0.5
13	2023-02-22	3	0.864	0.066	0.798	92.4	6.6	-0.2	7.2	-0.2
14	2023-02-27	2	1.180	0.073	1.107	93.8	3.4	-0.1	6.3	0.0
15	2023-03-01	2	1.155	0.116	1.039	90.0	5.0	0.2	5.3	0.0
16	2023-03-03	2	1.351	0.243	1.108	82.0	4.1	-0.2	5.0	0.0
17	2023-03-08	3	1.615	0.043	1.572	97.3	1.3	-0.6	1.6	-1.0
18	2023-03-13	3	0.405	0.025	0.380	93.8	11.3	0.2	11.4	0.3
19	2023-03-15	3	1.084	0.109	0.976	89.9	4.2	-0.2	6.6	0.0
20	2023-03-22	3	0.350	0.110	0.239	68.6	11.0	0.2	11.7	0.5
21	2023-03-28	3	0.777	0.129	0.648	83.4	7.0	-0.2	7.9	0.2
22	2023-03-29	3	0.401	0.042	0.359	89.5	10.7	0.0	12.2	0.3

Flexibility factor (*FF*) is calculated in the trial period for flexibility event days and non event days to understand how much individual heat pumps operate in periods of "low" or "high load" in the grid (a 4 hour interval 16:00-20:00 is used, for the reasons explained in Section 3). The distribution of *FF* values found during the field trial is reported in *Figure 5* for the treatment group (subset of the intervention group, following the definitions reported in *Table 4*). It can be seen in the figure how heat pumps, on non-flex days, typically operated in hours of "low load" (20 hours per day) with a median of 0.59 and mean of 0.57. The distributions does however include a left tail of values, that are near 0 and lower than 0, where flexibility becomes particularly relevant. On flexibility event days, the distribution is shown to shift somewhat to the right (towards 1), with a median of 0.64 and a mean of 0.60. This shift in distribution to the right indicates the effect of energy flexibity actions, with a reduction of heat pump demand during "high load" hours (16:00-20:00 during weekdays).

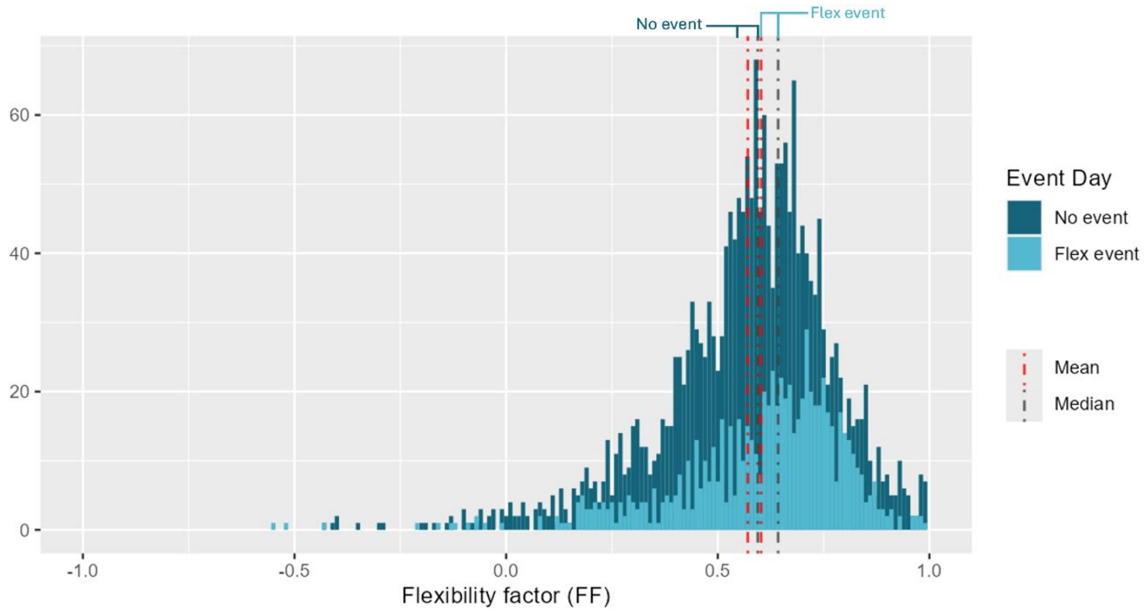

Figure 5: Distribution of Flexibility factor (FF) computed for individual buildings in flexibility event days and non-event days during the trial period.

In summary, the findings from the 22 events conducted indicate that the heat pumps may be effectively controlled remotely during flexibility events by a third-party. However, in certain instances, there is still a small power demand, although it should ideally be zero, in the treatment group. Further, the daily average ambient temperature during the event days varies between -2.0 ºC and 11.3 ºC, while the average ambient temperature during the event ranges from -0.8 ºC to 12.2 ºC. Therefore, the results reflect a representative range of temperature conditions that air source heat pumps may encounter in Southern and Mid England. In addition, the average daily difference in ambient temperature between the treatment and control groups ranges from -0.6 ºC to 0.4 ºC. During the event, the difference between the groups ranges from 1.0 ºC to 0.5 ºC. Indeed, this demonstrates the very modest difference in temperature conditions between the groups involved in the field trial.

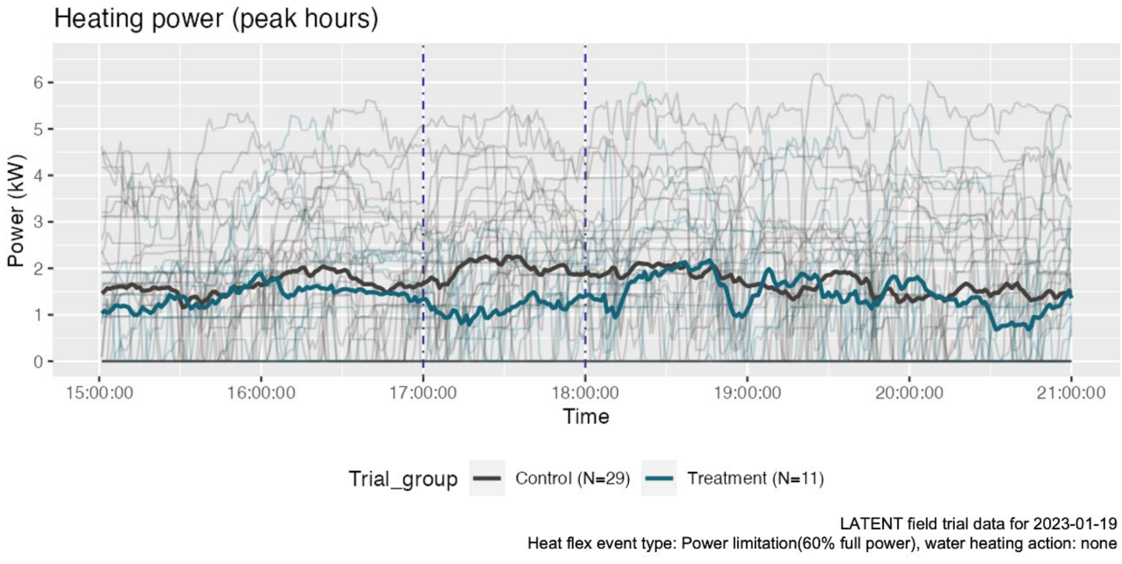

Figure 6: Heating power demand during peak hours, "power limitation" flexibility event 19th January 2023. Thin lines show power for individual households, thick lines show group mean values. The dotted lines show the start and end of the flexibility event.

In terms of energy flexibility event types, reported in *Table 6*, the "*power limitation"* mechanism was tested (these events are shown in blue in *Figure 4*, events 3-5 in *Table 7*) on 19th January 2023; as it is shown in *Figure 6*, the vertical blue dash-dot lines indicate the start and end of the flexibility event. During this event, the group power different $\Delta P_{el,HP,flex}$ was 0.859 kW. This was the most effective (in terms of kW reduction) event among the ones using "*power limitation"* mechanism although with a small number of active participants (n = 11), as indicated in *Figure 6*.

Figure 7 and *Figure 8* show heating power during two heat flexibility events lasting one hour and using the "*call-to-heat"* mechanism on the 2nd and 7th February respectively (these events are shown green in *Figure 4*, events 7-8 in *Table 7*). These events involved a greater number of treatment group participants: n = 22 and n = 20 respectively. The data from these events shows that the treatment and control groups exhibit similar mean power consumption profiles leading up to the heat flex events, although observed demand in the treatment group was slightly higher in the pre-event period on 7th February. The mean heating power demand in the treatment group reduces quickly following the start of both events and remains consistently below the power demand of the control group for the duration of the event, on average $\Delta P_{el,HP,flex}$ is 0.648 kW and 1.581 kW lower in the treatment group during the 2nd February and 7th February events respectively.

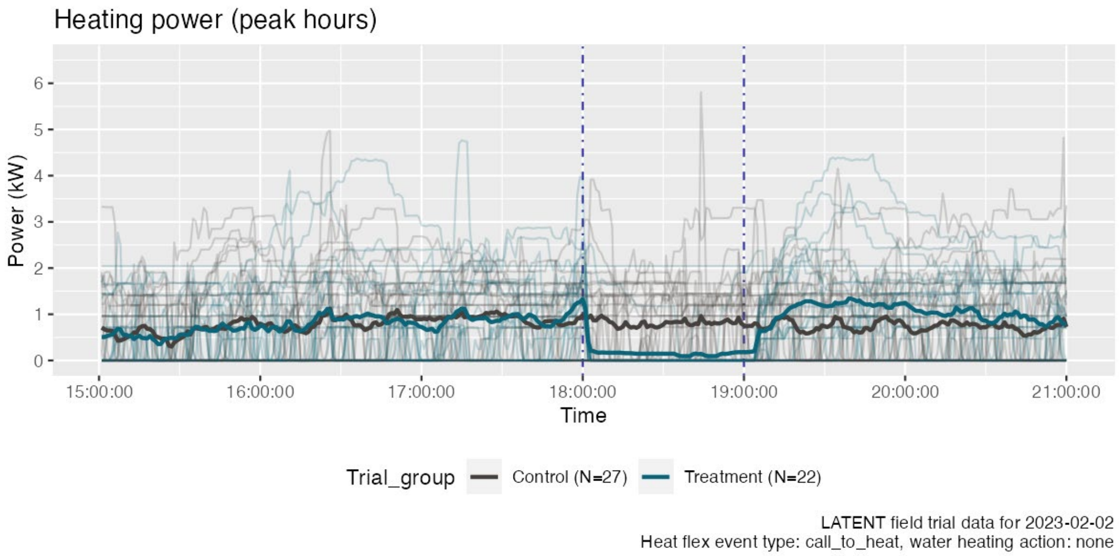

Figure 7: Heating power demand during peak hours, "call-to-heat" flexibility event 2nd February 2023. Thin lines show power for individual households, thick lines show group mean values. The dotted lines show the start and end of the flexibility event.

