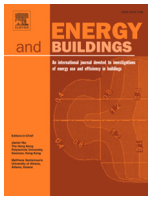
Energy & Buildings 354 (2026) 116975

Contents lists available at ScienceDirect

Energy & Buildings

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



Frequency control with aggregated heat pumps in buildings: A comparison of complex and simple control strategies

Manuel Koch [a,b,*], Paul Scharnhorst [c], Parantapa Sawant [b], Colin N. Jones [a]

[a] EPFL, STI IGM LA3, Lausanne, 1015, Vaud, Switzerland
[b] FHNW, Hofackerstrasse 30, Muttenz, 4132, Basel-Land, Switzerland
[c] CSEM, Rue Jaquet-Droz 1, Neuchâtel, 2002, Neuchâtel, Switzerland

ARTICLE INFO

Keywords:
Buildings
Heat pumps
Frequency control
System identification
Model predictive control

ABSTRACT

As the share of electricity generation from non-dispatchable sources like wind and photovoltaics grows, so does the need for active grid stabilization. For this purpose, we use aggregated heat pumps in buildings as a virtual power plant. Unlike other technologies used to stabilize the grid, such as pumped hydro or natural gas power plants, the aggregation of buildings is not hindered by geography, nor does it incur any additional strain on the environment. However, a large number of buildings is required to reach a substantial level of deployable power. The right trade-off between cost and performance in the aggregation method is essential to be competitive with the more established technologies in this field. To this end, we evaluate simple and complex frameworks for frequency control with a simulated cluster of 1000 buildings, based on real data, and estimate their respective cost-effectiveness. The most complex methods, using full data access and bi-level optimization, yield the highest total flexibility provided and the lowest tracking error. However, the lower system costs of the less complex methods, relying on limited data and simplified formulations of the optimization problem, outweigh their inferior performance in terms of expected profitability. With the assumed tariffs, we only expect a net profit when using the more cost-effective methods with large buildings. To achieve profitability with small buildings within the specifics of our simulations, the remunerations would have to increase by a factor of two to seven.

1. Introduction

To prevent system failure in the electric grid, a fine balance between production and consumption must be maintained, most commonly by adjusting the production to match the consumption. In case of a power deficit or surplus, the grid operator requests certain power plants, which are contracted for this purpose, to increase or decrease their power production. However, this becomes more challenging as the share of non-dispatchable electricity generation from wind and photovoltaics grows [1–3]. At the same time, the most well-suited power plants for variable generation face an uncertain future: Natural gas peaker plants are meant to be phased out under many countries carbon emission reduction goals [4,5]. Large hydropower plants can only be built in a few suitable locations, and may be increasingly hindered by droughts in the future [6,7]. Lastly, the currently existing battery technologies suitable for grid storage are expensive [8], and require rare materials with vulnerable supply chains [9].

On the other hand, it is also possible to align consumption with production in a similar manner. This is known as demand response and can be broadly classified into two groups [10]: In price-based programs, the utility provider charges higher tariffs during times of high demand or low supply, but it is left to the consumers to shift their consumption accordingly. This can be done with a fixed schedule (time-of-use, TOU) or dynamically (real-time pricing, RTP). In incentive-based programs, the consumers agree to follow a direct request from the provider to increase or decrease their consumption. They are paid for providing this service, or penalized if they fail to do so. The advantage of price-based programs is that they are easier to implement technologically and legally. The advantage of incentive-based programs is that they allow a more controllable modulation of the consumption. Naturally, demand response requires consumption with substantial flexibility in its scheduling.

One such category of consumption is the heating and cooling of buildings with electric heat pumps and chillers. Buildings account for around a third of the total primary energy consumption in developed countries, with roughly half of it being used for heating and cooling [11–15], resulting in around 15 % of the total primary energy being used for heating and cooling of buildings. Almost all cooling is provided by electric chillers [16,17], and electric heat pumps are becoming the

* Corresponding author.
E-mail address: manuelpascal.koch@epfl.ch (M. Koch).

https://doi.org/10.1016/j.enbuild.2026.116975
Received 29 October 2025; Received in revised form 23 December 2025; Accepted 6 January 2026
Available online 11 January 2026

Nomenclature

a	AGC signal
A	state matrix
b	binary activation variable
B	control matrix
C	observation matrix
I	radiation
j	scenario
k	discrete time step
M	number of scenarios
N	prediction horizon
P	power
R	flexibility envelope
T	temperature
u	control variable for building
μ	control variable for cluster
v	disturbance scenario
w	weather/disturbance
x	system state of building
y	observed variable
γ	flexibility band (one sided)
χ	dystem state of cluster
AGC	automatic generation control
ARX	autoregressive model with exogenous inputs
CHF	Swiss franc
HVAC	heating, ventilation & air conditioning
MPC	model predictive control
TCL	thermostatically controlled load

predominant method of heating in many countries [18,19], electrifying this segment of consumption as well. This large energetic volume[1] is usually complemented by substantial temporal flexibility[2] stemming from the buildings' thermal inertia. The usable flexibility of an individual building depends strongly on the construction style and HVAC system. Typical time constants for the indoor temperature may range from 10 min to 20 min for forced-air, 30 min to 120 min for wall-mounted radiators, and 2 h to 8 h for radiant slabs.

When buildings are used for demand response, they are often aggregated in large numbers to average out individual tracking errors, since HVAC equipment is limited in its ability to modulate power consumption accurately, and to address grid operators' unwillingness to engage with a large number of small consumers individually. The electric power of heat pumps and chillers in buildings typically ranges from a few kW for single-family houses to hundreds of kW for large commercial buildings, as opposed to power plants typically ranging from hundreds of MW to a few GW. From the many methods to aggregate buildings for demand response in the existing literature [20–22], we outline three major categories that will be invoked later in this paper: The most simple approach is to control each building individually and only aggregate them in terms of accounting. The algorithms are thus the same as for controlling one building only, which has been thoroughly studied [23]. To provide some specific examples relevant to this paper, reference [24] uses data-driven models of multiple rooms in an office building to conduct a series of short-term experiments. Reference [25] evaluates an adaptive system identification framework to account for seasonal changes in a longer-term simulation study with a residential building. Reference [26] demonstrates an adaptive direct data-driven approach in a series of experiments on an isolated classroom building.

[1] Total electric energy consumption of the relevant systems over the relevant time period.

[2] Ability to reschedule energy consumption in terms of an earth mover's distance.

On the other hand, there are methods based on flexibility envelopes, whose properties are well-suited for aggregated control frameworks. Different definitions of flexibility envelopes exist, from a set of feasible power trajectories through energy bounds [27], or time-based feasible power shifts [27–29], to feasible power augmentations for fixed time [30]. The commonality between the different types is that they are primarily used for flexibility quantification rather than direct control, e.g., [28] uses an MPC formulation to control a building with its appliances while reporting the envelopes to the system operator, and [31] presents a new way to consider uncertainty in the envelope prediction. Here, we follow the approach of [30] by computing individual flexibility envelopes and using them for the aggregated dispatch of flexibility requests.

Lastly, we introduce the category of thermostatically controlled loads (TCL), a defining feature of which is the indirect control of a device's electricity consumption via thermostat setpoints, hence the name. Examples may include refrigerators, freezers, air-conditioning units and heat pumps, the latter being the focus of our work. The key advantage of this indirect control is the ability to manipulate a device's electricity consumption while leaving the local low-level control in place. The key disadvantage is the uncertainty of the device's response stemming from said low-level controller. Despite this, the feasibility of using TCLs for demand response with a variety of control methods has been well-established [32,33]. The latter reports challenges when using TCLs for frequency control on a sub-minute time scale, which is faster than the time steps used by us. Furthermore, we take previously raised concerns regarding privacy and thermal comfort [34] into account by introducing a supervisory control framework using only aggregated data. While many more approaches to building aggregation exist, we restrict the scope of this paper to the ones discussed above.

Beyond the control algorithms per se, a meta-review of the literature [10,35–40] has revealed a number of more contextual topics that merit further investigation, two of which we address in this study: (1) The design of reliable and deployable hardware (sensors, communication, etc.) to harness buildings for demand response. (2) The evaluation of the cost effectiveness and energy efficiency of using buildings for demand response. To this end, we simulate incentive-based demand response with a cluster of 1000 buildings with heat pumps. The building models are based on real-world operational data, detailed in Section 2.1. The demand response scheme is based on secondary frequency control in the Swiss national grid. The operation of the heat pumps is shifted within a flexibility band around a baseline, as shown in Section 2.2. The two core research questions of this study are: (1) Can more complex control frameworks deliver better performance for frequency control, compared to simpler alternatives? (2) If so, does the improved performance justify the higher expected system costs? Correspondingly, we introduce four different frameworks in Sections 2.4–2.7, ranging from the upper to the lower end of a reasonable spectrum of complexity, as we judged it. Section 3 compares the results in terms of the amount of load shifting provided, as well as their tracking errors. Furthermore, we provide cost estimates for each variant to evaluate their respective cost-effectiveness, and gain insights into the marginal benefits of the more complex, i.e. more expensive, control frameworks. Section 4 discusses the limitations of our study, before final conclusions are drawn in Section 5.

2. Methodology

2.1. Building cluster model

In a series of selection steps, 1000 buildings are selected from an initial dataset of almost 60 000, identified from donated real-world operational data in [41]. The authors of the original paper fit linear gray-box models with one to five states to the data in an iterative process, publishing the most accurate one for each building. Since only temperature measurements and total heating runtime were collected, but not the actual heating power, a uniform heating input is reconstructed to identify

the models. Correspondingly, these models are inherently normalized in terms of heating power. For each building, the model structure, the parameters, their statistical significance, a quality of fit category, and a series of root mean square errors for open-loop predictions over varying horizons are given.

In the first step of the selection, we identify every building that meets the following criteria: (1) A fifth order model. This includes the majority of the buildings. (2) Located in Köppen climate zones Dfa[3] or Dfb[4]. This encompasses cities like Chicago, Toronto and Boston. (3) Fit qualified as "good" by the authors of [41], as opposed to "close" or "poor". (4) The presence of a heating system.

This reduces the selection to 9441 buildings. Considering that the dataset includes buildings with extreme parameters, whose validity the authors themselves call into question (despite the "good" fit), we then exclude every building that has a physical parameter in the top or bottom 5 % of the statistical distribution of this parameter out of the first selection. This further reduces the selection to 4692 buildings.

