Journal of Building Performance Simulation

ISSN: 1940-1493 (Print) 1940-1507 (Online) Journal homepage: www.tandfonline.com/journals/tbps20

Smart building HVAC control challenge: experience and solutions from the ADRENALIN project

Harald Taxt Walnum, Balázs András Tolnai , David Blum , Jicheng Shi , Colin N. Jones , Deep Dessai , Wenbin Wang , Wenjie Xu , Sebastien Gros & Igor Sartori

To cite this article: Harald Taxt Walnum, Balázs András Tolnai , David Blum , Jicheng Shi , Colin N. Jones , Deep Dessai , Wenbin Wang , Wenjie Xu , Sebastien Gros & Igor Sartori (30 Apr 2026): Smart building HVAC control challenge: experience and solutions from the ADRENALIN project, Journal of Building Performance Simulation, DOI: 10.1080/19401493.2026.2656363

To link to this article: https://doi.org/10.1080/19401493.2026.2656363

Published online: 30 Apr 2026.

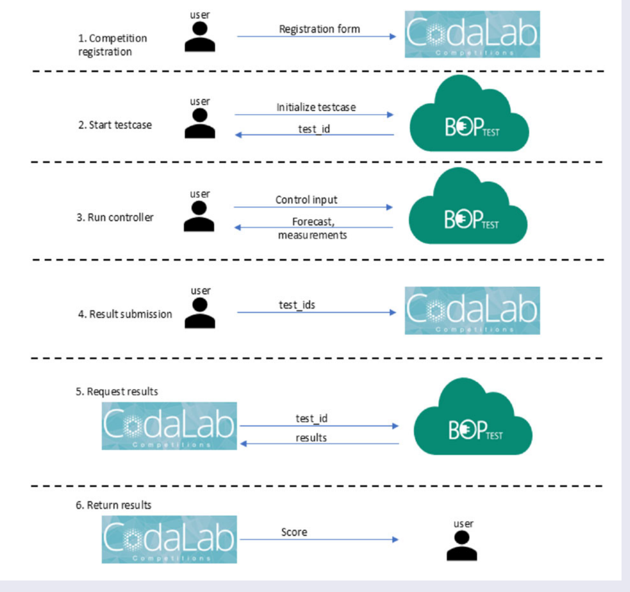
JOURNAL OF BUILDING PERFORMANCE SIMULATION
https://doi.org/10.1080/19401493.2026.2656363

Smart building HVAC control challenge: experience and solutions from the ADRENALIN project

Harald Taxt Walnum [a,b], Balázs András Tolnai[c], David Blum[d], Jicheng Shi[e], Colin N. Jones[e], Deep Dessai[f], Wenbin Wang[e], Wenjie Xu[e], Sebastien Gros [b] and Igor Sartori [a]

[a]Department of Architectural Engineering, SINTEF Community, Oslo, Norway; [b]Department of Engineering Cybernetics, NTNU, Trondheim, Norway; [c]SDU Center for Energy Informatics, Maersk McKinney Moeller Institute, The Faculty of Engineering, University of Southern Denmark, Odense, Denmark; [d]Building and Industrial Energy Systems Division, LBNL, Berkeley, USA; [e]Automatic Control Lab, EPFL, Lausanne, Switzerland; [f]Independent Researcher, Goa, India

ABSTRACT
A smart building HVAC control competition crowdsourced and compared algorithms on fair and equal ground using the standardized BOPTEST framework. The competition attracted 138 participants, but only 9% submitted valid solutions for the final stage, highlighting the complexity of advanced HVAC control design. The winning solutions showed significant potential to reduce energy use and cost by shifting demand, without compromising occupant comfort. Across scenarios, thermal energy cost reductions of 36–76% relative to a baseline, were achieved. In peak heat periods, the cost reduction leveraged limited energy use reduction (0–15%), but more significant energy price reduction (34–62%). This shows smart controls' ability to avoid as much as possible consumption during the morning peak hours, when spot prices are tendentially the highest. Hosting the competition has highlighted challenges in creating competitions that both are fair and promotes solutions that are transferable to real life implementation.



ARTICLE HISTORY
Received 4 November 2025
Accepted 26 March 2026

KEYWORDS
Buildings; HVAC; control; competition; BOPTEST

1. Introduction

Operation of buildings represents about 30% of energy consumption globally (IEA, GlobalABC, and UN Environmental Programme 2020) and 40% in the EU. With 75% of the EU building stock being still energy-inefficient, the bulk of it is due heating and cooling (Directive (EU) 2024). Cooling is already responsible for the summer peak load in the power system and, with the heating sector seeing an increase in heat pump adoption, heating is on its way to play the same role in winter, as it is already the case in

CONTACT Harald Taxt Walnum harald.walnum@sintef.no Department of Architectural Engineering, SINTEF Community, Pb 124, Blindern, 0314Oslo, Norway

Norway, for example. In the residential sector, the share of electricity in heating demand should grow to 40% by 2030 and to 50–70% by 2050; in the services sector, these shares are expected to be around 65% by 2030 and 80% by 2050 (COM 2020).