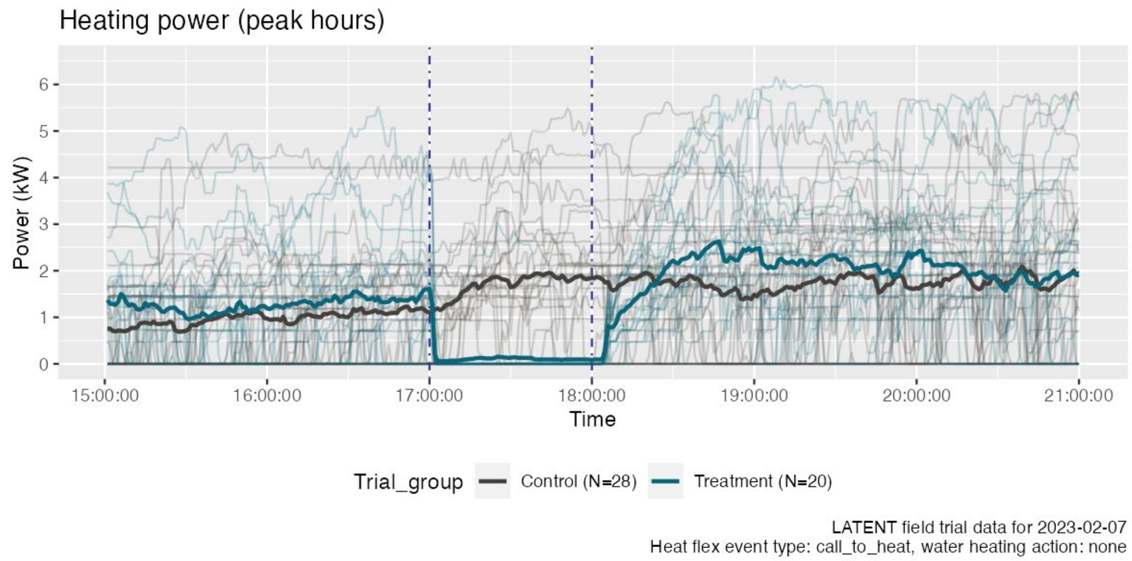

Figure 8: Heating power demand during peak hours, "call-to-heat" flexibility event 7th February 2023. Thin lines show power for individual households, thick lines show group mean values. The dotted lines show the start and end of the flexibility event.

Higher power demand was observed in the treatment group for a period following the energy flexibility events with the ramp-up taking between around 20 and 40 minutes. This is associated with the heat pumps in this group working to return to their normal operating parameters. This "snapback" effect could be prevented by combining *"call to heat"* and *"power limitation"* in sequence and a preliminary attempt was made on 22nd March 2023 event, discussed later. Inspecting the temperature records from weather data for the heat flex event days presented in *Figure 7* and *Figure 8* respectively, the mean ambient temperature was approximately 9.5 ºC at the start of the flex event on 2nd February (*Figure 9*). The ambient conditions were significantly colder during the event on 7th February with overnight temperatures below 0 ºC and temperature at the start of the heat flex event approximately 5 ºC (*Figure 9*). As can be seen in *Figure 9*, the mean ambient temperature for the treatment and control groups shows no significant differences. In turn, this enables a meaningful comparison between the performance of the two groups; indeed, as reported earlier in *Table 7*, the small differences in ambient temperature (daily average and during the events) enables an appropriate comparison of performance across all the flexibility events in the trial. The larger reduction in load recorded during the colder of the two events reflects the higher load of the heat pumps working to maintain target temperatures at lower ambient temperature conditions and a corresponding lower COP of the ASHP.

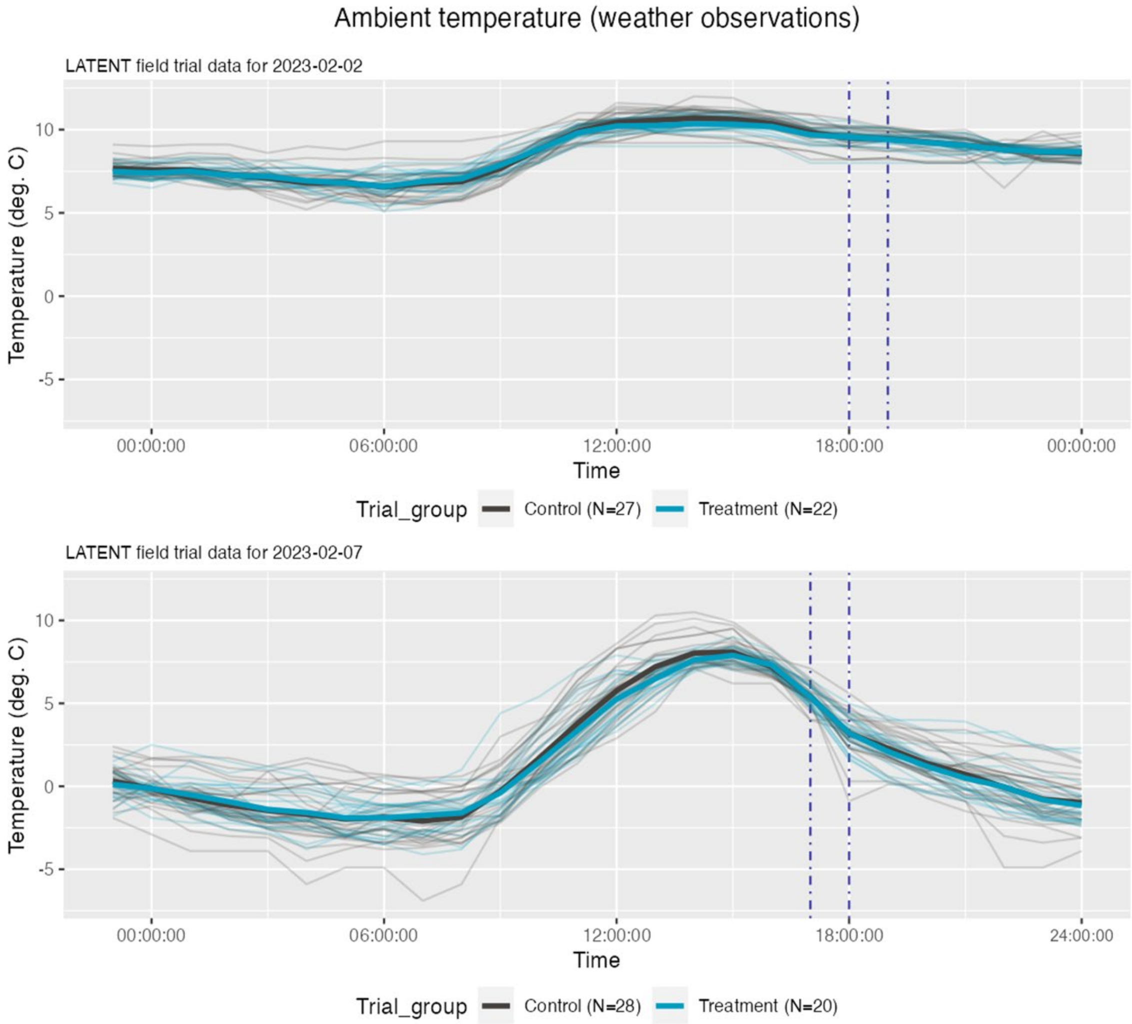

Figure 9: Ambient temperature for flexibility events on 2nd February 2023 (top) and 7th February 2023 (bottom). Thin lines show temperature for individual households, thick lines show group mean values and dotted lines show start and end of flexibility event.

Analysis of the power demand for domestic hot water (DHW) operation of the heat pumps revealed that when the heating operation was turned off using the "*call-for-heat*" mechanism, a subset of heat pumps would initiate a DHW heat cycle, which can be seen in *Figure 10* for the event on 2nd February 2023. A similar behaviour was observed on 7th February 2023 and is reported in Appendix A.1 in *Figure 15* for completeness.

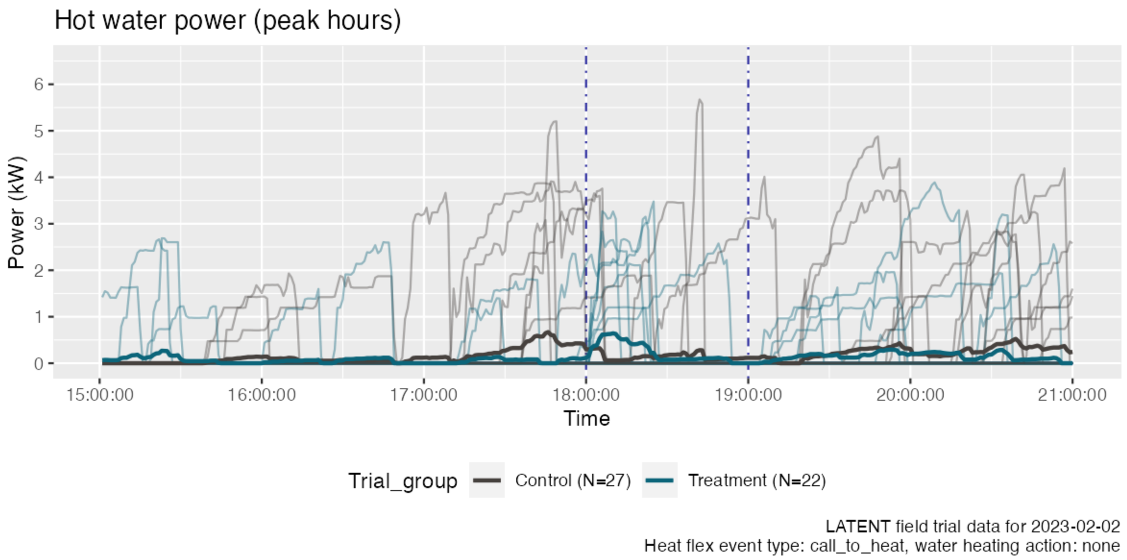

Figure 10: DHW power demand during peak hours, "call to heat" flexibility event 2nd February 2023. Thin lines show power for individual households, thick lines show group mean values. The dotted lines show the start and end of the flexibility event.

This effect can be attributed to the energy flexibility events as this behaviour is present in the treatment group and coincides with the varying start times of the heat flex events. These results also show a lower number of individual power demand traces overall compared to those found for heating power. This indicates that the operation of heat pumps for providing domestic hot water (DHW) was less likely to occur during the peak period (late afternoon/early evening during weekdays) than the operation for space heating. The initiation of a DHW cycle potentially reduces the effectiveness of the heat flex event, therefore the research team implemented changes to the heat deferral events scheduled to happen later in the field trial.