Next, we generate a continuous-time state-space model of the form $\dot{x} = Ax + B_w w + B_u u$ for each of them, with the normalized heating power as the input u, the ambient temperature and global horizontal solar radiation as disturbances $w = \begin{pmatrix} T_{amb} & I_{glo} \end{pmatrix}^\top$, the state $x = \begin{pmatrix} T_{sensor} & T_{mass} & T_{heat} & T_{extern} & T_{interior} \end{pmatrix}^\top$, representing the temperatures of the thermostat sensor, internal mass (walls and furniture), heating system, external walls and internal air. This is completed by the output $y_k = T_{sensor,k} = Cx_k$. Each building model is discretized using the *Matlab c2d* command[5] with the *zero-order hold* method and a step size of 1 min, which is sufficient to capture the fastest dynamics found in the cluster. Each model is completed by an on-off controller for the zone temperature with a hysteresis of 1 K.

In the third selection step, we simulate these buildings from November through February with randomly switching temperature setpoints. Corresponding to the aforementioned climate zones, we use EnergyPlus weather data for Chicago O'Hare[6]. Buildings with a mean absolute tracking error above 1 K over this four-month simulation are excluded, reducing the selection to 4476 buildings. In the fourth step, we exclude buildings whose average heating power over the simulation period is below 35 % or above 80 % of the nominal power, since buildings that need either very little heating, or almost permanent heating, are ill-suited for demand response. A building that requires almost permanent heating just to maintain a target indoor temperature is limited in its upward flexibility by its nominal power, and limited in its downward flexibility by its lower temperature constraint. The inverse applies to buildings that require almost no heating. In practice, it can be assumed that a provider of such a service to grid operators has some kind of screening process in place to exclude ill-suited buildings from their pool. This step strongly reduces the number of considered buildings to 1067. It is unclear if the excluded buildings in fact have such unusual behavior, or if poor operational data resulted in the identification of poor models. For the final selection step, we choose the first 1000 buildings, based on the order in which they are listed in the dataset, to comprise the building cluster for this study. We note that this is not a selection of 1000 random buildings, but a selection of 1000 buildings reasonably well-suited to demand response.

For the purpose of this study, we assume a heat pump for each building. As mentioned before, the original building models provided by [41] are normalized with regard to heating power. Since a collection of buildings is unlikely to have uniform heating power in reality, we randomly assign each building's heat pump a nominal electric power between 2 kW to 10 kW with equal probability, to diversify the cluster for the subsequent simulations. The total nominal power of the resulting building cluster is 5.9 MW. In the process of identifying linear models, various nonlinearities and disturbances that are present in real buildings are implicitly approximated in first order. For the scope of this paper, we thus omit the reintroduction of nonlinear components, such as internal gains or the temperature dependence of a heat pump's coefficient of performance. The impact thereof and the possibility of a follow-up study with more detailed building models are discussed in Section 4.

2.2. Secondary frequency control framework

As mentioned in the introduction, we aim to participate in the secondary level of frequency control of the Swiss national grid, which has three distinct levels: Primary frequency control, which is activated within seconds; secondary frequency control, which is activated within minutes; and tertiary frequency control, which is activated should an imbalance last longer than 15 min [42]. For buildings, we consider the secondary level to be the most appropriate one. The faster primary level would impose an excessive strain on the mechanical components of the heating system by requiring fast switching. The slower tertiary level would be sub-optimal for buildings with a light construction style, which have faster thermodynamic time constants than heavy brick- or concrete-based buildings. The general sensibility and feasibility of using buildings for frequency control, beyond the arguably simpler price-based schemes, has been previously established by multiple publications [43–48].

For our simulations, we implement the secondary frequency control based on the work of [24,49] as follows: A baseline power consumption and a flexibility band γ are calculated ahead of time at midnight for the following day, and communicated to the grid operator, corresponding to the first plot of the simplified visualization in Fig. 1. The flexibility band is symmetrically applied to both sides of the baseline and constant over each day. Within this band, the power consumption is then modulated based on the instantaneous automatic generation control (AGC) signal from the grid operator, which has a range of $[-1, 1]$, corresponding to the minimum and maximum power of the pre-calculated flexibility band (Plots 2 and 3). The resulting reference trajectory must then be tracked with a specified accuracy (Plot 4). Since the grid operator rewards the amount of flexibility provided, we seek to maximize the width of the flexibility band, occasionally at the expense of optimal energy efficiency.

From [24,49], we also have one year of AGC data with a 1 min resolution. This dataset was artificially generated to match the statistical properties of a sample of real data from the Swiss national grid operator. The frequency decomposition in Fig. 2 shows that some of the strongest components are in the range of minutes to hours, which buildings are well-suited to cover. Because we do not want to switch the heat pump more frequently than every 15 min, we resample the original AGC signal correspondingly. We note that doing so attenuates some of the sharper peaks in the original data. Fig. 3 shows the weather and AGC signal for the entire simulation period. The top plot shows harsh ambient temperatures of as low as -20°C and large temperature changes of up to 30 K within a few days. The middle plot shows moderately sunny conditions. The AGC signal in the bottom plot is close to zero for most of the time, with a few pronounced spikes around days 6, 27 and 44.

2.3. Control overview

Sections 2.4–2.7 detail the four control schemes used for this study in descending order of complexity. The *ideal* variant makes the most favorable assumptions regarding state knowledge and control fidelity, with a separate bi-level control scheme for each building, and serves as a semi-hypothetical best-case scenario. The *individual* variant uses the same overall architecture with individual control of each building, but non-ideal identified models. The *intermediate* variant restricts state knowledge to the heat pump power of each building, which it can override if needed, but uses an aggregated optimization for the whole cluster.

[3] Dfa: Continental, no dry season, hot summer
[4] Dfb: Continental, no dry season, warm summer
[5] https://mathworks.com/help/control/ref/dynamicsystem.c2d.html
[6] https://energyplus.net/weather-location/north_and_central_america_wmo_region_4/USA/IL/USA_IL_Chicago-OHare.Intl.AP.725300_TMY3

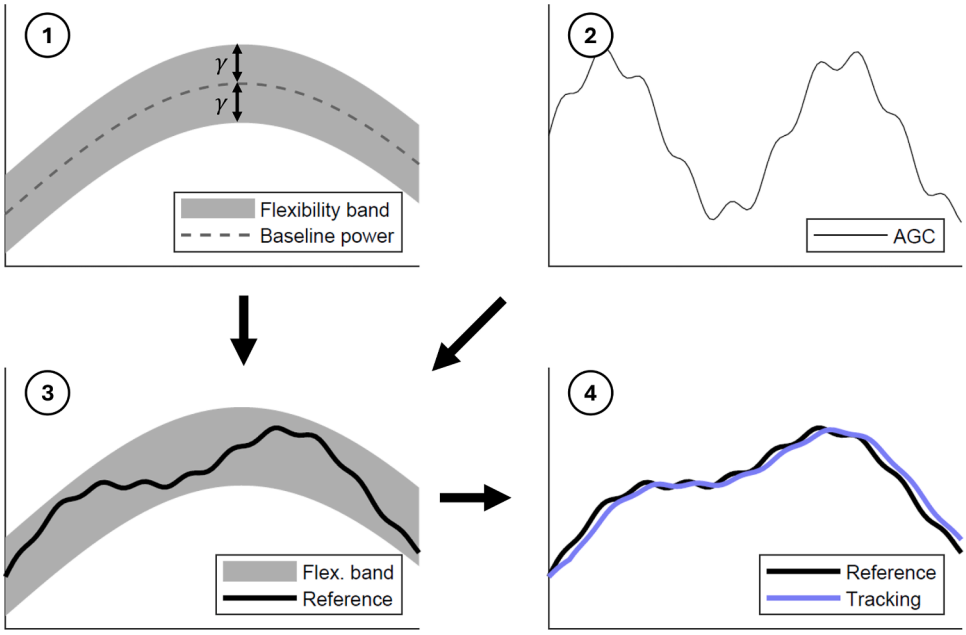


Fig. 1. Baseline power, flexibility band γ and AGC signal are merged into a reference power to be tracked with the heat pump.

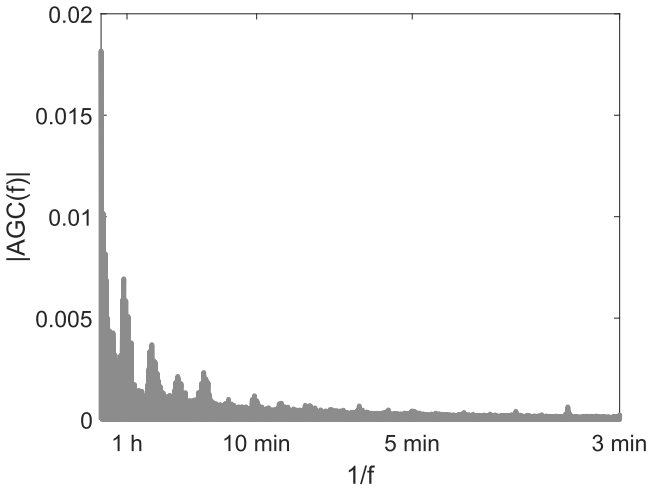


Fig. 2. Spectral decomposition of AGC signal.

The *minimalist* variant uses no building-level knowledge, only the aggregate power consumption. It only controls the heat pumps indirectly, by overriding the thermostat setpoint within a defined band. As the name suggests, we aim to keep this variant as simple as possible. A visual comparison is given in Fig. 4. For all four frameworks, the time step for optimization is 15 min. The zone temperature limits are 20°C to 24°C, with a default value of 22°C. The evaluation period covers January and February. For the variants that require system identification, it is performed from mid-November through December. The objective for all controllers is to maximize load shifting and minimize tracking error, while avoiding violations of the zone temperatures. All variants have access to ideal forecasts of the weather 36 h into the future, and the AGC signal 6 h into the future, except for the intermediate variant, which uses an AGC forecast of 24 h.

2.4. Ideal control

For the ideal control scheme, every building has its own day-ahead planner and a subordinate MPC for execution. Both use perfect models and full state knowledge. The control variable is the normalized compressor speed, i.e. electric power, of the heat pump, which is treated as a continuous variable, meaning that we neglect the minimum speed that a real compressor would have. As the name of the variant indicates, these are highly idealized assumptions, which are meant to approximate an upper bound of the theoretically achievable control performance.