However, buildings also have a large potential for flexibility thanks to their slow thermal inertia, and thus offer the potential to be a cost-effective solution for providing the flexible demand needed to support load growth and maintain grid stability (Satchwell et al. 2021). As the majority of the building stock that will be operating in 2050 is made of buildings that are already built today, enabling reduced and flexible demand from existing buildings is vital for operating cost-effective and reliable electric grids of the future. The paths for reaching energy efficiency targets typically assume a significant increase in the energy renovation rate; however these rates are challenging to reach (Sandberg et al. 2016), making the targets unrealistic without new approaches.

Predictive control approaches for heating, ventilation and air conditioning (HVAC) systems have been widely studied in the research community and have shown potential for cutting energy costs in buildings and enabling flexible operation. Several reviews demonstrate the abundance of theoretical studies on different approaches to smart control of HVAC systems (Drgoňa et al. 2020; Kathirgamanathan et al. 2021; Xin et al. 2024; Yao and Shekhar 2021). However, a widespread adoption in practice is still missing (Drgoňa et al. 2020). Blum et al. (2022) gives a literature review of field tests on commercial buildings documented in scientific literature between 2011 and 2021 and found 14 studies. The field tests demonstrate energy or cost saving in the range of 10–70%.

However, while the reported relative savings give indications on the MPC performance, the savings are calculated based on different baseline comparisons, building states, HVAC systems and weather conditions, making a comparison across buildings challenging. Furthermore, Kathirgamanathan et al. (2021) points out that there is a strong need for benchmarking datasets besides establishing processes and toolkits that facilitate benchmarking. Not having benchmarking tools makes it challenging to compare the performance and scalability of different data-driven techniques as the individual studies are presented for different buildings each having different boundary conditions. The same need is emphasized by Blum et al. (2021). For benchmarking of control algorithms, static databases are not sufficient, as the state of the building changes when the controls are applied.

Such framework is offered by the BOPTEST (Blum et al. 2021). To facilitate benchmarking and comparison of different control algorithms, BOPTEST constructs a standardized computational environment for building emulators, which exposes overwritable control points and available sensor points of the emulators using a familiar, language-independent HTTP Rest API (Application Programming Interface). For scalability purposes, it was important to focus on solutions that are transferable between buildings with as little tailoring as possible, which is often mentioned as a key barrier for a large-scale implementation of MPC (Bird et al. 2022). Comparative analysis of the algorithms resulting from the competition has helped in gaining a better understanding of the performance and scalability of different control algorithms in the building domain.

The ADRENALIN project was an international project with a balanced consortium of a dozen partners between academic, research organisation and industry partners, mostly from Northern Europe (see ADRENALIN project 2026 for further details), aimed at facilitating large scale roll-out of data services in the existing building stock by crowdsourcing to data challenge competitions the development of new algorithms. In particular, this paper is related to the 'BOPTEST Challenge': a smart building HVAC control challenge aimed to develop efficient and scalable control algorithms that can be applied in commercial buildings to reduce energy consumption and release their flexibility potential. The algorithms performance was evaluated through the BOPTEST platform. The competition winners have been engaged into a three month period of knowledge transfer to the sponsoring companies, with the aim to implement their solutions in real-life conditions on the digital platforms of the partner companies, and so test their general validity and replicability, and to demonstrate real-life performance.

Donoho (2024) discusses the importance of competitive challenges to drive the field of data sciences forward. Two examples of building HVAC control and energy flexibility challenges, are the CityLearn challenges, first launched in 2020 (Vázquez-Canteli et al. 2020), and the world championship in cybernetic building optimization (WCCBO), first launched in 2019 (Togashi, Masato, and Yamamoto 2020). The CityLearn challenges are based on the CityLearn platform (Nweye et al. 2025), designed for benchmarking of reinforcement learning (RL) algorithms for building energy coordination and demand response in cities. While the initial challenge only contained static pre-calculated load profiles for buildings, this has later been expanded with long short-term memory (LSTM) based models to allow for flexible operation of the building heating system. However, the models, designed to represent residential buildings with heat pumps, do not include the complicated HVAC systems typically present in modern non-residential buildings. The WCCBO focuses on reducing energy consumption and optimizing the indoor climate. The first competition used an emulator of a large office building that included models of the control and HVAC system (Togashi and Miyata 2019). The emulator could be run on a service and communicating through VPN and BACnet for real-time feedback control or locally with predefined control input sequences, however, only one team successfully connected to the server and was able to run with feedback control (Togashi, Masato, and Yamamoto 2020). The second competition focussed on control of a variable refrigerant volume (VRV) system (Togashi et al. 2024) and was hosted in the autumn of 2024. However, the lack of forecasting services makes it unsuited for evaluating flexibility and excludes predictive algorithms such as model predictive control (MPC)

This paper has a double objective: on the one hand to describe the process of outsourcing the development of smart building control algorithms via a competition and the lessons learned from it, as described in Section 2 and discussed in Section 5.2; on the other hand to present the winning algorithms and their results, as described in Section 4 and discussed in Section 5.1

2. Competition organization and rules

This section describes the conduction of the competition from both an organizational and a technical point of view.