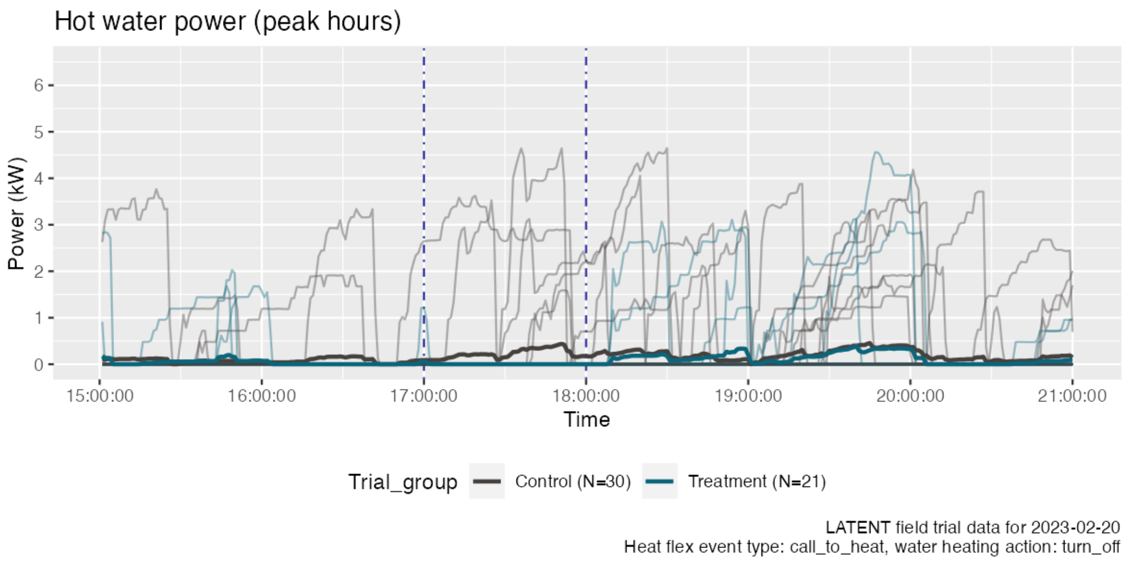

Figure 11: DHW power demand during peak hours, "call to heat - DHW off" flexibility event 20th February 2023. Thin lines show power for individual households, thick lines show group mean values. The dotted lines show the start and end of the flex event.

To preserve the load reduction obtained by turning off the space heating function (*"call to heat"* mechanism in *Table 6*), the DHW cycle was also turned off to prevent heat pumps from switching from heating to hot water cycles during the flex events (these events are shown red in *Figure 4* and described in *Table 6, "call to heat – DHW off"*). *Figure 11* shows the deferral of the hot water cycle clearly (shifted in time) during one such heat flex event between 17:00 and 18:00 on the 20th February. To visualise the impact on aggregate load, *Figure 12* shows the total load profile for both heating and hot water power demand during a flexibility event where only the heating function was deferred (7th February, top image), and an event where both heating and hot water functions were deferred (20th February, bottom image).

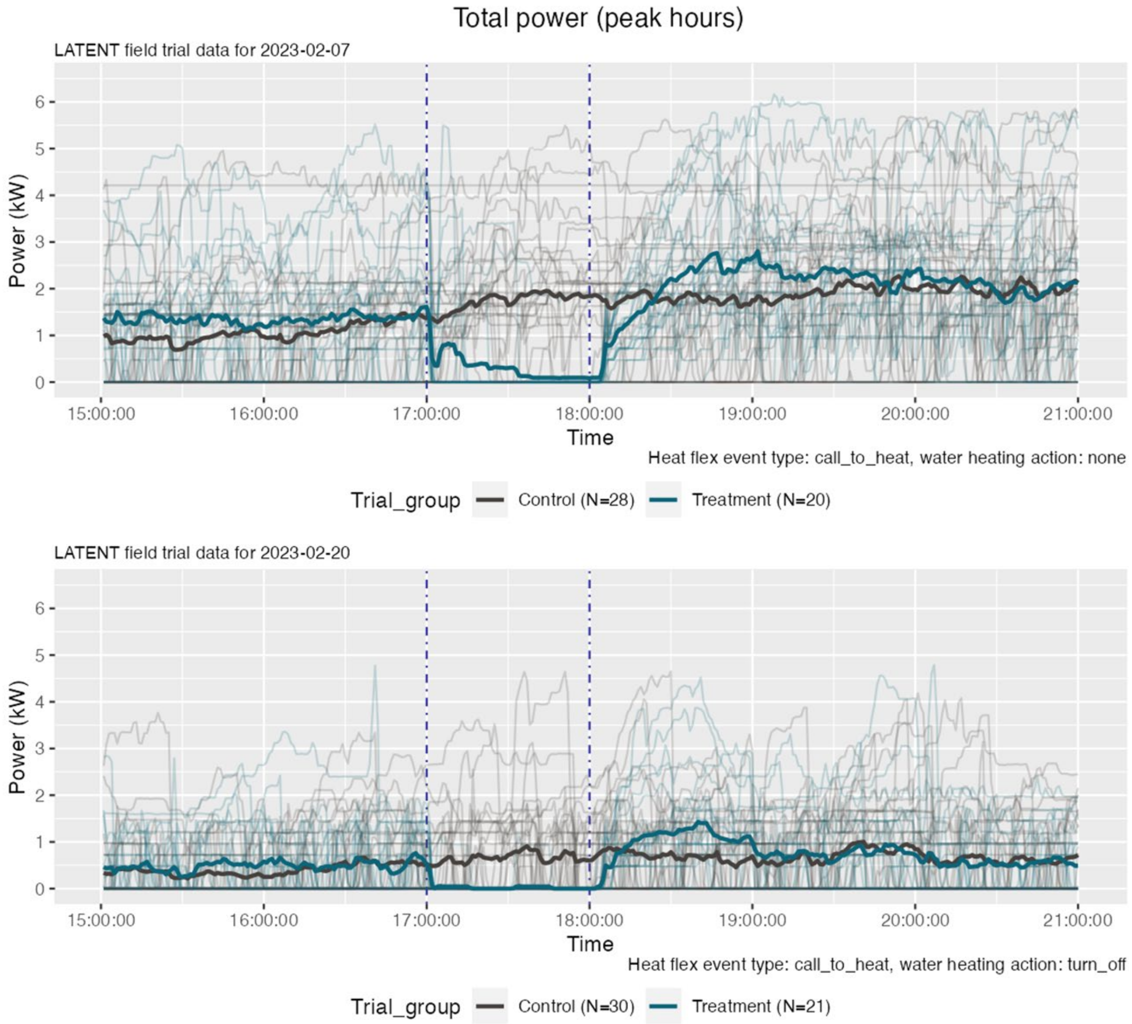

Figure 12: Total load profiles (heat and DHW) during "call to heat" flexibility event 7th February 2023 (top) and "call to heat - DHW off" flexibility event 20th February 2023 (bottom) . Thin lines show power for individual households, thick lines show group mean values. The dotted lines show the start and end of the flexibility event.

The impact of treatment group heat pumps initiating a hot water cycle is clearly observed soon after the start of the heat flex event on 7th February at 17:00; this feature is not present during the 20th February event, as shown in *Figure 12.* Unfortunately, the deferral of the hot water cycle was not reliably implemented during this field trial. On a number of occasions, the data shows that there was load for hot water operation during the events indicating that the deferral was not completely effective in reducing load to zero. The subsequent section reports the indoor temperature decrease that was observed during the flexibility event on 8th March 2023. Finally, in Appendix A.3, the outcomes of the event on 22nd March 2023, the day with the highest temperature, are presented. The event was characterised by an average daily ambient temperature of 11.0 ºC and an average ambient temperature of 11.7 ºC, leading to the smallest power reduction value $\Delta P_{el,HP,flex}$ of 0.239 kW.

5.2 Indoor air temperature change during energy flexibility events

Only a limited set of internal temperature data was collected for the control group in the field trial initially. This was due to the recruitment of the control group happening after the start of the active trial. While some back-dated data was obtained through the monitoring platform, the internal temperature and humidity sensors were installed later, thus affecting data coverage. Therefore, the data presented hereafter refer to energy flexibility events towards the end of the trial. *Figure 13* shows the measured internal temperatures for both the treatment group and the participants which were excluded because of i) an indoor air temperature lower than the threshold defined in Section 3, or ii) sending an override request pre-event. *Figure 13* also shows that the participating households have slightly higher indoor air temperatures compared to the control group, this effect persists even when data for the excluded households are combined into the treatment group. Interestingly, the data shows a wide range of internal temperatures across the sample households, indicating a significant variability in user behaviour and comfort preferences. *Figure 13* also reveals that the internal temperature of the treatment group dropped by 0.78 ºC during the flex event (minimum value was reached slighting after around 0.8 ºC), as recalled in *Table 8* later on. The mean values of the indoor temperature trajectories are reported with their confidence interval (95%) to facilitate the comprehension of the behaviour at the aggregated level, for a group of buildings. Individual indoor temperature trajectories are reported in Appendix A.2 to ensure completeness and to illustrate the spectrum of variability of operational conditions found at the individual building level. The ambient temperature graph in *Figure 13* illustrates the ambient temperature conditions, which exhibit a small temperature fluctuation throughout the day.

Finally, the bottom graph within *Figure 13* shows the change in internal temperature relative to the start of the flex event. The mean temperature in the control group rises slightly over the heat flex event while the mean temperature in the treatment group declines until approximately 20 minutes after the end of the event (indicated by the black dash-dot line in *Figure 13*), indicating a small lag of the heating system response when turned back on after the flexibility event. In *Figure 13* (bottom graph), the internal temperature change has been normalised to account for the difference between ambient (outdoor air) and indoor air temperatures recorded across households and is expressed as the change in lounge (living space) temperature per degree of temperature gradient (delta-t in charts, ΔT, difference between indoor and ambient temperature) for each household. The temperature drop observed at the end of the flexibility event on 8th March was the largest of all the events with a normalised average value of -0.045 °C per degree gradient ΔT, corresponding to an average value of -0.78 °C drop across the treatment group households relative to the start of the event. The average ambient daily temperature for the event day is 1.3 °C and the average ambient temperature during the event is 1.6 °C, representing one of the coldest days in the trial period, even though not the coldest one.