The planner (1) calculates a baseline consumption and flexibility band for the following day once per day at midnight. Doing so is a robust optimization problem of an unusual nature: The disturbance under which we seek to guarantee constraint satisfaction is not some undesired process noise, but the AGC signal we want to track, which is unknown at this stage of the optimization. Conventional robust optimization tends to be very conservative and assume uniform noise [50]. Since neither of these properties are desired for our application, we use the scenario-approach to robustness, which optimizes over a number of measured sample disturbances, called the *scenarios*, trying to achieve constraint satisfaction for all of them [51]. This method has been applied to buildings in several publications [52–54]. There is theoretical work on how many scenarios to use, but it is highly conservative. Reference [54] reports good performance with four scenarios, capturing the uncertainties of the weather forecast and internal gains, in a building control study, when the theory suggested to use 4799 scenarios. We use a selection of $M = 10$ scenarios, that is well-diversified and reasonably representative of the whole AGC dataset, shown in Fig. 5.

The planning horizon is 36 h, i.e. $N_p = 144$. This covers the 24 h of the current day, the 6 h horizon of the low-level controller at the end of the day, plus an additional 6 h to prevent turnpike behaviors at 30 h, as shown in Fig. 6.

These scenarios[7] υ are applied in the dynamics constraints (1b). Multiplied by the flexibility band γ, they act as an input disturbance, creating M different state trajectories, all of which should satisfy the output constraint (1c). We note that since constraint satisfaction cannot be guaranteed, it is actually implemented as a soft constraint, which is omitted from the equations for the sake of legibility. The input constraint (1d) assures the ability to track the reference in the moment, since γ is the maximum amount by which we will be asked to deviate

[7] To avoid confusion, we use the symbol a for the actual and predicted AGC signal, but υ for the scenarios, which are comprised of AGC signal trajectories.

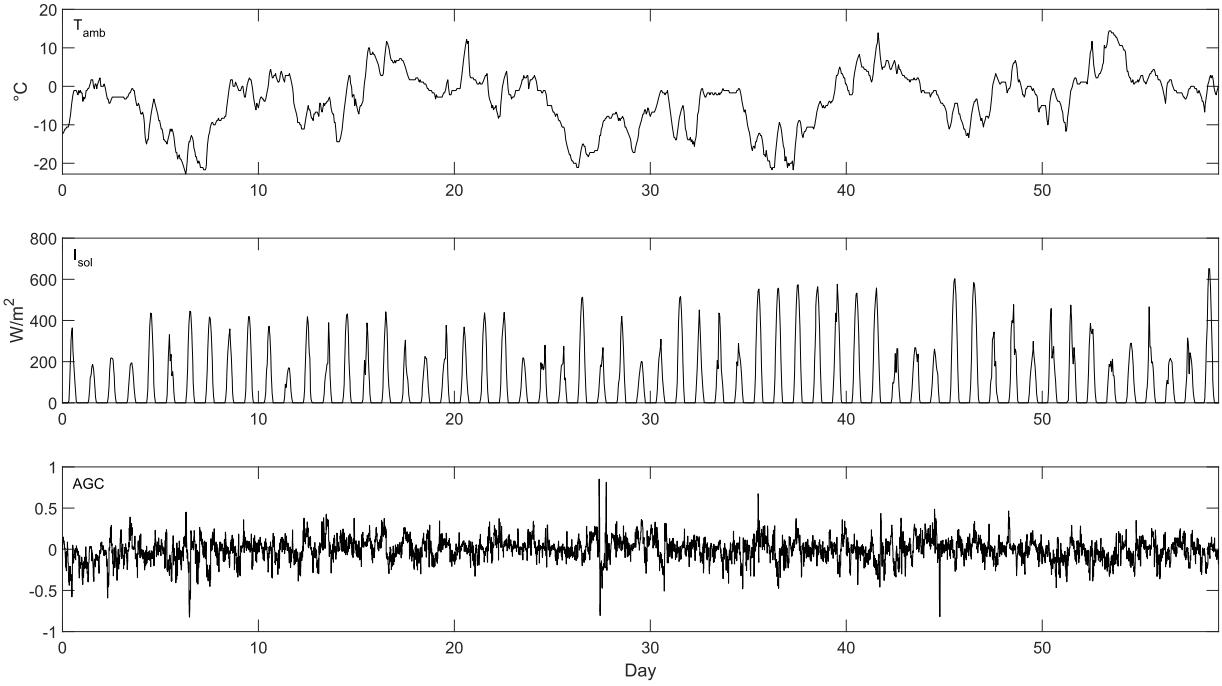


Fig. 3. Ambient temperature, global horizontal radiation and AGC signal for the whole simulation period.

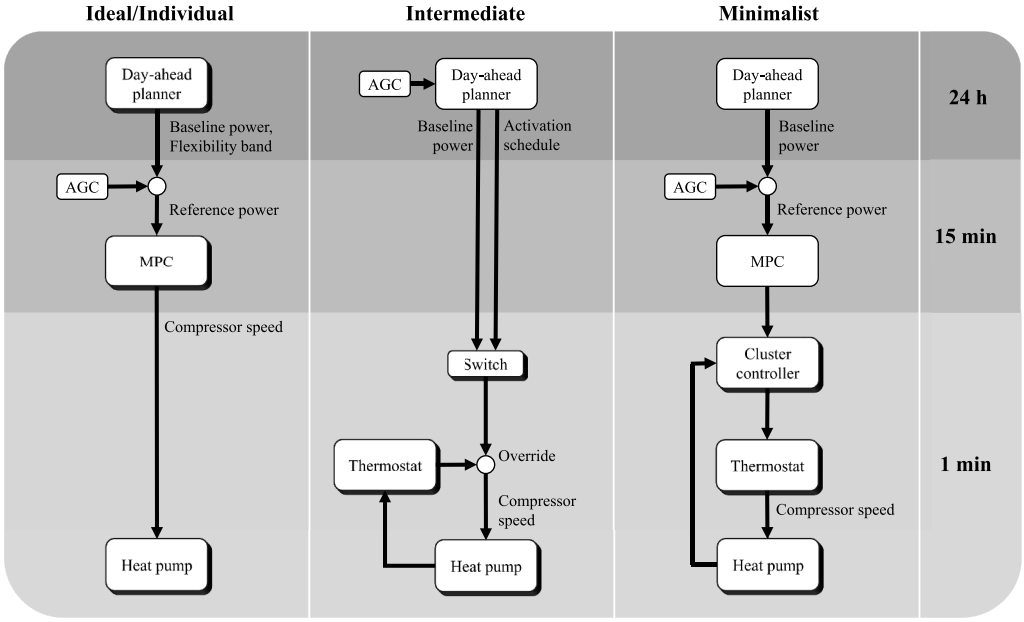


Fig. 4. Comparison of all four control schemes. Three-dimensional boxes indicate one instance for each building. Flat boxes indicate one instance for the whole cluster. Ideal and individual variants are equivalent in their general structure.

from the baseline power during execution.

$$\max_{x,u,\gamma} \quad \gamma \tag{1a}$$

$$\text{s.t.} \quad x_0 \text{ and } w_{[0:N_p-1]} \text{ known}$$

$$x^j_{k+1} = Ax^j_k + B_w w_k + B_u(u_k + \gamma v^j_k) \tag{1b}$$

$$T_{min} \le Cx^j_k \le T_{max} \tag{1c}$$

$$\gamma \le u_k \le 1 - \gamma \tag{1d}$$

$$\gamma \ge 0 \tag{1e}$$

$$\forall k \in \{0, \dots, N_p - 1\}$$

$$\forall j \in \{1, \dots, M\}$$

From the planner (1), we obtain an optimized baseline $P_{base} = P_{nom}u^*$ and flexibility band γ^*. Together with the ACG signal, we get the tracking reference $P_{ref,k} = P_{base,k} + \gamma^* a_k$. This reference is tracked by the subordinate MPC (2) with a shorter horizon of 6 h, i.e. $N_m = 24$. Since we use an actualized AGC signal at this point, the scenarios and flexibility band are not longer present in (2b) and (2c).

$$\min_{x,u} \quad \| P_{nom}u - P_{ref} \|_2^2 \tag{2a}$$

$$\text{s.t.} \quad x_0,\ w_{[0:N_m-1]} \text{ and } a_{[0:N_m-1]} \text{ known}$$

$$x_{k+1} = Ax_k + B_w w_k + B_u u_k \tag{2b}$$

$$T_{min} \le Cx_k \le T_{max} \tag{2c}$$

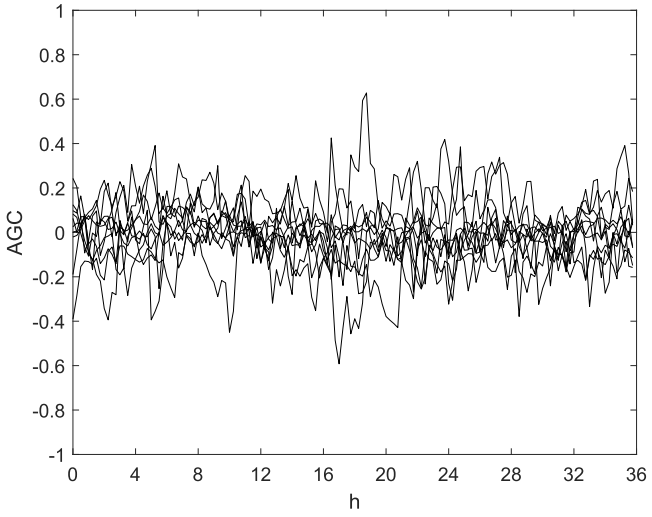


Fig. 5. The ten AGC scenarios for the day-ahead planner.

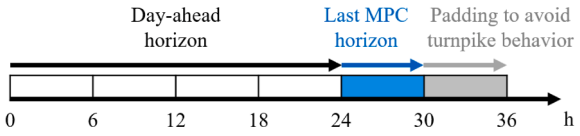


Fig. 6. Visualization of the 36 h planning horizon.

$$0 \le u_k \le 1 \qquad \forall k \in \{0, \dots, N_m - 1\} \tag{2d}$$

2.5. Individual control

The individual control variant works like the ideal variant, except that the perfect models and state estimates are replaced with identified sixth-order autoregressive models with exogenous inputs (ARX). These being input-output models, no state estimators are needed. Identification data is collected by applying random step inputs to the thermostats from mid-November through December. Even with this well-excited data, identifying models for the individual buildings proved challenging. Since the basic ARX methodology frequently resulted in models with slightly unstable poles, we enforce stability by constraining the autoregressive parameters to be larger or equal to zero and their sum to be smaller or equal to 0.999, forcing the state of the model to converge to zero over time[8]. The optimization problems for the day-ahead planner and MPC are the same as (1) and (2), except for the different models. We thus omit them for the sake of brevity.