2.1. Back-end simulation platform – BOPTEST

The Building Optimization Testing Framework (BOPTEST) provides open-source software to test, evaluate, and benchmark the performance of building HVAC control strategies (Blum et al. 2021), including advanced algorithms like MPC. The framework is comprised of a repository of publicly available building emulators (so-called 'test cases') and a run-time environment (RTE) that exposes an API for users to select a test case, set up testing scenarios, interact with the test case by reading and writing I/O points, obtain forecasts of exogenous variables, such as weather and energy prices, and request calculation of a standard set of key performance indicators (KPIs). Users of BOPTEST for controller testing can download the framework from GitHub and deploy it on their own computing resources, or access a publicly deployed instance on AWS which is managed by the BOPTEST project team (which we will call 'BOPTEST-Service').

While the competition created two new test cases not already publicly available as part of BOPTEST, the competition made extensive use of the RTE and BOPTEST-Service. Competition participants used the BOPTEST RTE and API, either locally or with BOPTEST-Service, to run controller tests for the competition, the practical process of which is described in more detail in the next section. Updates to BOPTEST-Service were made to allow competition administrators to make the two competition test cases available to users under a particular *namespace* created for the competition, separate from the existing publicly available test cases. In addition, BOPTEST-Service manages multiple client requests for running tests in parallel by assigning and distributing a unique test ID to each client upon their request for deployment of and access to a test case. Also, an administrator-only API for BOPTEST-Service allows for access to simulation and KPI data generated during tests stored in persistent storage for a provided test ID. This combination of features, therefore, allowed competition administrators to retrieve and verify test results generated by participants after participants supply the test IDs used for their specific tests, as described in the next section.

2.2. Competition hosting

For the Adrenalin BOPTEST Challenge, CodaLab served as the submission and competition management platform, handling the retrieval of results and the calculation of the final scores. CodaLab is an open source platform designed to facilitate the organization and execution of scientific and data-driven competitions in various research domains, including machine learning, control systems, and energy management (Pavao et al. 2023).

The core workflow for participation and submission is shown in Figure 1. After executing their algorithms on BOPTEST and receiving their test IDs, participants submitted them to CodaLab. To make this submission, participants had to organize their test IDs into an Excel file, which included the corresponding case name, scenario type, and results. CodaLab then fetched the KPIs for the test IDs from the BOPTEST service database, computed the score, according to Equations (1) and (2). These scores were then posted to CodaLab's leaderboard, allowing participants to see how their solutions ranked in real time. Participants could submit partial results by leaving some test IDs blank, allowing them to receive feedback even if they had not yet completed all scenarios. However, to be eligible for a winning position, a valid result was required for all four competition scenarios.

By automating only the retrieval of results and the comparison of scores with the baseline, CodaLab ensured a streamlined competition experience while maintaining fairness and transparency. This setup allowed participants to freely develop and test their control strategies in BOPTEST while benefiting from CodaLab's structured result management and ranking system. The seamless integration of these platforms provided an efficient, scalable, and reproducible framework for benchmarking smart building control solutions.