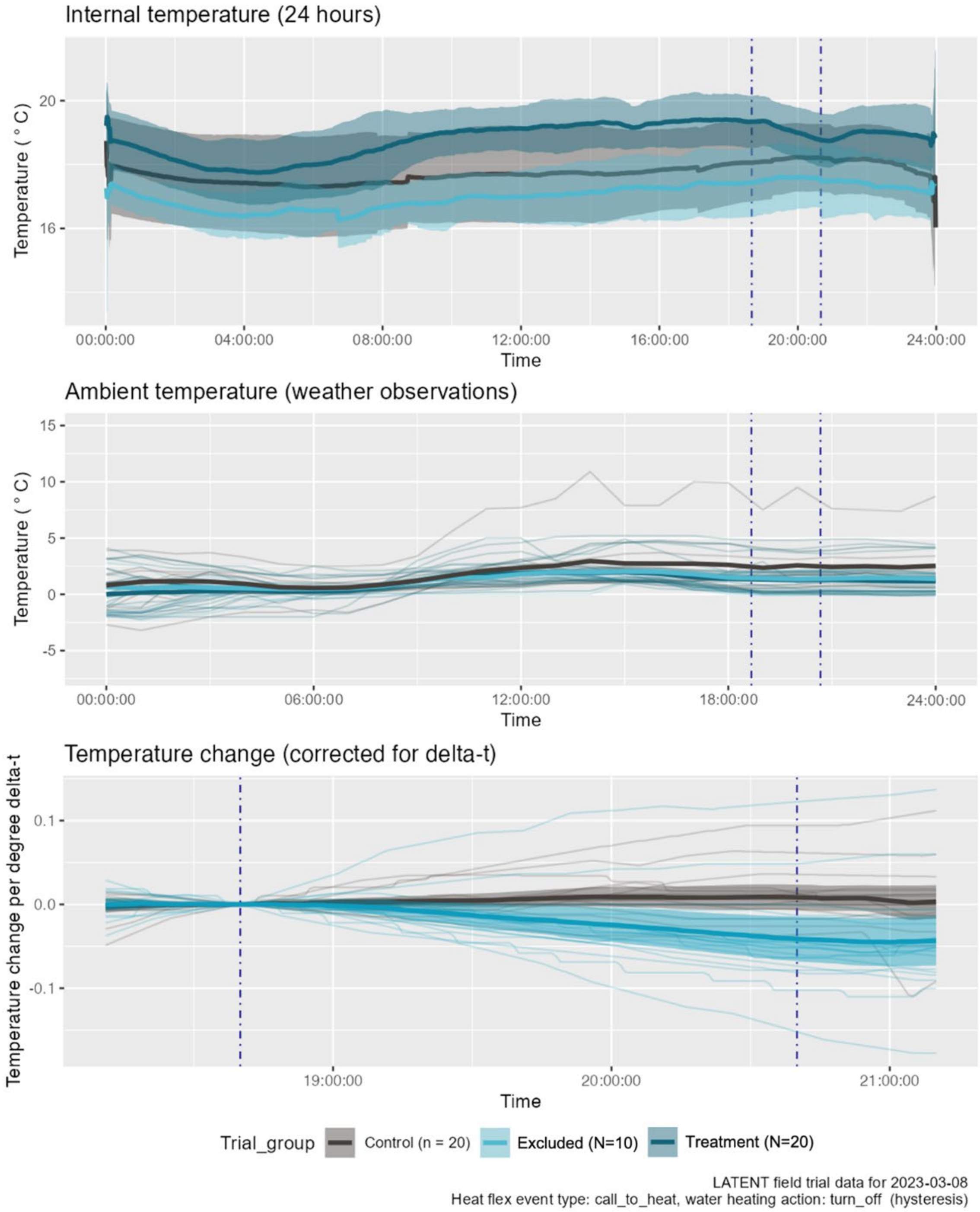

Figure 13: Measured internal temperatures (top), ambient temperature (middle), and temperature change measured in living space, normalised for internal-ambient temperature gradient (bottom) for trial households during a "call to heat - DHW off" flexibility event 8th March 2023. Thick lines show group mean values and areas indicate the 95% confidence interval of temperatures in the corresponding trial groups.

Table 8 reports the temperature change and the normalised temperature change for each flexibility event conducted during the trial. The values are all negative (therefore indicating a temperature drop), except in one case where a very small temperature increase is recorded, equal to 0.03 °C. The accuracy of measurement specified by the manufacturer of sensors (0.01 °C), reported in *Table 2* in Section 3, is used for the table to highlight even small difference; however a more conservative approach would be rounding the measured quantities to one decimal. Further, it should be noted that the moment of the day when events are running are characterised by high internal gains in the lounge (living space) due to the presence of people and appliances. This has clearly a positive impact, as it helps in reducing the temperature drop during the flexibility events.

Table 8: Temperature change and timing of minimum for flexibility events during the trial period (n=22).

N.	Date	Type	Temp. change	Normalised temp. change (per *ΔT*)	Start time	End time	Timing of min. temperature
			°C	-	-	-	-
1	2023-01-10	2	-0.02	-0.004	17:00	17:30	17:38
2	2023-01-12	2	0.03	0.003	17:30	18:30	18:59
3	2023-01-17	1	-0.08	-0.004	17:00	17:30	17:47
4	2023-01-19	1	-0.27	-0.015	17:00	18:00	18:06
5	2023-01-26	1	-0.13	-0.009	18:00	19:00	19:17
6	2023-01-31	2	-0.43	-0.030	19:00	20:00	20:39
7	2023-02-02	2	-0.05	-0.006	18:00	19:00	19:06
8	2023-02-07	2	-0.35	-0.021	17:00	18:00	18:23
9	2023-02-10	2	-0.26	-0.022	17:00	18:00	18:28
10	2023-02-14	2	-0.30	-0.022	17:00	18:00	18:35
11	2023-02-15	2	-0.06	-0.006	17:00	18:00	18:01
12	2023-02-20	3	0.23	0.014	17:00	18:00	18:01
13	2023-02-22	3	-0.25	-0.020	17:00	18:00	18:18
14	2023-02-27	2	-0.46	-0.034	16:00	18:00	18:13
15	2023-03-01	2	-0.59	-0.040	18:00	20:00	20:29
16	2023-03-03	2	-0.35	-0.025	17:30	19:30	19:36
17	**2023-03-08**	**3**	**-0.78**	**-0.045**	**18:40**	**20:40**	**21:00**
18	2023-03-13	3	-0.24	-0.027	16:00	18:00	18:16
19	2023-03-15	3	-0.54	-0.043	18:00	20:00	20:19
20	2023-03-22	3	-0.06	-0.009	16:00	18:00	18:05
21	2023-03-28	3	-0.50	-0.043	17:00	19:00	19:08
22	2023-03-29	3	-0.27	-0.033	17:15	19:00	19:08

5.3 Ambient temperature dependence of energy flexibility outcomes

Following the argumentation reported in Section 2, outdoor air temperature is essential to characterise load profiles when heating demand is present. Ambient temperature influences both the thermal demand of the building and the Coefficient of Performance (COP) of Air Source Heat Pumps (ASHP), leading to a higher power demand at lower temperatures. Therefore, if flexibility events are effective in reducing demand by turning-off heating system for the requested amount of time, higher power reductions would be expected to be achieved during colder days and when outdoor air temperature during the event are lower. *Figure* 14 shows that the difference in power demand of the treatment and control groups observed during the 22 energy flexibility events of this trial does indeed support this hypothesis and increase as the mean ambient temperature during the event decreases. The same figure also shows that the flex events using the *"power-limitation"* mechanism clearly do not achieve as large reductions as the *"call-to-heat"* mechanism, when the system it turned-off completely. For the *"call-to-heat"* mechanism, divergence of the two types of events (*DHW on* and *DHW off*) would be expected; however the results from these two sets of events are very similar as the unexpected behaviour described in Section 5.1 didn't happen for all the heat pumps monitored.
In brief, *"power limitation"* mechanism is nearly independent on outdoor air temperature (i.e. the third-party control modulates power reduction up to 60% of full power and this can happen independently on outdoor temperature), while *"call to heat"* mechanisms are intrinsically temperature dependent because of building thermal demand and COP of heat pump. In turn, these are intrinsically dependent on physical characteristics of the building

that will be explored more in detail in the future to enable a better characterisation of the building stock for predictive purpose, in combination with the gradient of indoor air temperature decay, which clearly impacts user comfort.

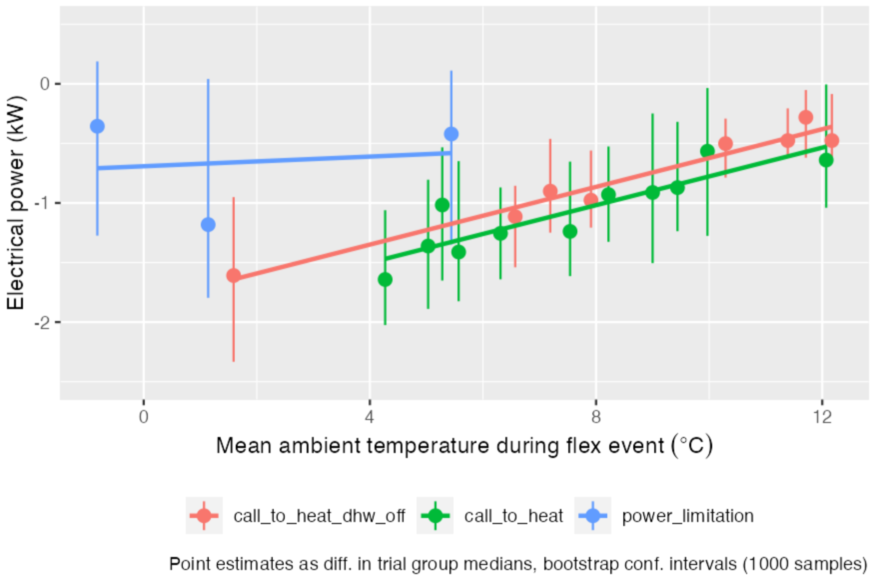

Figure 14: Scatter plot showing difference in group average total power demand during flex event against the average ambient temperature for all energy flexibility events in the trial, grouped by colour

At this stage, linear regression models of the power reduction achieved (on average as indicated in *Figure 14*) for the cluster of buildings have been fitted using the mean ambient temperature (during the flexibility event) and the mean daily ambient temperature (during the event day) as input variables. The results achieved in terms of Root Mean Square Error (*RMSE*), coefficient of determination R^2 and adjusted R^2 are presented in *Table 9* and indicate a good model fit for all the event mechanisms. The results have clearly limitations due to the number of events and buildings involved in the trial. However, this kind of modelling can assist in quantifying the impact of flexibility at an aggregated level, employing formulations similar to the ones adopted for predicting electric load profiles on a UK national scale [100,101]. At the same time, these formulations are useful for analysing flexibility behaviour at the individual level [58] using counterfactuals, or at large scale using Difference in Differences (DID) approach [72], as reported in Section 2.

Table 9: Regression model of flexibility behaviour at the cluster level. Difference in group average total power demand as a function of mean ambient temperature during the flexibility event and of the mean ambient daily temperature during the event day

Type	**Event mechanism**	**Input variables**	***RMSE***	***R^2***	***Adj-R^2***
			kW	-	-
2	*Call to heat*	Mean temp. event	0.164	0.759	0.732
		Mean temp. event, Mean temp. daily	0.131	0.863	0.828
3	*Call to heat - DHW off*	Mean temp. event	0.084	0.969	0.964
		Mean temp. event, Mean temp. daily	0.090	0.970	0.958
2/3	*Call to heat and Call to heat - DHW off*	Mean temp. event	0.153	0.855	0.846
		Mean temp. event,	0.123	0.912	0.901

Mean temp. daily

5.4 Override requests received during the field trial

Across the 22 scheduled energy flexibility events, there were a total of 26 requests to override the third-party control (17 during actual event days and 9 during non event days), which correspond to a very low incidence of requests in percentage terms, as discussed later. When householders requested an override of the third-party control, their heating system was returned to normal operation, as indicated earlier in Sections 3 and 4.