2.6. Intermediate control

To fill the gap between the ideal and minimalist variants, we implement an intermediate control variant based on the work of [29,30]. While the ideal variant assumes full knowledge of the individual building states, and the minimalist variant assumes no knowledge thereof, this variant assumes that we measure the power of the individual buildings, but estimate their thermal states using the virtual storage model approach presented in [29], which consists of three parts: The baseline consumption model predicts a building's power consumption based on the weather forecast, and functions independently of the following two parts. The baseline state model predicts the thermal state of charge of a building, which is bound by the zone temperature constraints of 20 °C to 24 °C, during normal operation. Since the default controllers of the buildings aim to maintain a constant zone temperature, this model tends to converge toward a neutral point, but is offset by hot or cold weather. Lastly, the virtual storage model predicts how a buildings' thermal state changes, as a function of requests to increase or decrease its heating power. This is the component that enables the optimization of load shifting. For a more comprehensive description, we refer the reader to the corresponding papers[9].

To identify the baseline consumption and state models, both of which are kernel ridge regression (KRR) models, data of normal operation is collected from mid-November to mid-December. To identify the virtual storage model, randomized requests to increase or decrease the heating power are applied during the second half of December.

From this virtual storage model, we estimate a flexibility envelope for each building as follows: Given the predicted baseline power of one building denoted by $P \in \mathbb{R}^{N_b}$, we define the minimum and maximum relative power levels (with respect to the baseline) $\underline{R}^t \in \mathbb{R}^{N_b}$ and $\bar{R}^t \in \mathbb{R}^{N_b}$ that can be held for t time steps without violating the constraints of the identified virtual storage model. The flexibility envelope R^t is then defined as $[\underline{R}^t, \bar{R}^t]$. Each building can be activated once per day for up to t time steps. In the experiments, we use $N_i = 96$ as optimization horizon, with a 15 min time step. Furthermore, we chose $t = 15$, so we assume the availability of each building for up to 3.75 h. In line with [30], we define the scheduling problem (3), which simultaneously computes the γ at the start of each day.

$$\max_{b,\gamma} \quad \gamma \tag{3a}$$

$$\text{s.t.} \quad b_{i,k} \in \{0,1\} \tag{3b}$$

$$\sum_{k=0}^{N_b-1} b_{i,k} \le 1 \tag{3c}$$

$$\sum_{i=1}^{n_{bld}} \sum_{l=k_s}^{k} b_{i,l} \underline{R}^t_{i,l} \le \gamma a_k \tag{3d}$$

$$\sum_{i=1}^{n_{bld}} \sum_{l=k_s}^{k} b_{i,l} \bar{R}^t_{i,l} \ge \gamma a_k \tag{3e}$$

$$0 \le \gamma \le \min\left\{\min\{P_{base}\}, P_{nom} - \max\{P_{base}\}\right\} \qquad \forall k \in \{0, \dots, N_b - 1\}, \ \forall i \in \{1, \dots, n_{bld}\} \tag{3f}$$

where $b_{i,k}$ is the binary variable denoting whether building i is activated at step k, $k_s = \max\{k - t - 1, 0\}$ is the starting index to consider activation of a building in the last t time steps, n_{bld} denotes the number of buildings, P_{base} denotes the aggregate baseline consumption of all buildings, and P_{nom} denotes the aggregate nominal power of all buildings. Constraint (3c) ensures that each building is only activated at maximum once per day. Constraints (3d) and (3e) ensure that the flexibility of all activated buildings covers the request with the optimal γ for all time steps, both in negative and positive direction.

The reference P_{ref} is then tracked by disaggregating the relative request between all active buildings as follows, while non-active buildings consume according to their predicted baseline. For $a_k < 0$, we have

$$u_{i,k} = \frac{P_{i,k} + b_{i,k} \frac{\gamma a_k}{\underline{P}_{tot,k}}}{P_{nom,i}}, \quad i = 1, \dots, n_{bld}$$

with $P_{i,k}$ being the baseline consumption of building i at step k, and $\underline{P}_{tot,k} = \sum_{i=1}^{n_{bld}} \sum_{l=k_s}^{k} b_{i,l} \underline{R}^t_{i,l}$ being the total decrease potential of the active buildings. Equivalently, the dispatch can be defined for $a_k > 0$ with $\bar{P}_{tot,k} = \sum_{i=1}^{n_{bld}} \sum_{l=k_s}^{k} b_{i,l} \bar{R}^t_{i,l}$ being the total increase potential. These inputs are applied as long as they do not lead to temperature constraint violations in the building, in which case the internal controller tries to counteract the violation.

[8] Proof in Appendix A.

[9] The original formulation in [29] uses a risk parameter α, which we omit for the scope of this study.

Note that this approach does not determine the baseline consumption in the optimization procedure, but instead uses the predicted baseline for each building. Because of this, the maximum achievable values for γ are bounded as in constraint (3f). Additionally, we restrict the duration for which a building is available to provide flexibility per day. For these reasons, this variant is more conservative in its utilization of the flexibility potential, compared to the ideal and individual variants.

2.7. Minimalist control

For the minimalist variant, we treat the entire building cluster as one plant. Each building has a minimum and maximum room temperature, with the mean thereof treated as the default value. The control variable $\mu \in [-1, 1]$ for the frequency control is the fraction of buildings being instructed to shift their thermostat setpoint from the default value to either the maximum (positive value, increased power) or minimum (negative value, decreased power). Hence, we treat the buildings as thermostatically controlled loads. The output of the system is the aggregated power of all the heat pumps. The activation of the buildings is sequential. If more buildings need to be activated, the ones with the lowest total activation time up to that moment are activated first. If fewer buildings need to be activated, the ones that have been in a state of activation the longest are deactivated first.

To identify a corresponding model of the building cluster, data is collected by applying random steps to μ from mid-November through December. We use the Hammerstein-Wiener model (4) with no input nonlinearity, a model order of $n_{hw} = 8$, and an identified output saturation, constraining the output to a range from $P_{\min} = 0$ to $P_{\max} = 5.9\,\mathrm{MW}$. The *Matlab* command *nlhw* is used with a regularization weight of 100.

$$P_k = \begin{cases} P_{max}, & \text{if } \mathbf{1}^\top \chi_k > P_{max} \\ P_{min}, & \text{if } \mathbf{1}^\top \chi_k < P_{min} \\ \mathbf{1}^\top \chi_k, & \text{otherwise} \end{cases} \tag{4a}$$

$$\chi_{k+1} = \sum_{k=1}^{n_{hw}} B_\mu \odot \tilde{\mu}_k + B_w \odot \tilde{w}_k - F \odot \tilde{\chi}_k \tag{4b}$$

With $\chi \in \mathbb{R}^3$ the internal state of the model[10], the shorthand notation $\tilde{\cdot}_k$ for $\begin{pmatrix} \cdot_k & \cdots & \cdot_{k-n_{hw}+1} \end{pmatrix}$, and $\odot$ the element-wise multiplication. In the following, we denote this model as $P_k = f(\chi_k)$ and $\chi_{k+1} = g(\tilde{\chi}_k, \tilde{\mu}_k, \tilde{w}_k)$.

Unlike the other variants, we cannot optimize the size of the flexibility band for this one. We thus calculate a baseline consumption over the planning horizon N_p from the aforementioned model and the weather forecast, assuming $\mu = 0$. Based on the results of the ideal variant in Section 3, we heuristically choose a fixed $\gamma = 2000\,\mathrm{kW}$. The actual γ of the ideal variant is greater than that around 75 % of the time, so we consider 2000 kW to be a moderately conservative value. Should adding this value to the baseline estimation result in a violation of the known power limits, it gets cropped correspondingly. The resulting reference power P_{ref} is tracked with the MPC (5) by optimizing the constrained control variable μ of the building cluster.

$$\min_{P,\,\mu,\,\chi} \quad \|P - P_{ref}\|_2^2 \tag{5a}$$

$$\text{s.t.} \quad \tilde{\chi}_0,\ \tilde{\mu}_0, w_{[-n_{hw}+1:N_m-1]} \text{ and } a_{[0:N_m-1]} \text{ known}$$

$$P_k = f(\chi_k) \tag{5b}$$

$$\chi_{k+1} = g(\tilde{\chi}_k, \tilde{\mu}_k, \tilde{w}_k) \tag{5c}$$

$$\mu_{min} \le \mu_k \le \mu_{max} \tag{5d}$$

$$\forall k \in \{0, \ldots, N_m - 1\}$$

The initial state of the model $\tilde{\chi}_0 \in \mathbb{R}^{3\times n_{hw}}$ comes from a log of the most recent predictions of χ_1^*, which we obtain from solving the MPC problem. To account for the prediction errors, which are expected when fitting a simple model to a complex system, and the process noise, which stems from the heat pumps switching on and off, the optimized output (i.e. power) trajectory from the MPC is used as a setpoint for a cascading PI controller[11] with a 1 min time step. The proportional gain is $1/(6 \times P_{nom})$, the integral gain $1/(P_{nom} \times 60\,\mathrm{s})$ and the integrator is limited to a range of ± 1. Incidentally, this controller also addresses the state estimation problem, albeit in an inverted manner. While a state estimator accelerates, if its prediction falls behind the measured output (and vice versa), the cascading controller increases the control input, if the measured output falls behind the prediction. Trials with a dedicated state estimator, as would be standard practice, resulted in decreased control performance. We explain this with the aforementioned counteracting nature of the cascading controller and the state estimator.

3. Results

The key performance indicators of all four variants are compared in Table 1. The total γ_{tot} is the mean value $\bar{\gamma}$ integrated over the evaluation period of 59 days. We introduce the total energy shifted for frequency control E_{fc}, the default energy consumption E_{def}, and the activation ratio $r_{act} = \gamma_{tot}/E_{def}$. For our building cluster $E_{def} = 3989\,\mathrm{MW\,h}$ from January through February with constant thermostat setpoints of 22 °C. The tracking error is defined as $\varepsilon = |P_{act} - P_{ref}|$. The total error is $E_{err} = \sum_{k=1}^{N_{smp}} \varepsilon_k$ and the normalized average error $\hat{\varepsilon} = \frac{1}{N_{smp}} \sum_{k=1}^{N_{smp}} \varepsilon_k/\gamma_k$, with N_{smp} the number of samples. Analogously to the activation ratio, we also define the error ratio $r_{err} = E_{err}/E_{def}$. The computation time t_{comp} includes the building simulation and optimization, with the latter being the dominant component.