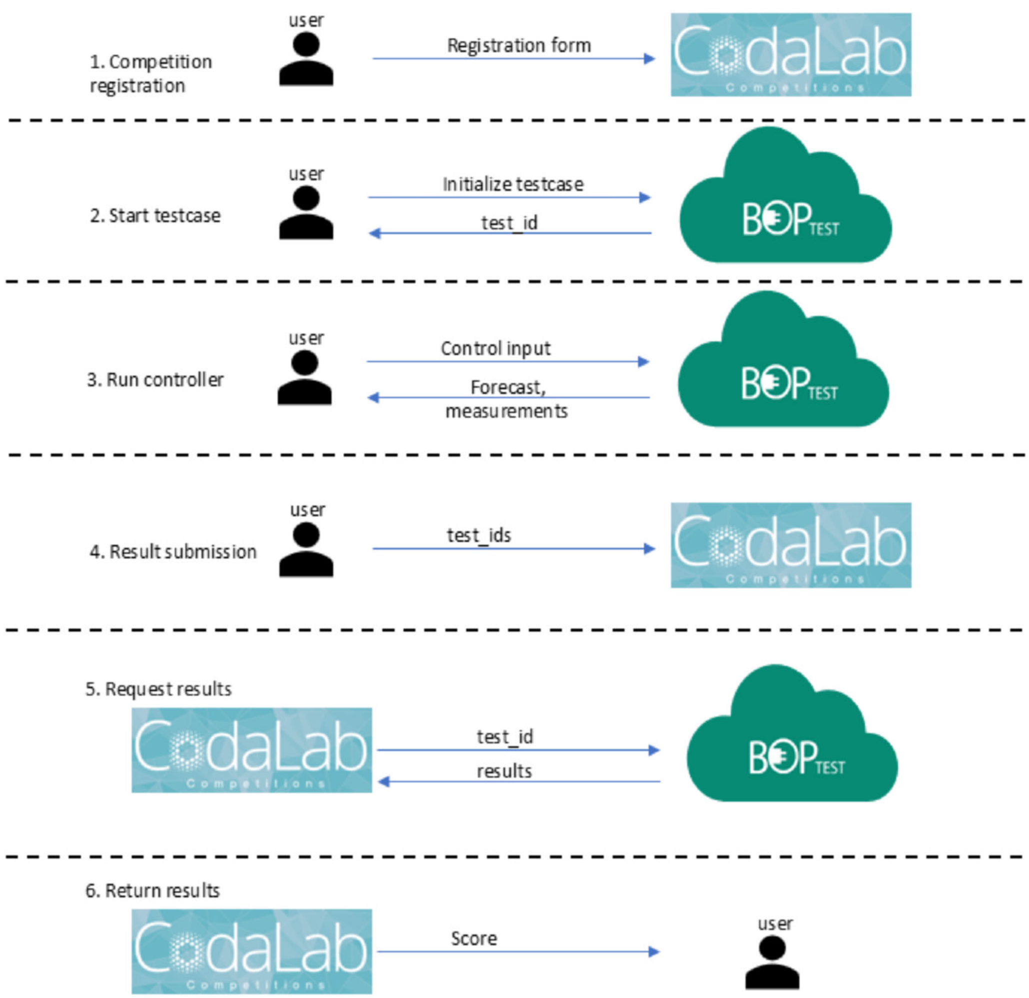


Figure 1. Workflow for participation, submission and retrieving results.

2.3. Organization

Two emulators were developed for the competition. The emulators are described in detail in Section 3. For each of the two emulators, two scenarios were defined: *peak_heat_day* and *typical_heat_day*. The *peak_heat_day* scenarios were defined as a 14-day period centred around the day with the highest heating demand when running with the baseline controller. The *typical_heat_day* scenarios were manually selected 14-day periods, representing days with typical heat demand.

The main competition was split into two stages: a training stage and a competition stage. The training stage lasted two months. In this phase of the competition, participants were allowed to download and run a compiled version of the emulators locally or through the BOPTEST service platform. The staring kit for the competition, including the emulators is publicly available (Adrenalin 2026). Running it locally also allowed for using the BOPTEST-gym environment (Arroyo et al. 2021) for training of RL algorithms. There were no limitations in the generation or usage of data from the emulators to train the models. However, only runs through the service platform could be submitted to the CodaLab platform. A valid submission for all four scenarios during the training stage was required for the participants to be eligible to participate in the competition stage.

The competition stage lasted one week. For this stage, a new version of each of the emulators were made available. The boundary conditions (weather and price data) were replaced with another dataset. In this phase, the local versions of the emulators were not available, so all runs had to be through the service platform.

2.4. Rules and evaluation

To allow a fair and open evaluation of the submissions, the main ranking of the submitted solutions was fully quantitative, based on a predefined cost function. In addition, participants were asked to provide a qualitative description of their controller.

Table 1. Description of KPIs, adopted from Blum et al. (2021).

Name	Unit	Description
$cost_{tot}$	EUR/m^2	Cost: operational cost associated with the HVAC energy usage. Energy $*$ price
$idis_{tot}$	ppm$*$h/zone	Indoor air quality (IAQ): the extent that the CO_2 concentration levels in zones exceed bounds of the acceptable concentration level, which are predefined within the test case FMU for each zone, averaged over all zones.
$pdih_{tot}$	kW/m^2	Peak district heating demand: the HVAC peak district heating demand (15 min).
$pele_{tot}$	kW/m^2	Peak electrical demand: defines the HVAC peak electrical demand (15 min).
$tdis_{tot}$	Kh/zone	Thermal discomfort: the cumulative deviation of zone temperatures from upper and lower comfort limits that are predefined within the test case FMU for each zone, averaged over all zones. Air temperature is used for air-based systems and operative temperature is used for radiant systems.

The quantitative evaluation used a selection of the standard KPIs from the BOPTEST framework and weighting them in a total score. The description of the relevant KPIs is shown in Table 1 below. Details on the calculation of each KPI can be found in Blum et al. (2021).

The score of a controller on a single scenario i is given by Equation (1). The score equation contains the energy cost plus a peak power penalty, representing a peak power tariff. The energy prices for electricity and district heating (DH) are the same and are based on electricity spot prices for the location of the building, to make them coherent with the local weather. However, as an important factor in the competition, was for the controller to demonstrate its ability to exploit the flexibility of the building, the spot price signal was modified to make it more volatile. For the competition stage, this transformation was performed by extracting the 4 h with the highest prices, and 8 h with the lowest prices for each day. The prices for high price and low price hours were then set to 0.2 EUR/kWh and 0 EUR/kWh respectively. The price for the remaining hours was set to the mean of the high and low price hours, equal to 0.067 kWh. Finally, a fixed energy cost of 0.025 EUR/kWh was added to all hours.