In order to compute percentage of override requests, it is necessary to compute first a potential number of override requests. Following the rules of the field trial, there is a maximum of 1 heat flexibility event per day and events are not run during weekends (participants have been advised that events would have not been run during weekend). Therefore, the number of potential requests for event days is computed as the number of events days (22) multiplied by the number of participants in the intervention group (n in the range 27-30) leading to a total of 633 potential requests, reported in *Table 10.* The potential number of override request during non event days is computed with reference to the number of days during the experimental trial period (79), subtracting the event days (22) and the number of Saturdays and Sundays (weekend days, 11), again multiplied by the number of participants in the intervention group (range 27-30). In this case, the potential number of requests is 1339. The results are reported in *Table 10*, showing that 1.1 % of the total potential number of potential requests was received during the event and 1.3 % after the event (during the same day); only 0.3% of requests were received before the event in the same day. Interestingly, the 0.7% of requests were received on days when there were no events.

Table 10: Number and proportion of override requests by timing of request

Request received	Event day	Non-event day	Event day: Before	Event day: During	Event day: After
Potential override requests	633	1339	633	633-2	633-(2+7)
Number of requests received	17	9	2	7	8
Percentage of request over potential override requests (%)	**2.7**	**0.7**	**0.3**	**1.1**	**1.3**

The requests received before events and during non event days are likely due to users feeling uncomfortably cold, even in normal operating conditions, or thinking ahead and pre-emptively over riding to ensure comfortable conditions are maintained. These behavioural aspects will be investigated in future research. The 1.2 % of requests received after the event (in the same day) is also interesting as a number of request occurred more than 2 hours after the end of the heat flex event. This indicates that the impact of energy flexibility with heat deferral may be perceived by some participants well after the event and during periods where the heating is running in normal conditions.

The number of override requests by event duration is reported in *Table 11***.** The 2 hour flexibility events are the ones with the highest proportion of requests, 4.4%. In this case the percentage is computed based on the number of days by event duration. The small percentage of requests received indicates that it is possible to extend the duration of events in future field trials, even though this may require a more careful consideration of the building characteristics and user traits.

Table 11: Number of override requests by event duration

Event duration	Number of requests	Number of events (days)	Potential override requests	Percentage of requests over potential requests (%)
30 mins	2	2	54	**3.7**
60 mins	3	11	309	**1.0**
120 mins	12	9	270	**4.4**
Non event	9	-	1339	**0.7**

Finally, due to the small amount of events run, it is impossible to derive a reliable correlation between the number of override requests and the event duration, considering also the impact of ambient temperature, discussed previously. This is also a relevant element to be considered in future research.

5.5 Summary of field trial findings, lessons learned and future work

The results achieved during the field trial are summarised to indicate the most relevant findings at this stage and inform further research developments. In particular, the energy flexibility mechanisms that were implemented and tested have shown that it is feasible to implement strategies that involve both *"power limitation"* (i.e., where the third-party constrains the thermal power output) and a complete turn-off of the heat pump, named *"call to heat"*. The third-party's flexibility intervention mechanisms at this stage of research were based on a Rule-Based Control strategy (RBC); however, more advanced options, such as Model Predictive Control (MPC) strategies, could be considered in the future whether they could provide a significant benefit in terms of duration of the event, for example by slightly increasing indoor temperatures during non-peak hours before the events and/or by charging DHW storage in non-peak hours (i.e. load shifting and valley filling strategies). Further, the field trial gave opportunity to identify some unexpected behaviour in the control of ASHP operation. For instance, heat pumps activated DHW production when heating was turned off, necessitating the development of an alternative mechanism in which both heating and DHW were turned off, name *"call to heat – DHW off"*. The collected data also indicated a significant increase in power demand following energy flexibility events, commonly referred to as the *"snapback"* effect. This effect can be limited by sequentially combining two mechanisms: *"call to heat"* followed by *"power limitation"*. The combined mechanism was tested preliminarily during the 20th event on 22nd March 2023. Overall, while operational anomalies were detected at the individual heat pump level during events using the criteria reported in Section 3, the results at the aggregate level were still significant (88.2% power reduction for *"call to heat"* and *"call to heat – DHW off"*), due to the non coincidence of anomalies in heat pump response.

From the user perspective, the number of override requests received during the event was 1.1 % of the total potential number of requests, and 1.2 % after the event (0.3% before the event), during the same day. It is noteworthy that 0.7 % of potential requests were received on days when there were no events. This issue will be explored further using upcoming surveys and semi-structured interviews with trial participants. In terms of events' duration, the 2 hour flexibility events received the highest proportion of override requests, 4.4%, a percentage significantly higher than the one computed at the aggregated level. Nonetheless, this percentage is still relatively small, suggesting that there is potential for extending the duration of events in future field trials, as well as verifying if the small amount of requests is consistent with other types of user traits and buildings'characteristics. Some of the participants in the intervention group were excluded from the treatment group due to the fact that indoor air temperatures were significantly lower than anticipated (below 18 ºC, the threshold for participation in the flexibility event set during this trial). Users were contacted afterwards and confirmed the fact that in some cases they kept indoor temperatures which were lower than the comfort requirements normally assumed. This highlights the significance of having sufficient internal temperature monitoring as a safeguarding mechanism and the challenge of creating models that are customised for individual users and enhances the appeal of modelling clusters of buildings and communities to understand their aggregated behaviour, rather than the individual one. Insights are summarised hereafter in *Table 12* indicating also potential future directions for research and implementation.

Table 12: Main research finding and future research potential

Issue	Findings	Future research potential
Energy flexibility mechanisms testing	3 rule-based control (RBC) mechanisms are tested *"power limitation"*, *"call to heat"*, *"call to heat - DHW off"*, the latter developed based on unexpected behaviour of HPs and combined with pre-heating of DHW.	Control strategies can be improved, e.g. predictive pre-heating of building and DHW storage, but they have to consider the problem of simplicity and scalability to be effectively deployed.
Roll-out, duration and timing of flexibility events	Events with different durations have been tested in the late afternoon/early evening periods with duration of 30 min, 1 h and 2 h. A small amount of override requests has been received (1.1% during the event, 1.3% after the event).	More in depth analysis of indoor temperature drops and temperature dependence of energy flexibility achievable in relation to buildings' and user characteristics. A better characterisation can provide additional information that could enable successful flexibility in longer events.
User behaviour and comfort requirements	The number of third-party control override requests received is small number, indicating a seamless experience for participants.	More insights on the user perception and understanding of the overriding mechanism can be collected. Better characterisation of comfort preferences based on customer traits is possible.

The research results obtained during this trial align well with the fundamental research questions set in the project. First, the effectiveness of third-party control was demonstrated by the significant reduction in power demand during the events, with only a modest drop in indoor temperature. Second, the study quantified the flexible load achievable for a cluster of buildings, highlighting how the aggregated average power reduction achieved, correlated with outdoor air temperature, is substantial. Third, customer response to energy flexibility events was positive, as evidenced by the small number of override requests, suggesting that users did not frequently notice when energy flexibility mechanisms were in operation. Lastly, regarding the duration of flexibility events tolerable by customers, the longer events tested lasted for up to 2 hours. Although these longer events had a slightly higher percentage of override requests, the overall number remained small, indicating the potential for even longer events in future trials.

6 Conclusions

Within the broader context of accelerated decarbonisation of both the electric grid and building stock in the UK, this research investigates the potential of third-party energy flexibility control in clusters of residential buildings that are transitioning from gas boilers to air source heat pumps. During the field trial involving 30 buildings (intervention group) and 22 flexibility events run between 10/01/2023 and 29/03/2023, two fundamental control mechanisms were tested, *"power limitation"* and *"call to heat"*, focused essentially on power demand reduction. Some issue involving the interaction between third-party flexibility control and default operation and control settings of heat pumps were observed. In general, the *"call to heat"* and *"call to heat - DHW off"* mechanisms resulted in a significantly greater reduction compare to *"power limitation"*, with average power reduction in percentage equal to 88.2% and a maximum power demand reduction of 1.581 kW (average value throughout the cluster of buildings). However, the data also revealed evidence of a substantial rebound in heating power after energy flexibility event occurrences, often called *"snapback"* effect. This can be limited by combining the two mechanisms in sequence first "call-to-heat" and then "power limitation"; this was tested in the 20th event on 22nd March 2023.

Overall, the field trial was successful as the flexibility control strategies were proven to be effective (delivering a relevant power demand reduction) and acceptable for customers. Indeed, during event days, the third-party control override requests received were 1.1 % of the total potential request during the event, and 1.3% after the event during the same day (0.3% before the event); these percentages are just slightly higher than the request received during non event days, equal to 0.7 %. In terms of event duration, the requests received were 4.4% for 2 hour events and smaller percentages for shorter events; however the correlation between duration of events and override requests needs to be studied further, with a larger number of events and considering ambient temperature dependence and other concomitant factors.

Nonetheless, the results obtained indicate the potential of a seamless third-party control which can run as a background service related to heat pump installation, employing a simple and scalable control strategy. Despite the limitations discussed in Section 5.4, the research provides the evidence for further research focused on the definition of business models (in terms of tariffs, incentives, etc.), in the complex value chain of energy flexibility, involving clusters of buildings (customers) and third-parties such ESCOs, aggregators, installers and manufacturers. These entities can partner to propose energy flexibility services alongside energy efficiency services involving the replacement of building technologies, in a context of building stock decarbonisation.

Future research should primarily address the limitations identified and concentrate on a deeper examination of buildings in terms of energy flexibility in relation to ambient temperature, as well as the indoor air temperature drop. This research can be guided by essential information gathered for the buildings in the trial, which can then be compared to statistical reference buildings and energy benchmarks that are available for residential buildings in the UK. In summary, the findings from this study can be used to guide future trials that assess the effectiveness and acceptability of third-party energy flexibility control. This control capabilities can be embedded into widely adopted technologies such as ASHPs, and together with energy efficiency measures, on-site generation and storage, they can play a significant role in the decarbonisation of the building stock, while contributing to the optimal operation of electric infrastructures.

Appendix A

A.1 Flexibility event 7th of February 2023, Domestic Hot Water (DHW) power demand

Hereafter, the results obtained in the flexibility event on the 7th of February 2023 with mechanism *"call to heat"* are reported. The results complement the ones discussed earlier in Section 5.1 and *Figure 15* indicates a behaviour of the DHW power demand similar to *Figure 10* for the event on the 2nd of February 2023. In both cases, a subset of heat pumps would initiate a DHW heat cycle after the mechanism *"call to heat"* is activated, this led to the definition of another mechanism *"call to heat - DHW off"* tested in later events.