Unsurprisingly, the ideal variant has the highest cluster activation, the smallest tracking error and the highest computation time. A tracking error of zero might be expected for the ideal variant, however, the ideal model includes the dynamics of the air and temperature sensors, which are often faster than the time step of the MPC, which leads to numerical problems, resulting in suboptimal solutions. Moreover, the scenario-approach is only an approximation of the disturbances from the AGC signal. A particularly large request to shift loads may force an unforeseen correction. The individual variant has a slightly lower activation ratio and computation time, but a much larger tracking error of 1 %. In contrast to the ideal variant, this can be explained by the identified models, which are not a perfect match, and fail to capture some of the buildings' faster dynamics, but pose fewer numerical problems.

Compared to the variants with a separate optimization for each building, the more aggregated intermediate variant shows a significantly lower activation ratio, which can be explained by its more conservative approach to planning flexibility. However, it slightly outperforms the individual variant in terms of tracking error, which can also be explained with its more conservative approach. The total computation time is improved by a full order of magnitude, which is unsurprising, considering its more aggregated nature. The fully aggregated minimalist variant shows a higher activation ratio than the intermediate variant, which is due to its more aggressive, partially hand-tuned approach to the flexibility planning. As was expected, due to its indirect control of the heat pumps, it has by far the largest tracking errors. Its fully aggregated approach results in the fastest computation time in the comparison, being around three times faster than the intermediate variant.

The ideal, individual and minimalist variants are implemented in *Matlab* and executed using all eight cores of an *Intel Core i9-11900K* processor. The intermediate variant is a product under development at

[10] The *Matlab* command uses a model order equal to the number of inputs.

[11] In an effort to create an even simpler control scheme, we implemented a power tracking with a pure PI controller, instead of the MPC and the cascading controller. However, the tracking periodically broke down when the control variable reached its upper or lower limit. In contrast, the MPC leverages one of its inherent strengths, which is to account for constraints, and avoids this problem.

Table 1
Load shifting, tracking error, constraint violation and computation time for all variants.

	Unit	Ideal	Individual	Intermediate	Minimalist
$\bar{\gamma}$	kW	2129	2066	1656	1724
γ_{tot}	MWh	3015	2952	2344	2441
r_{act}	%	75.6	74.0	58.8	61.2
E_{fc}	MWh	305.0	288.1	213.2	243.4
E_{err}	MWh	4.2	25.8	20.7	33.8
r_{err}	%	0.11	0.65	0.52	0.85
$\hat{\epsilon}$	%	0.14	1.0	0.95	1.48
T^{up}_{viol}	Kh/d	0	0.01	0	0.76
T^{low}_{viol}	Kh/d	0.15	0.15	0.18	0.28
t_{comp}	s	60 430	51 849	5379	1805

the institution of the second author. For reasons of protecting intellectual property, the corresponding simulations were contributed by them, using *Python* and a ten-core *Intel Core i7-1355U* processor, which has roughly 60 % of the performance of the *i9-11900K*[12]. The computation times would likely be different on the hardware used for actual deployment, but for the scope of this study, the values reported in Table 1 provide a reasonable approximation of their relative computational cost.

Lastly, T^{up}_{viol} and T^{low}_{viol} denote the average upper and lower temperature constraint violations, expressed in K h per day and building. Even with default control, some buildings fail to maintain $T_{min} = 20\,°C$ when the ambient temperature drops very low. The resulting temperature violations are equal to the ideal variant, meaning that the ideal variant introduces no additional constraint violation. The individual variant adds slight violations of the upper bound. The intermediate variant adds a similarly negligible violation to the lower bound. The minimalist variant adds more significant violations of both bounds, which is unsurprising, considering the thermostat setpoints are shifted to the bounds and the low-level heat pump controllers have a hysteresis of ±0.5 K. Nevertheless, we consider a combined constraint violation of 1.04 K h per day to be acceptable.

The three-day extract of active frequency control for all four variants in Fig. 7 reveals additional insights into their differences. The ideal and individual variants, which aim to utilize each building as much as possible, show some jittering on their flexibility bands, compared to the intermediate and minimalist variants, which lump all the buildings into one collective optimization, resulting in a smoother trajectory. In the bottom plot around day 27.4, the minimalist variant fails to track the positive spike in the AGC signal, which is unsurprising, considering it is the only variant relying on purely indirect control of the heat pumps. Some grid operators impose rules about the allowed normalized tracking error. Fig. 8 shows that no variant can contain all errors to less than 1 %. Among the non-ideal variants, only 50 % to 80 % of the errors are below 1 %, with the minimalist variant performing by far the worst. If a more relaxed threshold of 5 % error is considered, all non-ideal variants converge toward a probability of around 95 %, while the ideal variant reaches a reliability of over 99 %. However, the performance of the ideal variant is of little practical relevance in this context, only serving as a high-end benchmark for the non-ideal variants.

3.1. Cost estimate

In this section, we provide an estimate of the cost-effectiveness of each variant, assuming a number of generally favorable conditions, which are specified below and whose impact on the conclusions is discussed in Section 4. First, we introduce Eq. (6) for the expected financial gain per year G.

$$G = c_{fc}\,\gamma_{tot} - c_{err}\,E_{err} - C_{cen} - n_{bld}\,C_{bld} \tag{6}$$

Table 2
Annual components of cost estimate.

	Unit	Individual	Intermediate	Minimalist
Central computing	CHF	250 - 500	0	250 - 500
Cluster identification	CHF	0	0	500 - 1000
Building computing	CHF/bld	250 - 500	0.25 - 0.5	0
Building identification	CHF/bld	10 - 20	10 - 20	0
Administration	CHF/bld	20 - 50	20 - 50	20 - 50
Remuneration c_{fc}	CHF/MWh	10	10	10
Penalty c_{err}	CHF/MWh	100	100	100

With the remuneration for flexibility provided c_{fc}, the financial penalty for the tracking error c_{err}, the centralized system costs C_{cen} and the system costs per building C_{bld}. The variables γ_{tot} and E_{err} refer to annual values in this section. Table 2 shows cost estimates in Swiss Francs[13] (CHF) for the centralized and building-level system components, and the payment structure for frequency control. The ideal variant is omitted since we do not believe that there is a meaningful cost estimate for a perfect building model and state estimator.

We assume that the building cluster is managed by a pre-existing utility provider, so the technical staff and office space are already available. Based on the simulation times in Table 1, we account for one CPU core in an existing server system to execute all the centralized computation. The minimalist variant also requires the identification and maintenance of a cluster model. Since the methodology described in Section 2.7 can be mostly automated, we assume between five and ten work hours per year at a rate of 100 CHF/h.

On the building level, we assume that basic *smart home* capabilities are already given. This includes an internet connection, a thermostat, and a power meter for the heat pump[14]. The necessary computation power per building varies greatly between the variants. We assume one CPU core on a cloud service per building for the individual variant, one core per 1000 buildings for the intermediate variant, and one centralized core for the entire cluster for the minimalist variant. The cores for the intermediate variant are expressed as a cost per building, but would actually be centralized. The individual and intermediate variants require the identification and maintenance of a model for each building. Based on our experiences implementing Sections 2.5 and 2.6, we assume one work hour per year at a rate of 100 CHF/h for 10 % to 20 % of the buildings to manage potential malfunctions, respectively inaccurate models. The last cost component is a general administration cost for every building enrolled in the cluster. The remuneration and penalty cost are based on data from the Swiss national grid operator [55]. For better generalizability, we use the aforementioned two parameters to approximate the true payment mechanism, which is specific to the grid operator, and may differ significantly from the methods used by other providers.

Since the cost-effectiveness depends on the number and size of the buildings in the cluster, which may vary greatly in reality, we aim to decouple the cost estimate from the specifications at hand by dividing the total energy of flexibility provided into the activation ratio, the number of buildings and the averaged default energy consumption per building $\gamma_{tot} = r_{act}\,n_{bld}\,\bar{E}_{def}$, and analogously the error $E_{err} = r_{err}\,n_{bld}\,\bar{E}_{def}$. For the scope of this estimate, we treat the activation and error ratios in Table 1 as variant-specific constants. Furthermore, we assume the number of buildings in the cluster to be large enough to neglect the central

[12] https://www.cpubenchmark.net/compare/5317vs3904/Intel-i7-1355U-vs-Intel-i9-11900K

[13] On the 25th of February, 2025, the exchange rate according to *Google Finance* is 1 CHF = 1.11 USD = 1.07 EUR.

[14] Without this assumption, we almost never expect a profit, since their installation by a technician alone would likely cost over 1000 CHF. With an amortization time of 10 years, this would add over 100 CHF/a per building.