In addition, the score equation contains a penalty for violating the comfort constraints. The constraint violation penalties were manually tuned to ensure that the cost of violating the constraint is higher than the cost savings from reduced energy consumption, but avoid to high costs, that would favour robust control approaches. The tuning process for the thermal comfort constraint penalty was performed by running the scenarios with the temperature setpoint of the baseline controller slightly lower than the default. Then the thermal comfort penalty could be set so that the scenario score was slightly higher than with the default controller for all scenarios. A similar approach was used for the IAQ penalty, but by modifying the CO_2 setpoint of the demand controlled ventilation (DCV) controller.

The total score of a submission was then calculated as the sum of the difference between the scores of the baseline controller and submission controller for all scenarios, as shown in Equation (2). The baseline controllers are price-ignorant.

$$\begin{aligned}\mathrm{Score}_i = {} & \mathrm{cost}_{\mathrm{tot},i} + 4.25 \times (\mathrm{pdih}_{\mathrm{tot},i} + \mathrm{pele}_{\mathrm{tot},i}) \\ & + 0.005 \times \mathrm{tdis}_{\mathrm{tot},i} + 0.0001 \times \mathrm{idis}_{\mathrm{tot},i}\end{aligned} \tag{1}$$

$$\mathrm{Score}_{\mathrm{Total}}^{\mathrm{controller}} = \sum_{i=1}^{4} \left(\mathrm{Score}_i^{\mathrm{baseline}} - \mathrm{Score}_i^{\mathrm{controller}}\right) \tag{2}$$

To be eligible for the prize money, the submission needed positive scores on all scenarios. The three participants with the highest total score were awarded a total 10,000 Euros each.

3. Emulators

This section describes the building emulators used for the Adrenalin competition. First a general overview and common properties between the two emulators are discussed, followed by detailed discussions of each emulator in separate subsections.

The two emulators were chosen and designed to be representative of modern non-residential buildings in Northern Europe. Both buildings have DH as heat source for a radiator based hydronic heating system and DCV. The idea, though not forced on the participants, was that the same controller approach could be used for both emulators.

As an important part of the competition was for the controllers to demonstrate exploitation of the thermal flexibility in the building envelope and the HVAC system, some extra effort was put in modelling the dynamic response realistically. While the thermal mass of the building and the radiators are inherently modelled in the Modelica components used in the BOPTEST emulators, additional piping segments were included between the heating central and the decentralized HVAC equipment. The piping segments also included heat loss, meaning that the temperature decays towards the surroundings when no heat is added from the heating central. This creates a dynamic effect with both a delay, and a time constant very different to the envelope, and can result in high peaks when the heating system is turned on. This effect

Table 2. Weather forecast parameters.

Name	Unit	Description
TDryBul	K	Ambient dry bulb temperature at ground level
HGloHor	W/m^2	Horizontal global radiation
winDir	rad	Wind direction
winSpe	m/s	Wind speed
relHum	%	Relative humidity

is typically seen in buildings with DH systems (Walnum et al. 2025).

3.1. Forecasts

The BOPTEST framework, by default, provides forecasts for all weather and occupancy parameters to be used by external control algorithms. For the Adrenalin competition, the available parameters were reduced to a selection that was found more realistic for a real case. The available weather forecast parameters are shown in Table 2, no forecast was given on occupancy, in addition, the indoor temperature bounds and CO_2-level limits were included in the forecast.

3.2. Adrenalin1 – single zone university building

The Adrenalin 1 emulator was based on a previous version of the standard BOPTEST emulator 'singlezone_commercial_hydronic', and has now replaced this emulator in BOPTEST. This emulator is based on a highly energy efficient university building at the University of Southern Denmark (SDU). Details about the real building can be found in Jradi et al. (2017), while details about the initial emulator model are described in Yang et al. (2020). The building envelope model is unchanged, but the HVAC system has been modified to provide more realistic behaviour and fit with the requirements of the Adrenalin competition. The resulting configuration and control are discussed below.

3.2.1. HVAC-system

The layout of the ventilation system is similar to the original BOPTEST emulator, but the performance characteristics and control of some of the components are different.

Figure 2 show a principle flow sheet for the air handling unit (AHU), as modelled in the emulator. It shows the placement of the sensors and components and the control principle for the baseline controller.

- The supply fan's relative speed (0-1) is controlled by the CO_2 level in the building. Minimum relative fan speed is set to 30% in the baseline control.
- The extract fan's relative speed (0-1) is controlled by the supply air flow rate, to keep equal supply and extract flow. Minimum relative fan speed is set to 20% in the baseline control.
- The speed of the rotary wheel in the heat recovery unit is controlled to reach the supply temperature set point temperature.
- The valve on the hydronic heating battery is controlled to reach the supply temperature set point.

The layout of the hydronic heating system is shown in Figure 3. Piping segments are modelled between the main distribution pipes and the control valves for both the ventilation and the radiator circuits.

Heat is supplied by a DH system. The temperature of the DH is fixed to 65 °C. The supply temperature of the hydronic heating system is controlled by the flow on the primary side of the DH heat exchanger. The set point follows an outdoor temperature compensation curve.