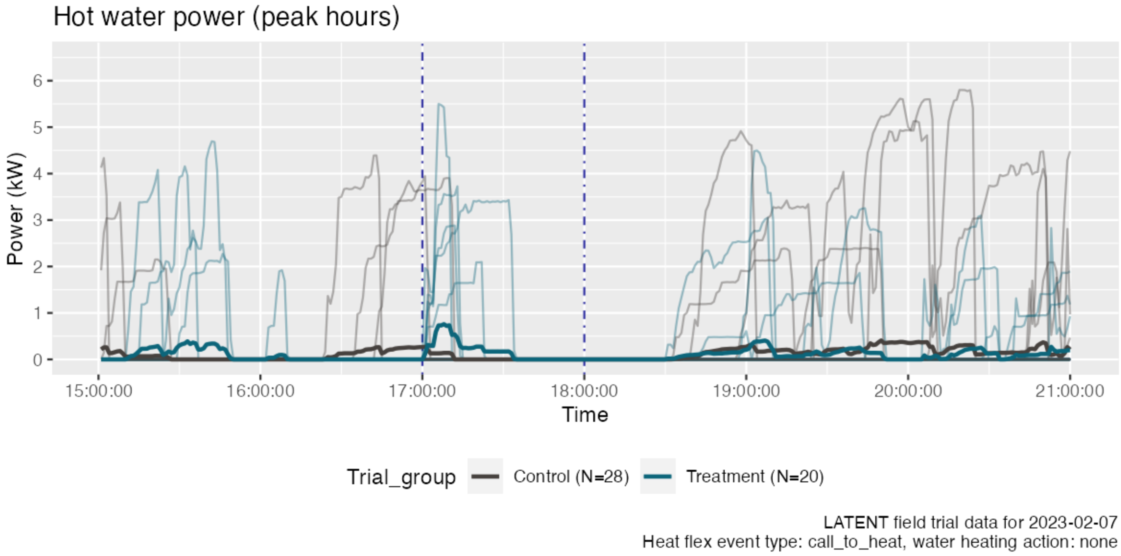

Figure 15: DHW power demand during peak hours, "call to heat - DHW off" flexibility event 7th of February 2023. Thin lines show power for individual households, thick lines show group mean values. The dotted lines show the start and end of the flexibility event.

A.2 Flexibility event 8th of March 2023, total load profile and internal temperatures

Hereafter, the results obtained in the flexibility event on the 8th of March 2023 with mechanism *"Call to heat - DHW off"* are reported. The results complement the ones discussed earlier in Section 5.2 and represent the total load profiles and the individual indoor air temperature trajectories (*Figure 16*), used as a basis to compute the confidence intervals depicted in *Figure 13* and the relative change (normalised by internal-ambient temperature difference) also depicted in *Figure 13*.

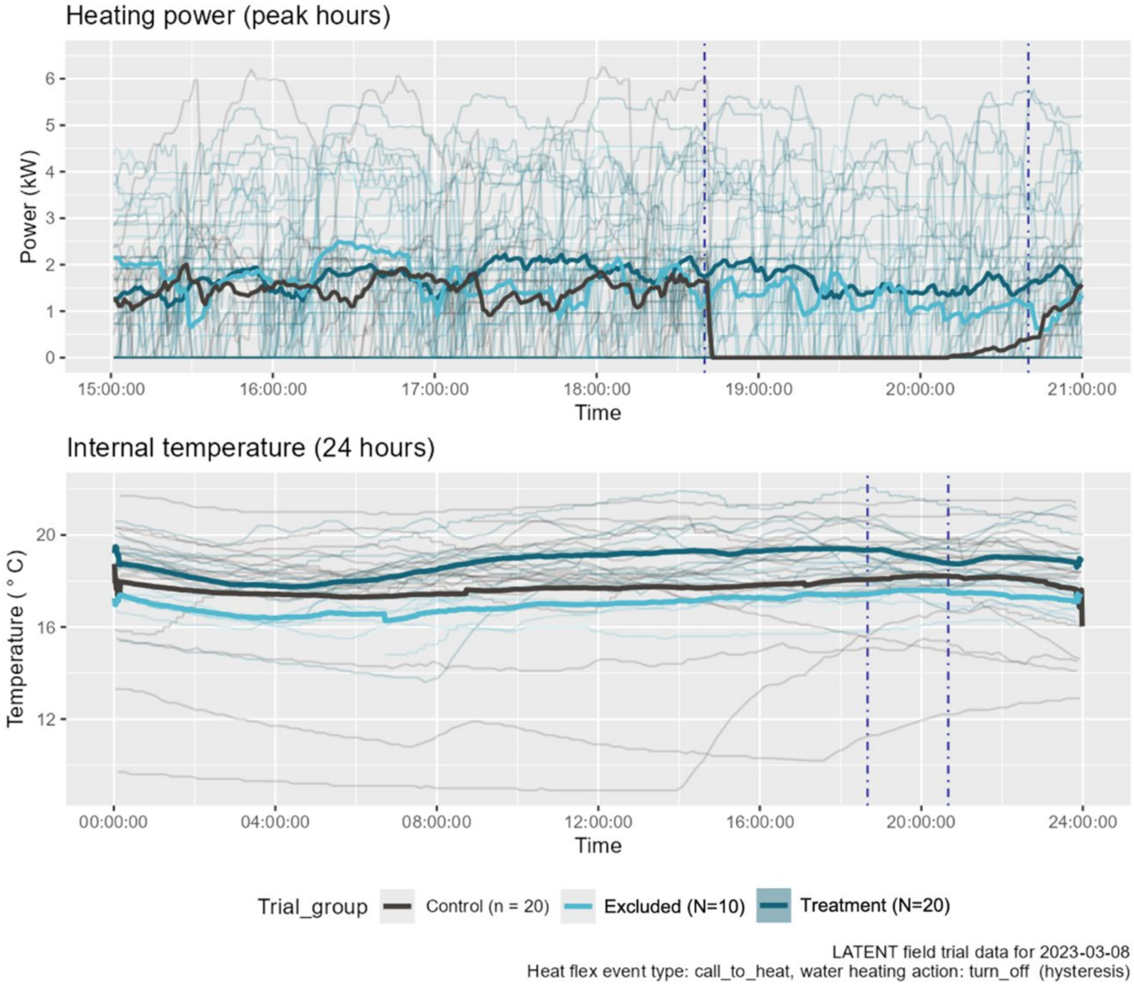

Figure 16: Total load profiles (heat and DHW) for trial households (top) and measured internal temperatures (bottom), for a "call to heat - DHW off" flexibility event 8th March 2023. Thin lines show power/temperature for individual households, thick lines show group mean values. The dotted lines show the start and end of the flexibility event.

A.3 Flexibility event 22nd of March 2023, total load profile, internal and ambient temperatures

Hereafter, the results obtained in the flexibility event on the 22nd of March 2023, which was the day with the highest mean temperature during the event, 11.7 ºC, and a mean daily temperature of 11.0 ºC, are reported. The mechanism run was *"call to heat - DHW off"* and the total load profiles, measured internal temperatures and ambient temperatures are reported in *Figure 17*.

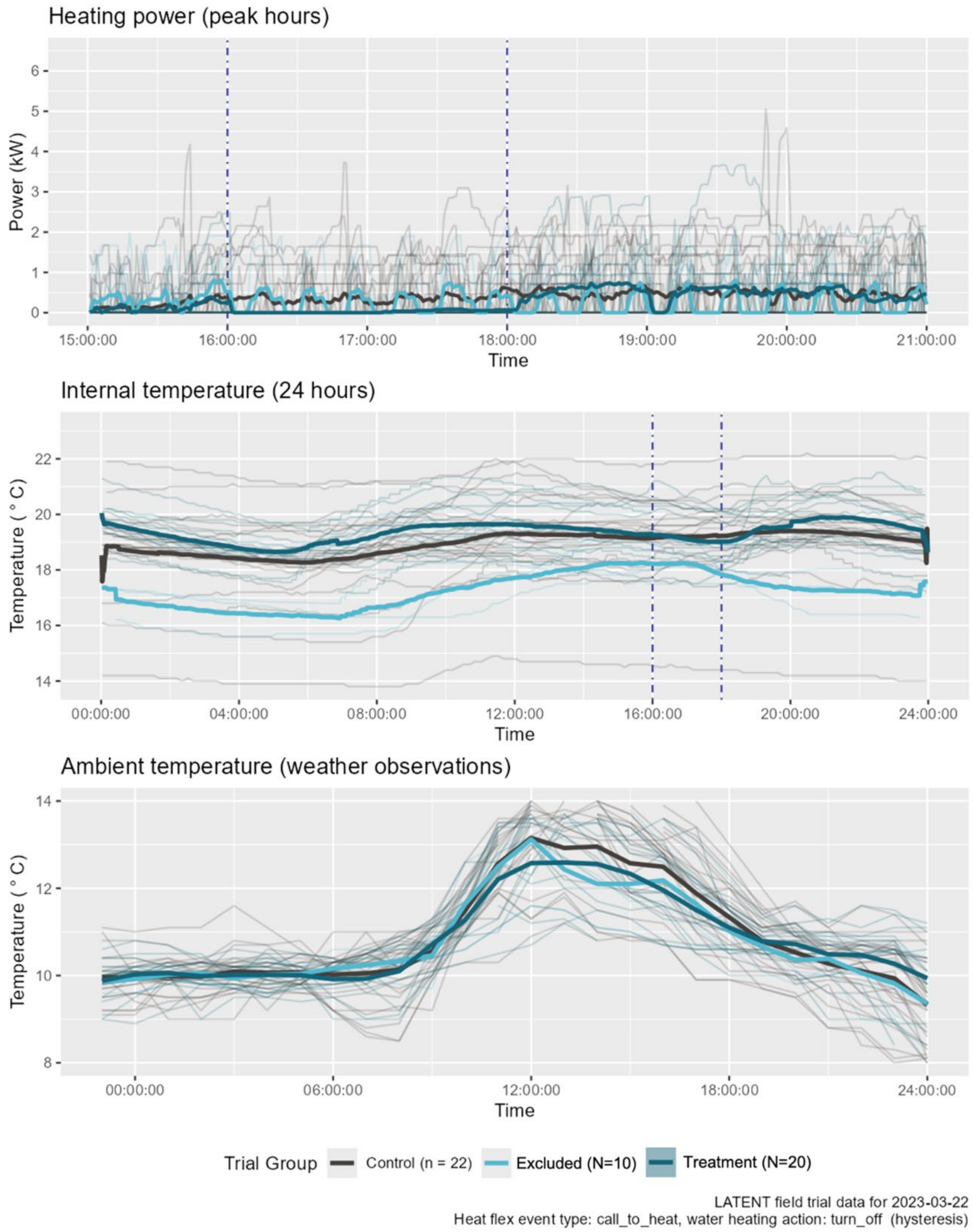

Figure 17: Total load profiles (heat and DHW) (top), measured internal temperatures (middle) and ambient temperature (bottom) for trial households, "call to heat - DHW off" flexibility event 22nd March 2023. Thin lines show power/temperature for individual households, thick lines show group mean values. The dotted lines show the start and end of the flexibility event.

Ethical approval

This project was approved by the University of Southampton Ethics Board (FEPS/70136).

Acknowledgements
This study and project is financially supported by EPSRC Reference: EP/T023074/1, project title “LATENT: Residential heat as an energy system service”.