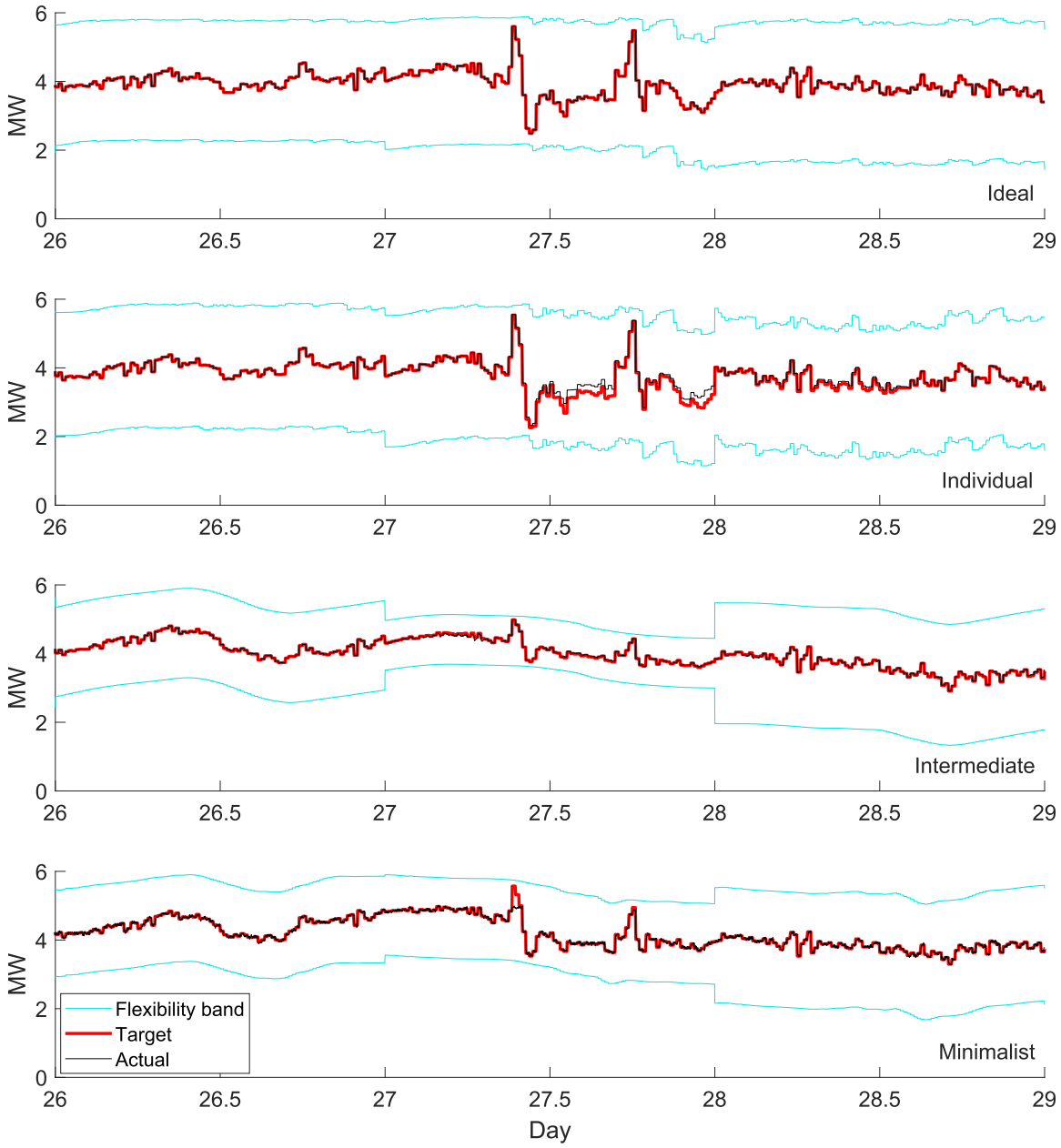


Fig. 7. Extract of flexibility band and tracking performance for all four variants.

cost components[15]. We thus obtain the simplified expected profit per building (7).

$$\bar{G} = (c_{fc}\, r_{act} - c_{err}\, r_{err})\, \bar{E}_{def} - C_{bld} \tag{7}$$

Fig. 9 shows the expected annual profit range per building for the three non-ideal variants as a function of the building's annual energy consumption. As a reference, the average default annual energy consumption of a building in our cluster is 10.6 MW h, a newly built single-family house in Switzerland is around 2 MW h. With the assumed remuneration, the expected profits for the individual variant are negative. The intermediate and minimalist variant are expected to break even at a size of around 10 MWh/a. We thus invert the question and calculate the remuneration necessary to break even for the intermediate and minimalist variants in Fig. 10, maintaining a penalty-to-remuneration ratio of 10. It can be seen that the remunerations would have to increase by a factor of around two to seven to break even with small buildings. It is unclear how likely such a stark increase of the remuneration is in the near future. At the current time, we thus conclude that the expected profit is only positive when using the simpler control frameworks with larger buildings.

4. Discussion

For the scope of this study, we made a number of simplifying assumptions. The models generated from [41] are linear, which is not a true representation of real buildings and heat pumps. Moreover, the authors of [41] state that many of the building models have been identified from sub-optimal data, and show physically implausible parameters. While we made an effort to exclude extreme outliers, as described

[15] For the minimalist variant with 1000 buildings, using the high-end estimate for central costs and the low-end estimate for building-level costs, the central costs account for 7 % of the total costs. We consider 1000 buildings to be toward the lower end of reasonable cluster sizes.

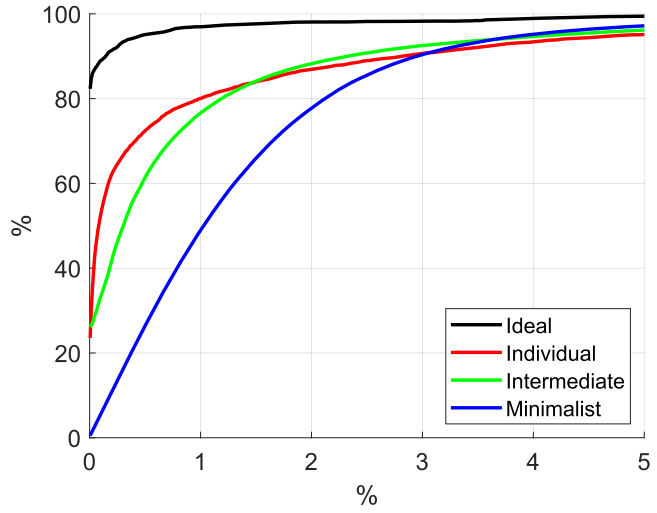


Fig. 8. Cumulative distribution function of the normalized tracking errors.

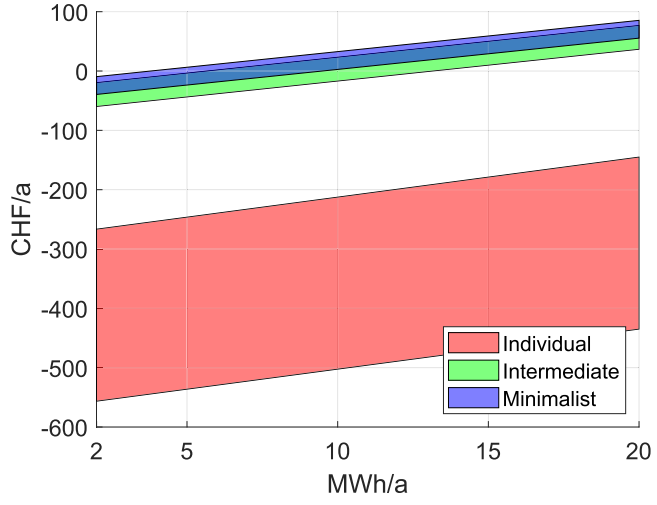


Fig. 9. Range of expected annual profit as a function of building energy consumption and control variant.

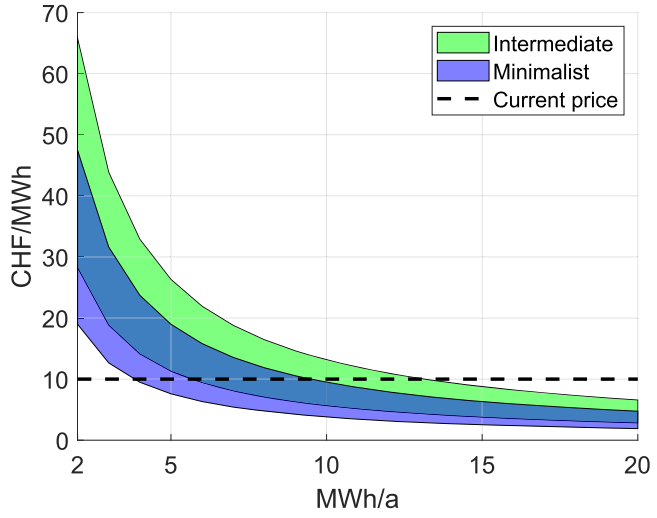


Fig. 10. Range of flexibility remuneration necessary to break even as a function of building energy consumption and control variant.

in Section 2.1, some questions regarding the accuracy of the models remain. Among them are the very powerful heating systems of the building cluster, running at around 30 % of their cumulative maximum power on average during ambient temperatures of -20°C to -10°C. On the other end of the spectrum are the some building models that occasionally fail to maintain a zone temperature of 20 °C. However, we argue that for the purposes of this study, a few inaccurate building models are acceptable, as long as the cluster is sufficiently diverse. Buildings failing to maintain their target zone temperature under extreme weather conditions cannot be ruled out in the real world, thus the control algorithms should be resilient to it to a reasonable extent. Furthermore, the prevalent construction style in North America differs from that of Switzerland, which we base the grid service framework and cost estimates on. A more geographically appropriate building cluster may yield different result, although it is unclear to which extent. While building models of higher quality and more representative of the Swiss (or European) building stock would have been preferred, no corresponding dataset is publicly available at the current time, to the best of our knowledge.

In one particular aspect, the simplifying assumptions made for our simulation model actually aggravate a problem, potentially to the detriment of one of the control variants: The process noise from the switching of the heat pumps for the minimalist variant is rather strong, because we use binary on-off controllers on the building-level, and because the cluster size is moderate with 1000 buildings. Heat pumps with internal variable-speed controllers, as well as a cluster with more buildings would mitigate the noise. In turn, the minimalist variant would likely be more negatively affected by a cluster with fewer buildings, compared to the other variants.

Regarding the boundary conditions for the simulations, two simplifying assumptions made by us are the ideal forecasts of the weather and the AGC signal, since we do not have access to corresponding datasets with real forecasting errors, and we are not aware of an established method to artificially generate such data for either the weather or the AGC signal. The control performance may have benefited from these idealized conditions, although estimating the impact thereof, compared to more realistic forecasts, is no trivial task. Since the ideal, individual and minimalist variants use the same ideal forecasts, they are unlikely to have caused a bias in the direct comparison of their performance. An exception to this is the intermediate variant, which uses a 24 h forecast of the AGC signal, as previously mentioned in Section 2.3. Its relative performance, compared to the other three variants, would likely decrease if realistic AGC forecasts were used.

As indicated in Sections 2 and 3, we believe there is room for improvement for all the data-driven system identification methods used in this study. The same model structure is applied to every building in the individual variant. Considering the diversity of building's thermal behavior, one may consider an iterative tailoring of the models for each building. Different model orders and stability constraints could be evaluated for each model. While less important for the simplified building models used in this study, nonlinear models may be worth considering for real buildings. Based on our experiences with the intermediate and minimalist variants, we propose KRR or Hammerstein-Wiener models with input and/or output saturations as potential candidates. Another direction would be to use different time steps depending on the time constants of the building. For the sake of aggregation, it may be beneficial to choose steps sizes which are multiples of each other, i.e. 15 min, 7.5 min and 5 min.