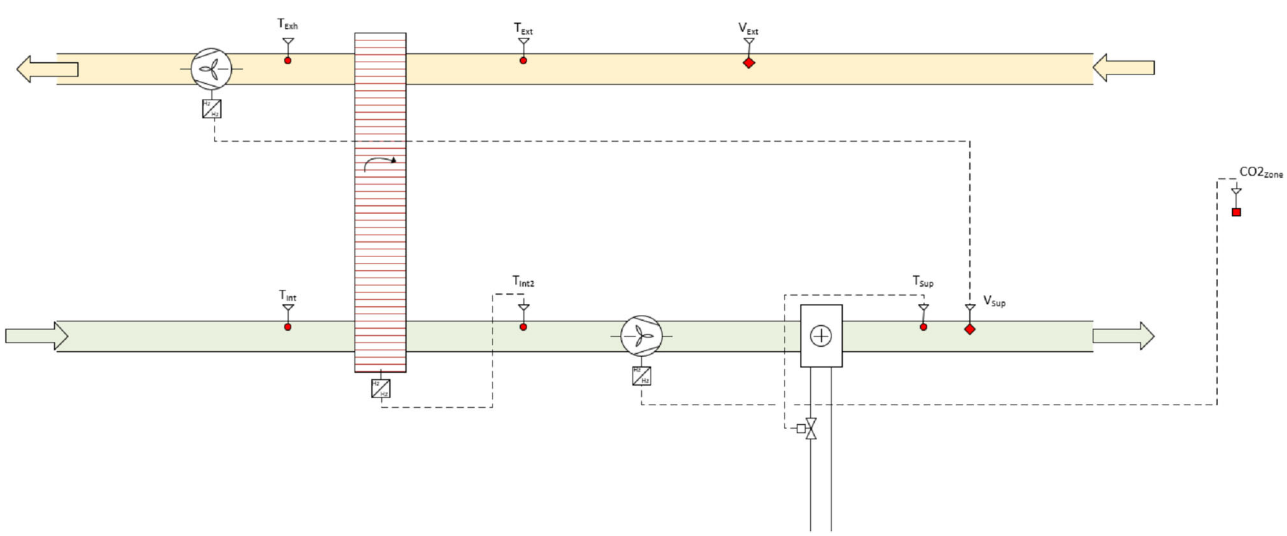


Figure 2. Principle flow sheet for AHU in emulator 1.

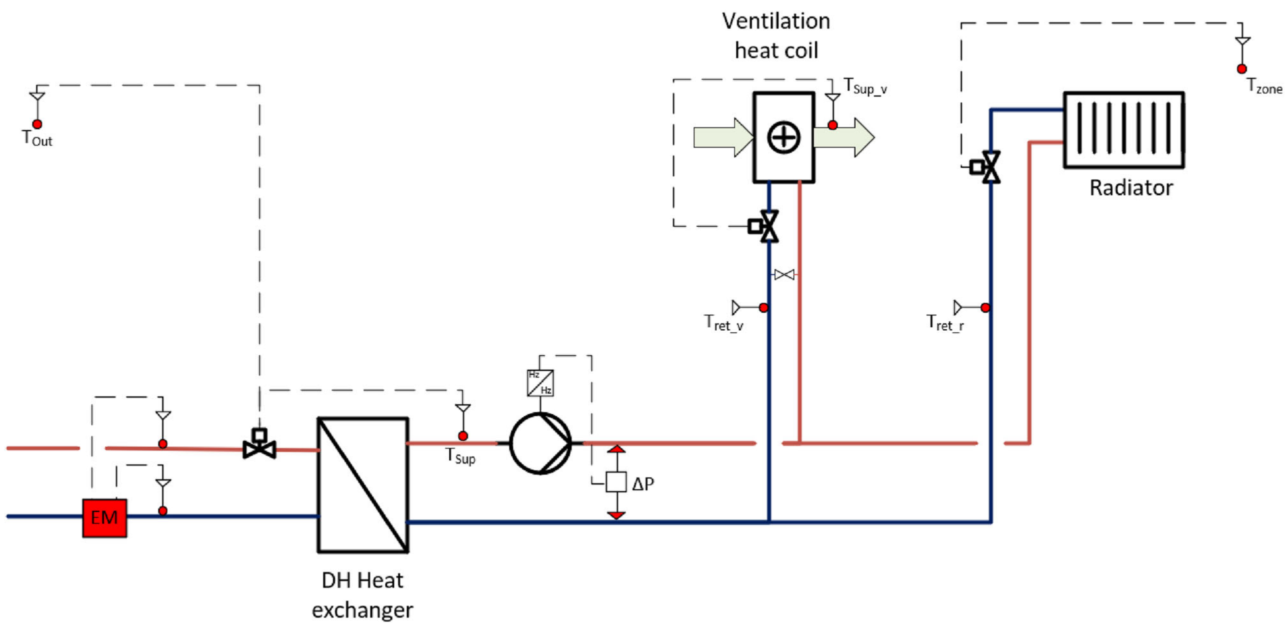


Figure 3. Principle hydronic flow sheet for emulator 1.

The main circulation pump sustains a fixed pressure across the distribution pipes. The baseline set point is 50,000 Pa. This is also the maximum possible pressure difference of the pump.