References

[1] Markard J, Geels FW, Raven R. Challenges in the acceleration of sustainability transitions. Environmental Research Letters 2020;15:81001. https://doi.org/10.1088/1748-9326/ab9468.
[2] Markard J, Hoffmann VH. Analysis of complementarities: Framework and examples from the energy transition. Technol Forecast Soc Change 2016;111:63–75. https://doi.org/https://doi.org/10.1016/j.techfore.2016.06.008.
[3] Department for Business, Energy & Industrial Strategy - Modelling 2050: Electricity system analysis 2020.
[4] Rosenow J, Eyre N. Reinventing energy efficiency for net zero. Energy Res Soc Sci 2022;90:102602. https://doi.org/https://doi.org/10.1016/j.erss.2022.102602.
[5] D’Ettorre F, Banaei M, Ebrahimy R, Pourmousavi SA, Blomgren EMV, Kowalski J, et al. Exploiting demand-side flexibility: State-of-the-art, open issues and social perspective. Renewable and Sustainable Energy Reviews 2022;165:112605. https://doi.org/10.1016/j.rser.2022.112605.
[6] Langevin J, Harris CB, Satre-Meloy A, Chandra-Putra H, Speake A, Present E, et al. US building energy efficiency and flexibility as an electric grid resource. Joule 2021;5:2102–28. https://doi.org/https://doi.org/10.1016/j.joule.2021.06.002.
[7] Jackson R, Zhou E, Reyna J. Building and grid system benefits of demand flexibility and energy efficiency. Joule 2021;5:1927–30. https://doi.org/https://doi.org/10.1016/j.joule.2021.08.001.
[8] Li H, Wang Z, Hong T, Piette MA. Energy flexibility of residential buildings: A systematic review of characterization and quantification methods and applications. Advances in Applied Energy 2021;3:100054. https://doi.org/10.1016/j.adapen.2021.100054.
[9] Wang Y, Wang J, He W. Development of efficient, flexible and affordable heat pumps for supporting heat and power decarbonisation in the UK and beyond: Review and perspectives. Renewable and Sustainable Energy Reviews 2022;154:111747. https://doi.org/https://doi.org/10.1016/j.rser.2021.111747.
[10] Ramirez-Mendiola JL, Mattioli G, Anable J, Torriti J. I’m coming home (to charge): The relation between commuting practices and peak energy demand in the United Kingdom. Energy Res Soc Sci 2022;88:102502. https://doi.org/https://doi.org/10.1016/j.erss.2022.102502.
[11] Papadis E, Tsatsaronis G. Challenges in the decarbonization of the energy sector. Energy 2020;205:118025. https://doi.org/https://doi.org/10.1016/j.energy.2020.118025.
[12] Levesque A, Pietzcker RC, Baumstark L, Luderer G. Deep decarbonisation of buildings energy services through demand and supply transformations in a 1.5°C scenario. Environmental Research Letters 2021;16:54071. https://doi.org/10.1088/1748-9326/abdf07.
[13] Rosenow J, Hamels S. Where to meet on heat? A conceptual framework for optimising demand reduction and decarbonised heat supply. Energy Res Soc Sci 2023;104:103223. https://doi.org/https://doi.org/10.1016/j.erss.2023.103223.
[14] Eyre N, Fawcett T, Topouzi M, Killip G, Oreszczyn T, Jenkinson K, et al. Fabric first: is it still the right approach? Buildings and Cities 2023. https://doi.org/10.5334/bc.388.

[15] Le Dréau J, Lopes RA, O'Connell S, Finn D, Hu M, Queiroz H, et al. Developing energy flexibility in clusters of buildings: A critical analysis of barriers from planning to operation. Energy Build 2023;300:113608. https://doi.org/https://doi.org/10.1016/j.enbuild.2023.113608.
[16] Sala J, Li R, Christensen MH. Clustering and classification of energy meter data: A comparison analysis of data from individual homes and the aggregated data from multiple homes. Build Simul 2021;14:103–17. https://doi.org/10.1007/s12273-019-0587-4.
[17] Martinez S, Vellei M, Le Dréau J. Demand-side flexibility in a residential district: What are the main sources of uncertainty? Energy Build 2022;255:111595. https://doi.org/10.1016/j.enbuild.2021.111595.
[18] Boßmann T, Staffell I. The shape of future electricity demand: Exploring load curves in 2050s Germany and Britain. Energy 2015;90:1317–33. https://doi.org/https://doi.org/10.1016/j.energy.2015.06.082.
[19] Naghiyev E, Shipman R, Goulden M, Gillott M, Spence A. Cost, context, or convenience? Exploring the social acceptance of demand response in the United Kingdom. Energy Res Soc Sci 2022;87:102469. https://doi.org/https://doi.org/10.1016/j.erss.2021.102469.
[20] Aragon V, Gauthier S, Warren P, James PAB, Anderson B. Developing English domestic occupancy profiles. Building Research & Information 2019;47:375–93. https://doi.org/10.1080/09613218.2017.1399719.
[21] Ruhnau O, Hirth L, Praktiknjo A. Time series of heat demand and heat pump efficiency for energy system modeling. Sci Data 2019;6:189. https://doi.org/10.1038/s41597-019-0199-y.
[22] Gibb D, Rosenow J, Lowes R, Hewitt NJ. Coming in from the cold: Heat pump efficiency at low temperatures. Joule 2023;7:1939–42. https://doi.org/https://doi.org/10.1016/j.joule.2023.08.005.
[23] Adams S, Kuch D, Diamond L, Fröhlich P, Henriksen IM, Katzeff C, et al. Social license to automate: A critical review of emerging approaches to electricity demand management. Energy Res Soc Sci 2021;80:102210. https://doi.org/10.1016/j.erss.2021.102210.
[24] Michellod JL, Kuch D, Winzer C, Patel MK, Yilmaz S. Building Social License for Automated Demand-Side Management—Case Study Research in the Swiss Residential Sector. Energies (Basel) 2022;15. https://doi.org/10.3390/en15207759.
[25] Ribó-Pérez D, Heleno M, Álvarez-Bel C. The flexibility gap: Socioeconomic and geographical factors driving residential flexibility. Energy Policy 2021;153:112282. https://doi.org/https://doi.org/10.1016/j.enpol.2021.112282.
[26] James PAB, Gauthier S, Bahaj AS, Manfren M. Grants on the web 2023.
[27] de-Borja-Torrejon M, Mor G, Cipriano J, Leon-Rodriguez A-L, Auer T, Crawley J. Closing the energy flexibility gap: Enriching flexibility performance rating of buildings with monitored data. Energy Build 2024;311:114141. https://doi.org/https://doi.org/10.1016/j.enbuild.2024.114141.
[28] Lopes RA, Chambel A, Neves J, Aelenei D, Martins J. A Literature Review of Methodologies Used to Assess the Energy Flexibility of Buildings. Energy Procedia 2016;91:1053–8. https://doi.org/10.1016/j.egypro.2016.06.274.
[29] Tang H, Wang S, Li H. Flexibility categorization, sources, capabilities and technologies for energy-flexible and grid-responsive buildings: State-of-the-art and future perspective. Energy 2021;219:119598. https://doi.org/10.1016/j.energy.2020.119598.

[30] Chen Y, Xu P, Gu J, Schmidt F, Li W. Measures to improve energy demand flexibility in buildings for demand response (DR): A review. Energy Build 2018;177:125–39. https://doi.org/10.1016/j.enbuild.2018.08.003.
[31] Yan X, Ozturk Y, Hu Z, Song Y. A review on price-driven residential demand response. Renewable and Sustainable Energy Reviews 2018;96:411–9. https://doi.org/10.1016/j.rser.2018.08.003.
[32] Dharani R, Balasubramonian M, Babu TS, Nastasi B. Load Shifting and Peak Clipping for Reducing Energy Consumption in an Indian University Campus. Energies (Basel) 2021;14. https://doi.org/10.3390/en14030558.
[33] Luc KM, Heller A, Rode C. Energy demand flexibility in buildings and district heating systems – a literature review. Advances in Building Energy Research 2019;13:241–63. https://doi.org/10.1080/17512549.2018.1488615.
[34] Vigna I, Pernetti R, Pasut W, Lollini R. Literature review on energy flexibility definitions and indicators for building clusters (https://www.annex67.org/media/1863/literature-review-on-energy-flexibility-definitions-and-indicators-for-building-clusters.pdf). 2019.
[35] Clauß- J, Finck C, Vogler-Finck P, Beagon P. Control Strategies for Building Energy Systems to Unlock Demand Side Flexibility – A Review. Proceedings of Building Simulation 2017: 15th Conference of IBPSA, vol. 15, San Francisco, USA: IBPSA; 2017, p. 1750–9. https://doi.org/https://doi.org/10.26868/25222708.2017.462.
[36] Péan TQ, Salom J, Costa-Castelló R. Review of control strategies for improving the energy flexibility provided by heat pump systems in buildings. J Process Control 2019;74:35–49. https://doi.org/https://doi.org/10.1016/j.jprocont.2018.03.006.
[37] Kathirgamanathan A, De Rosa M, Mangina E, Finn DP. Data-driven predictive control for unlocking building energy flexibility: A review. Renewable and Sustainable Energy Reviews 2021;135:110120. https://doi.org/https://doi.org/10.1016/j.rser.2020.110120.
[38] Lu X, Fu Y, O'Neill Z. Benchmarking high performance HVAC Rule-Based controls with advanced intelligent Controllers: A case study in a Multi-Zone system in Modelica. Energy Build 2023;284:112854. https://doi.org/https://doi.org/10.1016/j.enbuild.2023.112854.
[39] de Wilde P, Aly D, Cho S, Kim J-H, Kim S, Park C-S. Occupant behavioural freedom in building energy use. Appl Energy 2025;377:124682. https://doi.org/https://doi.org/10.1016/j.apenergy.2024.124682.
[40] Li R, Satchwell AJ, Finn D, Christensen TH, Kummert M, Le Dréau J, et al. Ten questions concerning energy flexibility in buildings. Build Environ 2022;223:109461. https://doi.org/10.1016/j.buildenv.2022.109461.
[41] Sousa J, Soares I. Benefits and barriers concerning demand response stakeholder value chain: A systematic literature review. Energy 2023;280:128065. https://doi.org/10.1016/j.energy.2023.128065.
[42] Hamwi M, Lizarralde I, Legardeur J. Demand response business model canvas: A tool for flexibility creation in the electricity markets. J Clean Prod 2021;282:124539. https://doi.org/10.1016/j.jclepro.2020.124539.
[43] Lu R, Hong SH. Incentive-based demand response for smart grid with reinforcement learning and deep neural network. Appl Energy 2019;236:937–49. https://doi.org/https://doi.org/10.1016/j.apenergy.2018.12.061.
[44] Hoicka CE, Lowitzsch J, Brisbois MC, Kumar A, Ramirez Camargo L. Implementing a just renewable energy transition: Policy advice for transposing