For the KRR models used for the intermediate variant, some adjustments were made to the hyperparameters, after we found the models to perform poorly for some buildings, which did not happen to the same extent in the previous studies conducted with this method [29,30], indicating that there is potential left to improve its robustness. For the minimalist variant, we identify two potential areas for improvement: Since process noise is inherent to the indirect control of the heat pumps, a stochastic model may help to manage it more effectively. Furthermore, the heuristically chosen, fixed flexibility band of this variant may be

replaced with a dynamic solution in the future, similar to the virtual state model from Section 2.6. These two aspects may interfere with each other in non-trivial ways. It may even be sensible to merge the quantification of the process noise with the optimization of the flexibility band. However, most of the alternative methodologies proposed above would increase the computational requirements. Considering the significance of the computational component of the cost estimates, the cost-effectiveness for each of these alternatives would have to be reevaluated, even if superior performance were achieved. All four variants use hand-tuned hyperparameters. It may be possible to improve the performance of each by introducing some form of auto-tuning. Because the sphere of potentially applicable algorithms is extensive and intricate, we refrain from speculating about the extent of their hypothetical impact on the performance.

Should one desire to practically implement the minimalist variant in its current form, i.e. without any of the aforementioned extensions, a simpler replacement for the fixed flexibility band derived from the results of the ideal variant would have to be found. A possible solution is to commence operation with a conservatively chosen band, maybe 10 % of the nominal power. Subsequently, this band may be gradually widened until increased tracking errors for spikes in the AGC signal begin to emerge, at which point the system has probably reached its limit.

For the cost estimate, we chose values based on our experience, and some consultation by industrial partners, in Switzerland. They may be less appropriate in the future or for different locations. We leave it to the interested reader to repeat our calculations with alternative values, based on their own experiences and estimations. Moreover, we assume that a connected thermostat and a power meter for the heat pump are already present in each building. Without these assumptions, the costs per building are far too high, considering the low remuneration for the provided load shifting. Currently, this assumption strongly limits on the pool of suitable buildings, but may do so less in the future, as *smart home* automation systems become more prevalent [56–58]. Lastly, a hard limit on the maximum allowed tracking error could pose a significant problem for our methods, since the error distributions shown in Fig. 8 have high variances, including some normalized tracking errors above 5 %.

Calling back to the findings of the literature search in Section 1 regarding the reliability, costs and practical validation of demand response with buildings, we believe that there is significant value in a continuation of this line of research, further improving the control performance and particularly the cost-effectiveness. First in simulations, but eventually in real buildings as well.

5. Conclusion

Comparing four different methodologies for secondary frequency control with a building cluster, we find an unsurprising positive relationship between the complexity and performance of each variant under the assumed regulatory framework, with the ideal variant having a normalized tracking error of 0.15 %, compared to 1 % for the individual variant, 0.95 % for the intermediate variant and 1.48 % for the minimalist variant. Similarly, we find the ideal variant to have the highest average activation of the building cluster's load shifting potential with an average flexibility band of 2129 kW, compared to 2066 kW, 1656 kW and 1724 kW for the individual, intermediate and minimalist variants respectively, compared to a nominal power of 5.9 MW. Should strict limits on the maximum allowed tracking error be applied, none of the non-ideal variants are likely to meet the corresponding accuracy thresholds.

Taking the estimated costs of the three non-ideal variants into account, we find the minimalist and intermediate variant to deliver their performance at a fraction of the cost of the individual variant, mostly stemming from the difference in computational requirements. However, with the assumed remunerations for providing frequency control, we expect the financial gains to be insufficient to offset the system costs in most cases. We expect a positive net profit when using the simple control variants for buildings with a large energy consumption, because they have the potential to shift greater loads, while we assume the system costs per building to be independent of size. For small, energy-efficient buildings, the remunerations would have to increase by a factor of two to seven to result in a net profit within the specifics of our simulations, even with the most cost-effective control variants.

CRediT authorship contribution statement

Manuel Koch: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization; **Paul Scharnhorst:** Writing – original draft, Software, Methodology, Formal analysis; **Parantapa Sawant:** Writing – review & editing, Project administration, Funding acquisition; **Colin N. Jones:** Writing – review & editing, Project administration, Methodology, Funding acquisition, Conceptualization.

Data availability

The authors do not have permission to share data.

Declaration of competing interest

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

Acknowledgments

The authors would like to thank Vaibhav Gupta and Alireza Karimi from EPFL for their advice regarding the identification and implementation of the Hammerstein-Wiener models. Part of the research published in this report was carried out with the support of the Swiss Federal Office of Energy SFOE as part of the SWEET Lantern project.

Appendix A. Proof of stability constraint

Theorem 1. *Consider the autoregressive system*

$$y_{k+1} = a_0\, y_k + \cdots + a_n\, y_{k-n}. \tag{A.1}$$

If $\sum_{i=0}^{n} a_i < 1$ *and* $a_i \geq 0, i = 1, \ldots, n$, *then the discrete-time dynamical system given by*

$$x_{k+1} = A x_k \tag{A.2}$$

with $x_k = \begin{pmatrix} y_k & y_{k+1} & \cdots & y_{k-n} \end{pmatrix}^\top$ *and* $A = \begin{bmatrix} a_0 & a_1 & \ldots & a_{n-1} & a_n \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$ *is asymptotically stable.*

Proof. To show that (A.2) is asymptotically stable, we need to show that $|\lambda_i| < 1$ for all eigenvalues λ_i of A.

First, we use that $\rho(A) \leq \|A\|$ for any natural matrix norm $\|\cdot\|$, where $\rho(A)$ denotes the spectral radius of A. Since $\|A\|_\infty = \max_{\|x\|_\infty = 1} \|Ax\|_\infty$ is a natural matrix norm, we have $\rho(A) \leq 1$.

Next, we assume that $a_n > 0$. This is not a restrictive assumption, as we could otherwise consider the largest i with non-zero a_i and prove stability for the reduced system. $a_n > 0$ guarantees that the matrix A is irreducible. Together with the non-negativity of the entries of A, we can use the Perron-Frobenius theorem, stating that $\rho(A) = \bar{\lambda}$ where $\bar{\lambda}$ is a simple eigenvalue of A with a corresponding eigenvector v with positive entries.

Assume $\bar{\lambda} = 1$. Let $v_k = \max_{i=0,\ldots,n} v_i$ be the maximum entry of v and consider the scaled vector $\bar{v} = \frac{1}{v_k} v$. Note that $\bar{v}$ is still an eigenvector of A with eigenvalue $\bar{\lambda}$, and $\bar{v}_k = 1, \bar{v}_j < 1$ for $j \neq k$.

Since we assume $\bar{\lambda} = 1$, we can use the eigenvalue equation recursively to obtain

$$A^k \bar{v} = \bar{\lambda}\bar{v} = \bar{v}. \tag{A.3}$$

We focus on the k-th row of the equation. Due to the structure of A, we have $(A^k)_k = \begin{pmatrix} a_0 & a_1 & \dots & a_{n-1} & a_n \end{pmatrix}$. Expanding the k-th row of the left-hand side of (A.3), we get

$$(A^k \bar{v})_k = \begin{pmatrix} a_0 & a_1 & \dots & a_{n-1} & a_n \end{pmatrix} \bar{v} = \sum_{i=0}^{n} a_i \bar{v}_i < \sum_{i=0}^{n} a_i < 1 = \bar{v}_k. \tag{A.4}$$

This leads to a contradiction with the eigenvalue property. Therefore $\bar{\lambda} < 1$, which concludes the proof. □