The control valve of the radiator tracks the measured indoor temperature. The baseline controller implements a night setback of 3 °C outside operating hours. After night set-back, the set point is reset to 22 °C 2 h before the next operating hour (4 h on Mondays) to ensure the building has reached the set point temperature before the building is occupied again.

3.2.2. Occupancy and internal gains

The occupancy profile is based on the collected data described in Yang et al. (2020), with minor modifications to adjust for missing data. Internal heat gains of 120 W/person are added to the thermal zone, split into radiative (40%), convective (40%), and latent gains (20%). The internal gains from electrical equipment and lighting are estimated to be part of this but are not measured separately.

3.2.3. Inputs and outputs

Table 3 shows the available control inputs for emulator 1. The zone temperature set point, supply temperature set point and zone CO_2 set point, are set point inputs to internal PID controllers, while the rest are direct control signals. However, all internal actuators are modelled with a second order delay filter, to avoid unrealistic behaviour. Table 4 shows the available measurement points. In addition, measurements from an internal weather station were available.

Table 3. Emulator 1 available control inputs.

Description	Unit	Min	Max
AHU return fan speed control signal	–	0	1
AHU supply fan speed control signal	–	0	1
Zone CO_2 concentration set point	ppm	400	1000
Pump differential pressure	Pa	0	50,000
AHU supply air temperature set point for heating	K	288.15	313.15
Zone temperature set point	K	283.15	303.15
AHU heating coil valve control signal	–	0	1
Radiator valve control signal	–	0	1
Supply temperature set point for heating	K	283.15	333.15

Table 4. Measured outputs from emulator 1.

Description	Unit
AHU supply air volume flow rate	m^3/s
AHU extract air volume flow rate	m^3/s
AHU heating coil supply water temperature	K
AHU air temperature exiting heat recovery unit	K
AHU return air temperature	K
AHU supply air temperature	K
Zone CO_2 concentration	ppm
AHU electrical power consumption	W
Pump electrical power consumption	W
DH thermal power consumption	W
AHU heating coil return water temperature	K
Zone air temperature	K

3.3. Adrenalin2 – multi zone office floor

The Adrenalin 2 emulator was based on the BOPTEST 'Multizone_office_simple_air' emulator, but the HVAC system has been rebuilt to a hydronic heating system. Details about the original emulator can be found at the BOPTEST webpage. (https://ibpsa.github.io/project1-boptest/docs-testcases/multizone_office_simple_air/index.html.) The envelope model is a model of a single

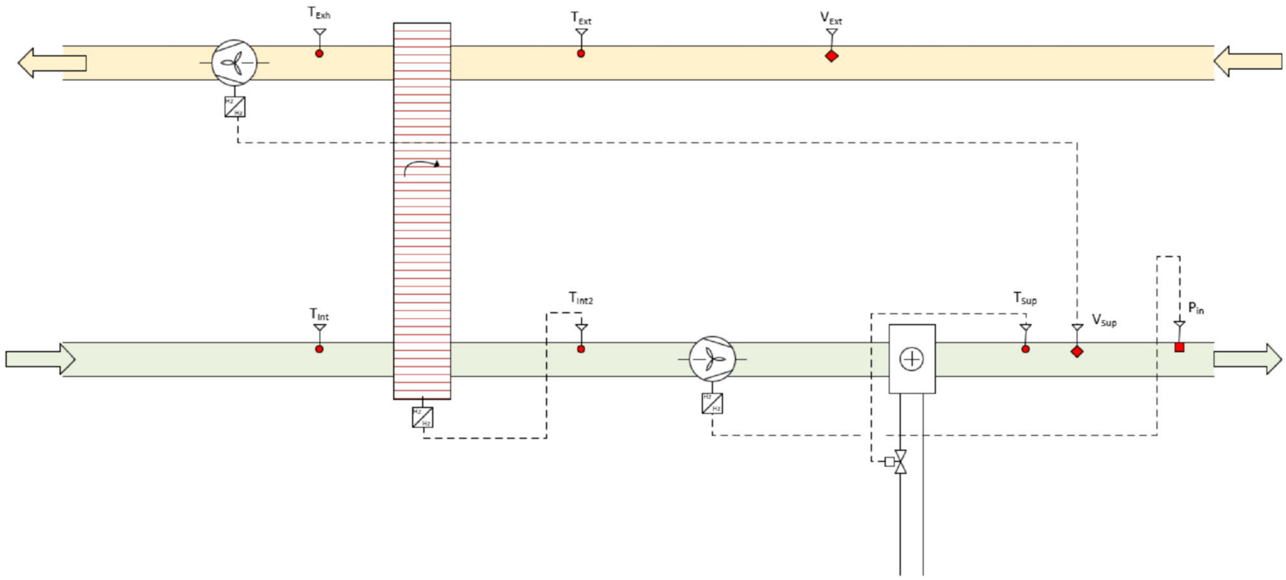


Figure 4. Principle flow sheet for AHU in emulator 2.

floor of an office building, divided into five zones: one per façade and a central zone, based on the DOE medium office building for Chicago IL (Deru et al. 2011). However, the façade construction was modified to the following: 10 mm wood, 250 mm insulation and 150 mm concrete.