the new European rules for renewable energy communities. Energy Policy 2021;156:112435. https://doi.org/10.1016/j.enpol.2021.112435.
[45] European Commission. Directive (EU) 2018/2001 of the European Parliament and of the Council of 11 December 2018 on the Promotion of the Use of Energy from Renewable Sources (Recast). 2018. (https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018L0844&from=IT). 2018.
[46] THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION. Directive (EU) 2019/944 of the European Parliament and of the Council on common rules for the internal market for electricity and amending Directive 2012/27/EU (recast). 2019.
[47] Golden M, Scheer A, Best C. Decarbonization of electricity requires market-based demand flexibility. The Electricity Journal 2019;32:106621. https://doi.org/https://doi.org/10.1016/j.tej.2019.106621.
[48] Chen Y, Guo M, Chen Z, Chen Z, Ji Y. Physical energy and data-driven models in building energy prediction: A review. Energy Reports 2022;8:2656–71. https://doi.org/https://doi.org/10.1016/j.egyr.2022.01.162.
[49] Rouws AL, Loonen RCGM, Hensen JLM. The role of the electricity grid in operation-induced greenhouse gas emissions by a residential building: A multi-year retrospective simulation study. Energy Build 2024;324:114870. https://doi.org/https://doi.org/10.1016/j.enbuild.2024.114870.
[50] Strbac G, Woolf M, Pudjianto D, Zhang X, Walker S, Vahidinasab V. The role of active buildings in the transition to a net zero energy system. Active Building Centre Research Programme 2020.
[51] Adhikari RS, Aste N, Manfren M. Multi-commodity network flow models for dynamic energy management – Smart Grid applications. Energy Procedia 2012;14:1374–9. https://doi.org/http://dx.doi.org/10.1016/j.egypro.2011.12.1104.
[52] Manfren M. Multi-commodity network flow models for dynamic energy management – Mathematical formulation. Energy Procedia 2012;14:1380–5. https://doi.org/http://dx.doi.org/10.1016/j.egypro.2011.12.1105.
[53] Göke L. A graph-based formulation for modeling macro-energy systems. Appl Energy 2021;301:117377. https://doi.org/https://doi.org/10.1016/j.apenergy.2021.117377.
[54] Le Dréau J, Heiselberg P. Energy flexibility of residential buildings using short term heat storage in the thermal mass. Energy 2016;111:991–1002. https://doi.org/10.1016/j.energy.2016.05.076.
[55] Askeland M, Georges L, Korpås M. Low-parameter linear model to activate the flexibility of the building thermal mass in energy system optimization. Smart Energy 2023;9:100094. https://doi.org/https://doi.org/10.1016/j.segy.2023.100094.
[56] Alzetto F, Pandraud G, Fitton R, Heusler I, Sinnesbichler H. QUB: A fast dynamic method for in-situ measurement of the whole building heat loss. Energy Build 2018;174:124–33. https://doi.org/https://doi.org/10.1016/j.enbuild.2018.06.002.
[57] Papachristou C, Hoes P-J, Loomans MGLC, van Goch TAJ, Hensen JLM. Investigating the energy flexibility of Dutch office buildings on single building level and building cluster level. Journal of Building Engineering 2021;40:102687. https://doi.org/https://doi.org/10.1016/j.jobe.2021.102687.
[58] Crawley J, Martin-Vilaseca A, Wingfield J, Gill Z, Shipworth M, Elwell C. Demand response with heat pumps: Practical implementation of three different

control options. Building Services Engineering Research and Technology 2022;44:211–28. https://doi.org/10.1177/01436244221145871.
[59] Kazmi H, Fu C, Miller C. Ten questions concerning data-driven modelling and forecasting of operational energy demand at building and urban scale. Build Environ 2023;239:110407. https://doi.org/https://doi.org/10.1016/j.buildenv.2023.110407.
[60] Price P. Methods for analyzing electric load shape and its variability, Lawrence Berkeley National Laboratory Report LBNL-3713E. 2010.
[61] Mathieu JL, Price PN, Kiliccote S, Piette MA. Quantifying changes in building electricity use, with application to demand response. IEEE Trans Smart Grid 2011;2:507–18.
[62] Borgeson SD. Targeted efficiency: Using customer meter data to improve efficiency program outcomes. University of California, Berkeley. 2013.
[63] OpenEEmeter (https://github.com/openeemeter/eemeter).
[64] CalTRACK. CalTRACK Methods (http://docs.caltrack.org/en/latest/methods.html).
[65] Granderson J, Fernandes S, Crowe E, Sharma M, Jump D, Johnson D. Accuracy of hourly energy predictions for demand flexibility applications. Energy Build 2023;295:113297. https://doi.org/https://doi.org/10.1016/j.enbuild.2023.113297.
[66] Mirfin A, Xiao X, Jack MW. TOWST: A physics-informed statistical model for building energy consumption with solar gain. Appl Energy 2024;369:123488. https://doi.org/https://doi.org/10.1016/j.apenergy.2024.123488.
[67] Manfren M, James PAB, Aragon V, Tronchin L. Lean and interpretable digital twins for building energy monitoring – A case study with smart thermostatic radiator valves and gas absorption heat pumps. Energy and AI 2023;14:100304. https://doi.org/https://doi.org/10.1016/j.egyai.2023.100304.
[68] Manfren M, Gonzalez-Carreon KM, James PAB. Interpretable Data-Driven Methods for Building Energy Modelling - A Review of Critical Connections and Gaps. Energies (Basel) 2024;17. https://doi.org/10.3390/en17040881.
[69] Interpretable machine learning, Section 3.2 Taxonomy of Interpretability Methods, Christopher Molnar (https://christophm.github.io/interpretable-ml-book/taxonomy-of-interpretability-methods.html), accessed on 24/05/82023.
[70] Lazzarin R, Busato F, Noro M. Heat pumps in refurbishment of existing buildings. J Fed Eur Heat Vent Air Cond Assoc REHVA 2012;6:45–9.
[71] Manfren M, Nastasi B. Parametric Performance Analysis and Energy Model Calibration Workflow Integration—A Scalable Approach for Buildings. Energies 2020;13. https://doi.org/10.3390/en13030621.
[72] Glass J, Suffian S, Scheer A, Best C. Demand Response Advanced Measurement Methodology - Analysis of Open-Source Baseline and Comparison Group Methods to Enable CAISO Demand Response Resource Performance Evaluation. 2022.
[73] Wang G, Hamad R, White JS. Advances in Difference-in-differences Methods for Policy Evaluation Research. Epidemiology 2024;35.
[74] Roth J, Sant'Anna PHC, Bilinski A, Poe J. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. J Econom 2023;235:2218–44. https://doi.org/https://doi.org/10.1016/j.jeconom.2023.03.008.
[75] Creswell JW. Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications; 2017.

[76] Reynders G, Diriken J, Saelens D. Generic characterization method for energy flexibility: Applied to structural thermal storage in residential buildings. Appl Energy 2017;198:192–202. https://doi.org/10.1016/j.apenergy.2017.04.061.
[77] Johra H, Goupy M, Wittchen KB. Thermal storage capacity in the entire building stock of Denmark for energy flexibility strategies and building-to-grid services. REHVA Journal 2024;61:14–7.
[78] Badiei A, Allinson D, Lomas KJ. Automated dynamic thermal simulation of houses and housing stocks using readily available reduced data. Energy Build 2019;203:109431. https://doi.org/https://doi.org/10.1016/j.enbuild.2019.109431.
[79] Wang A, Li R, You S. Development of a data driven approach to explore the energy flexibility potential of building clusters. Appl Energy 2018;232:89–100. https://doi.org/10.1016/j.apenergy.2018.09.187.
[80] Clauß J, Brozovsky J, Georges L. Demonstrating the load-shifting potential of a schedule-based control in a real-life educational building. Energy Build 2024;316:114321. https://doi.org/https://doi.org/10.1016/j.enbuild.2024.114321.
[81] Vigna I, Lollini R, Pernetti R. Assessing the energy flexibility of building clusters under different forcing factors. Journal of Building Engineering 2021;44:102888. https://doi.org/10.1016/j.jobe.2021.102888.
[82] Ghofrani A, Nazemi SD, Jafari MA. HVAC load synchronization in smart building communities. Sustain Cities Soc 2019;51:101741. https://doi.org/10.1016/j.scs.2019.101741.
[83] Manfren M, Nastasi B, Piana E, Tronchin L. On the link between energy performance of building and thermal comfort: An example. AIP Conf Proc 2019;2123:20066. https://doi.org/10.1063/1.5116993.
[84] Aste N, Buzzetti M, Caputo P, Manfren M. Local energy efficiency programs: A monitoring methodology for heating systems. Sustain Cities Soc 2014;13:69–77. https://doi.org/http://doi.org/10.1016/j.scs.2014.04.006.
[85] Chen X, Navidi T, Rajagopal R. Energy resource control via privacy preserving data. Electric Power Systems Research 2020;189:106719. https://doi.org/https://doi.org/10.1016/j.epsr.2020.106719.
[86] National Grid - Projects: EQUINOX (Equitable Novel Flexibility Exchange) n.d.
[87] Centre for Net Zero - HeatFlex UK.
[88] NEDO, Implementation Report for Smart Community Demonstration Project in Greater Manchester, UK, Manchester, (2017).
[89] Crawley J, Johnson C, Calver P, Fell M. Demand response beyond the numbers: A critical reappraisal of flexibility in two United Kingdom field trials. Energy Res Soc Sci 2021;75:102032. https://doi.org/https://doi.org/10.1016/j.erss.2021.102032.
[90] Gupta R, Morey J. Evaluation of residential demand response trials with smart heat pumps and batteries and their effect at the substation feeder. J Clean Prod 2023;403:136760. https://doi.org/10.1016/j.jclepro.2023.136760.
[91] Pelka S, Preuß S, Stute J, Chappin E, de Vries L. One service fits all? Insights on demand response dilemmas of differently equipped households in Germany. Energy Res Soc Sci 2024;112:103517. https://doi.org/https://doi.org/10.1016/j.erss.2024.103517.
[92] Crawley J, Moore G, Higginson S, Elwell C, Eyre N. The Role of Domestic Heat Pumps in Providing Flexibility to the UK Electricity System. Energies (Basel) 2024;17. https://doi.org/10.3390/en17122995.
[93] ThingsBoard - Open-source IoT Platform (https://thingsboard.io/).
[94] Visual Crossing Weather (https://www.visualcrossing.com/).

[95] TADO (https://www.tado.com/gb-en).
[96] Hall M, Geissler A. Comparison of Flexibility Factors and Introduction of A Flexibility Classification Using Advanced Heat Pump Control. Energies (Basel) 2021;14. https://doi.org/10.3390/en14248391.
[97] Department for Levelling Up, Housing and Communities - English Housing Survey 2021 to 2022: headline report. National Statistics. 2022.
[98] Office for National Statistics - Census 2021. 2023.
[99] Department for Work and Pensions - Family Resources Survey: financial year 2020 to 2021. 2022.
[100] Elexon - BSC - Profiling (https://www.elexon.co.uk/settlement/profiling/).
[101] Elexon – BSC - Load Profiles and their use in Electricity Settlement (https://bscdocs.elexon.co.uk/guidance-notes/load-profiles-and-their-use-in-electricity-settlement#what-is-a-load-profile).