References

[1] O.W.i. Data, Electricity Mix, (2020), https://ourworldindata.org/electricity-mix.
[2] P. Grünewald, J. Torriti, Demand response from the non-domestic sector: early UK experiences and future opportunities, Energy Policy 61 (2013) 423–429. https://doi.org/10.1016/j.enpol.2013.06.051
[3] H.C. Gils, Economic potential for future demand response in germany – modeling approach and case study, Appl. Energy 162 (2016) 401–415. https://doi.org/10.1016/j.apenergy.2015.10.083
[4] B.f. Energie, Energiestrategie 2050, (2018), https://www.bfe.admin.ch/bfe/de/home/politik/energiestrategie-2050.html.
[5] E. Commission, A European Green Deal, (2024), https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en.
[6] M.T.H. van Vliet, J. Sheffield, D. Wiberg, E.F. Wood, Impacts of recent drought and warm years on water resources and electricity supply worldwide, Environ. Res. Lett. 11 (12) (2016) 124021. https://doi.org/10.1088/1748-9326/11/12/124021
[7] J. Jääskeläinen, N. Veijalainen, S. Syri, M. Marttunen, B. Zakeri, Energy security impacts of a severe drought on the future finnish energy system, J. Environ. Manage. 217 (2018) 542–554. https://doi.org/10.1016/j.jenvman.2018.03.017
[8] Lazard, 2023 Levelized cost of energy+, 2023, https://www.lazard.com/research-insights/2023-levelized-cost-of-energyplus/.
[9] X. Sun, H. Hao, P. Hartmann, Z. Liu, F. Zhao, Supply risks of lithium-Ion battery materials: an entire supply chain estimation, Mater. Today Energy 14 (2019) 100347. https://doi.org/10.1016/j.mtener.2019.100347
[10] B. Parrish, P. Heptonstall, R. Gross, B.K. Sovacool, A systematic review of motivations, enablers and barriers for consumer engagement with residential demand response, Energy Policy 138 (2020) 111221. https://doi.org/10.1016/j.enpol.2019.111221
[11] Enerdata, Residential buildings: energy efficiency & consumption evolution in Europe, 2021, https://www.enerdata.net/publications/executive-briefing/households-energy-efficiency.html.
[12] U.S.E.I. Administration, U.S. Energy Facts Explained - Consumption and Production, (2024a), https://www.eia.gov/energyexplained/us-energy-facts/.
[13] U.S.E.I. Administration, Use of energy in commercial buildings in depth, (2023b), https://www.eia.gov/energyexplained/use-of-energy/commercial-buildings-in-depth.php.
[14] U.S.E.I. Administration, Use of energy in homes, (2023c), https://www.eia.gov/energyexplained/use-of-energy/homes.php.
[15] B.f. Energie, Energieverbrauch nach Verwendungszweck, 2025, https://www.bfe.admin.ch/bfe/de/home/versorgung/statistik-und-geodaten/energiestatistiken/energieverbrauch-nach-verwendungszweck.html.
[16] O. Abedrabboh, M. Koç, Y. Biçer, Sustainability performance of space-cooling technologies and approaches, Energy Sources Part A 44 (2022) 9017–9042.
[17] E. Elnagar, S. Pezzutto, B. Duplessis, T. Fontenaille, V. Lemort, A comprehensive scouting of space cooling technologies in europe: key characteristics and development trends, Renew. Sustain. Energy Rev. 186 (2023) 113636. https://doi.org/10.1016/j.rser.2023.113636
[18] B.f. Statistik, Bau- und Wohnungswesen, Gebäude, Energiebereich, (2025), https://www.bfs.admin.ch/bfs/de/home/statistiken/bau-wohnungswesen/gebaeude/energiebereich.html.
[19] I.E. Agency, The future of heat pumps – analysis, 2022, https://www.iea.org/reports/the-future-of-heat-pumps.
[20] A. Wang, R. Li, S. You, Development of a data driven approach to explore the energy flexibility potential of building clusters, Appl. Energy 232 (2018) 89–100. https://doi.org/10.1016/j.apenergy.2018.09.187
[21] M. Hu, F. Xiao, Quantifying uncertainty in the aggregate energy flexibility of high-rise residential building clusters considering stochastic occupancy and occupant behavior, Energy 194 (2020) 116838. https://doi.org/10.1016/j.energy.2019.116838
[22] F. Plaum, R. Ahmadiahangar, A. Rosin, J. Kilter, Aggregated demand-side energy flexibility: a comprehensive review on characterization, forecasting and market prospects, Energy Rep. 8 (2022) 9344–9362. https://doi.org/10.1016/j.egyr.2022.07.038
[23] R. Jurjevic, T. Zakula, Demand response in buildings: a comprehensive overview of current trends, approaches, and strategies, Buildings 13 (10) (2023) 2663. Publisher: Multidisciplinary Digital Publishing Institute, https://doi.org/10.3390/buildings13102663
[24] T.T. Gorecki, L. Fabietti, F.A. Qureshi, C.N. Jones, Experimental demonstration of buildings providing frequency regulation services in the swiss market, Energy Build. 144 (2017) 229–240. https://doi.org/10.1016/j.enbuild.2017.02.050
[25] M. Koch, P. Sawant, R. Eismann, C.N. Jones, A commissioning-oriented approach to data-driven modeling of buildings with heat pumps for predictive demand response (2025) 113016. https://doi.org/10.1016/j.jobe.2025.113016
[26] J. Shi, Y. Lian, C. Salzmann, C.N. Jones, Adaptive data-driven prediction in a building control hierarchy: a case study of demand response in Switzerland 333 (2025) 115498. https://doi.org/10.1016/j.enbuild.2025.115498
[27] R. D'hulst, W. Labeeuw, B. Beusen, S. Claessens, G. Deconinck, K. Vanthournout, Demand response flexibility and flexibility potential of residential smart appliances: experiences from large pilot test in belgium, Appl. Energy 155 (2015) 79–90. https://doi.org/10.1016/j.apenergy.2015.05.101
[28] J. Gasser, H. Cai, S. Karagiannopoulos, P. Heer, G. Hug, Predictive energy management of residential buildings while self-reporting flexibility envelope, Appl. Energy 288 (2021) 116653. https://doi.org/10.1016/j.apenergy.2021.116653
[29] P. Scharnhorst, B. Schubnel, R.E. Carrillo, P.J. Alet, C.N. Jones, Uncertainty-aware flexibility envelope prediction in buildings with controller-agnostic battery models, in: 2023 American Control Conference (ACC), 2023, pp. 583–590. https://doi.org/10.23919/ACC55779.2023.10156041
[30] P. Scharnhorst, B. Schubnel, R.E. Carrillo, P.J. Alet, C.N. Jones, Risk-aware scheduling and dispatch of flexibility events in buildings, Sustain. Energy Grids Netw. 39 (2024) 101512. https://doi.org/10.1016/j.segan.2024.101512
[31] J. Rousseau, H. Cai, P. Heer, K. Orehounig, G. Hug, Uncertainty-aware energy flexibility quantification of a residential building, in: 2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE), 2023, pp. 1–6. https://doi.org/10.1109/ISGTEUROPE56780.2023.10407489
[32] K. Xie, H. Hui, Y. Ding, Review of modeling and control strategy of thermostatically controlled loads for virtual energy storage system, Prot. Control Mod. Power Syst. 4 (4) (2019) 1–13. https://doi.org/10.1186/s41601-019-0135-3
[33] S. Nema, V. Prakash, H. Pandžić, The role of thermostatically controlled loads in power system frequency management: a comprehensive review, Sustain. Energy Grids Netw. 42 (2025) 101680. https://doi.org/10.1016/j.segan.2025.101680
[34] R. Nematirad, M.M. Ardehali, A. Khorsandi, A. Mahmoudian, Optimization of residential demand response program cost with consideration for occupants thermal comfort and privacy, IEEE Access 12 (2024) 15194–15207. https://doi.org/10.1109/ACCESS.2024.3358404
[35] F.L. Müller, B. Jansen, Large-scale demonstration of precise demand response provided by residential heat pumps, Appl. Energy 239 (2019) 836–845. https://doi.org/10.1016/j.apenergy.2019.01.202
[36] R. El Geneidy, B. Howard, Contracted energy flexibility characteristics of communities: analysis of a control strategy for demand response, Appl. Energy 263 (2020) 114600. https://doi.org/10.1016/j.apenergy.2020.114600
[37] Z.E. Lee, Q. Sun, Z. Ma, J. Wang, J.S. MacDonald, K. Max Zhang, Providing grid services with heat pumps: a review, ASME J. Eng. Sustain. Build. Cities 1 (1) (2020). https://doi.org/10.1115/1.4045819
[38] K. Kaspar, M. Ouf, U. Eicker, A critical review of control schemes for demand-side energy management of building clusters, Energy Build. 257 (2022) 111731. https://doi.org/10.1016/j.enbuild.2021.111731
[39] F. D'Ettorre, M. Banaei, R. Ebrahimy, S.A. Pourmousavi, E.M.V. Blomgren, J. Kowalski, Z. Bohdanowicz, B. Łopaciuk-Gonczaryk, C. Biele, H. Madsen, Exploiting demand-side flexibility: state-of-the-art, open issues and social perspective, Renew. Sustain. Energy Rev. 165 (2022) 112605. https://doi.org/10.1016/j.rser.2022.112605
[40] J. Le Dréau, R.A. Lopes, S. O'Connell, D. Finn, M. Hu, H. Queiroz, D. Alexander, A. Satchwell, D. Österreicher, B. Polly, A. Arteconi, F. de Andrade Pereira, M. Hall, T. Kırant-Mitić, H. Cai, H. Johra, H. Kazmi, R. Li, A. Liu, L. Nespoli, M.H. Saeed, Developing energy flexibility in clusters of buildings: a critical analysis of barriers from planning to operation, Energy Build. 300 (2023) 113608. https://doi.org/10.1016/j.enbuild.2023.113608
[41] C. Vallianos, J. Candanedo, A. Athienitis, Thermal modeling for control applications of 60,000 homes in North America using smart thermostat data, Energy Build. 303 (2024) 113811. https://doi.org/10.1016/j.enbuild.2023.113811
[42] Swissgrid, Principles of ancillary services products, 2022, https://www.swissgrid.ch/dam/swissgrid/customers/topics/ancillary-services/as-documents/D220824-AS-Products-V19-en.pdf.
[43] P. Zhao, G.P. Henze, S. Plamp, V.J. Cushing, Evaluation of commercial building HVAC systems as frequency regulation providers, Energy Build. 67 (2013) 225–235. https://doi.org/10.1016/j.enbuild.2013.08.031
[44] I. Beil, I. Hiskens, S. Backhaus, Frequency regulation from commercial building HVAC demand response, Proc. IEEE 104 (4) (2016) 745–757. https://doi.org/10.1109/JPROC.2016.2520640
[45] M.M. Olama, T. Kuruganti, J. Nutaro, J. Dong, Coordination and control of building HVAC systems to provide frequency regulation to the electric grid, Energies 11 (7) (2018) 1852. https://doi.org/10.3390/en11071852
[46] H. Wang, S. Wang, R. Tang, Development of grid-responsive buildings: opportunities, challenges, capabilities and applications of HVAC systems in non-residential buildings in providing ancillary services by fast demand responses to smart grids, Appl. Energy 250 (2019) 697–712. https://doi.org/10.1016/j.apenergy.2019.04.159
[47] H. Wang, S. Wang, A disturbance compensation enhanced control strategy of HVAC systems for improved building indoor environment control when providing power grid frequency regulation, Renew. Energy 169 (2021a) 1330–1342. https://doi.org/10.1016/j.renene.2021.01.102
[48] H. Wang, S. Wang, A hierarchical optimal control strategy for continuous demand response of building HVAC systems to provide frequency regulation service to smart

power grids, Energy 230 (2021b) 120741. https://doi.org/10.1016/j.energy.2021.120741
[49] T.T. Gorecki, Predictive Control Methods for Building Control and Demand Response, Ph.D. thesis, EPFL, 2017. https://doi.org/10.5075/epfl-thesis-7738
[50] M.B. Saltık, L. Özkan, J.H.A. Ludlage, S. Weiland, P.M.J. Van den Hof, An outlook on robust model predictive control algorithms: reflections on performance and computational aspects, J. Process Control 61 (2018) 77–102. https://doi.org/10.1016/j.jprocont.2017.10.006
[51] G.C. Calafiore, L. Fagiano, Robust model predictive control via scenario optimization, IEEE Trans. Automat. Contr 58 (1) (2013) 219–224. https://doi.org/10.1109/TAC.2012.2203054
[52] A. Parisio, M. Molinari, D. Varagnolo, K.H. Johansson, A scenario-based predictive control approach to building HVAC management systems, in: 2013 IEEE International Conference on Automation Science and Engineering (CASE), 2013, pp. 428–435. https://doi.org/10.1109/CoASE.2013.6654024
[53] X. Zhang, K. Margellos, P. Goulart, J. Lygeros, Stochastic model predictive control using a combination of randomized and robust optimization, in: 52nd IEEE Conference on Decision and Control, 2013a, pp. 7740–7745. https://doi.org/10.1109/CDC.2013.6761118
[54] X. Zhang, G. Schildbach, D. Sturzenegger, M. Morari, Scenario-based MPC for energy-efficient building climate control under weather and occupancy uncertainty, in: 2013 European Control Conference (ECC), 2013b, pp. 1029–1034. https://doi.org/10.23919/ECC.2013.6669664
[55] SwissGrid, Tenders, 2025, https://www.swissgrid.ch/en/home/customers/topics/ancillary-services/tenders.html.
[56] Statistia, Smart home - statistics & facts, 2024, https://www.statista.com/topics/2430/smart-homes/#topicOverview.
[57] I.Q. Electro, Smart home statistics by device category, most owned brands by country and facts, 2024, https://electroiq.com/stats/smart-home-statistics/.
[58] GreenMatch, Smart home statistics: key insights and trends, 2024, https://www.greenmatch.co.uk/blog/smart-home-statistics.