An important modification is also the addition of solar shading on the façade. The shading is automatically activated if the solar irradiation on the façade is above 150 W/m^2, but the position of the shading (0-1) can be directly overwritten by the external control algorithm.

3.3.1. HVAC-system

The floor is ventilated with demand controlled ventilation (DCV). The air handling unit (AHU) has a rotary wheel heat recovery unit, hydronic heating coil and speed controlled fans.

Figure 4 show a principle flow sheet for the AHU, as modelled in the emulator. It shows the placement of the sensors and component and the control principle for the baseline controller.

- The supply fan's relative speed (0-1) is controlled by a pressure difference input (supply pressure (Pin) – atmospheric pressure). Default (baseline) is 200 Pa. Minimum relative fan speed is set to 20%.
- The extract fan's relative speed (0-1) is controlled by the supply air flow rate, to keep equal supply and extract flow.
- The speed of the rotary wheel in the heat recovery unit is controlled to reach the supply temperature set point.
- The hydronic heating battery control valve is controlled to reach the supply temperature setpoint.

The supply air to the zones is controlled by individual dampers. The dampers are controlled to keep the CO_2 level below the threshold. Default is 950 ppm.

The layout of the hydronic heating system is shown in Figure 5. Piping segments are modelled between the main distribution pipes and the control valves for both the ventilation and the radiator circuits.

Heat is supplied by a DH system. The temperature of the DH is fixed to 65 °C. The supply temperature of the hydronic heating system is controlled by the flow on the primary side of the DH heat exchanger. The set point follows an outdoor temperature compensation curve.

The main circulation pump sustains a fixed pressure lift. The baseline set point is 35 kPa. This is also the maximum possible pressure difference of the pump.

The control valve of the radiator regulates after the measured indoor temperature. The baseline controller implements a fixed set point of 22 °C.

3.3.2. Occupancy and internal gains

The occupancy and internal gains profiles are set with a fixed schedule for workdays and weekends based on normalized profiles from the Norwegian standard (Standard Norge 2020), but with an added Gaussian noise term according to the following equation:

$$v' = v + \epsilon \cdot v \quad \text{with} \quad \epsilon \sim \mathcal{N}(\mu, \sigma^2) \quad \text{and}$$
$$\epsilon = \min(\max(\epsilon, -1), 1) \tag{3}$$

Where v' is the resulting internal gains, v is the original normalized internal gains. ϵ is sampled from the Gaussian distribution $\mathcal{N}$ with zero mean (μ) and variance (σ^2) of 0.2. The sampled noise (ϵ) is truncated between -1 and

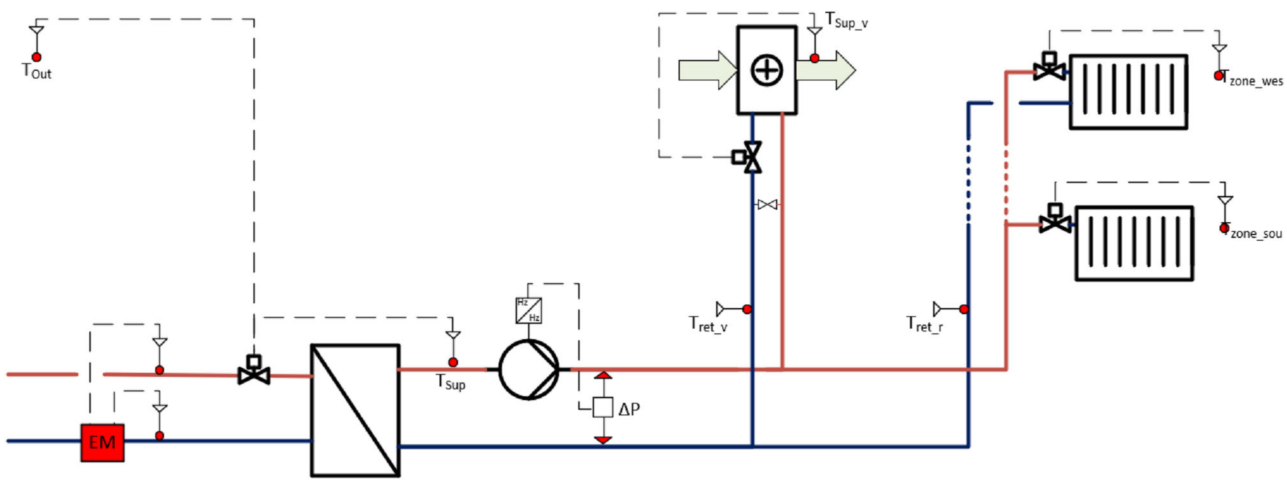


Figure 5. Principle hydronic flow sheet for emulator 2.

Table 5. Emulator 2 available control inputs.

Description	Unit	Min	Max
AHU return fan speed control signal	–	0	1
AHU supply fan speed control signal	–	0	1
Zone CO_2 concentration set point [5]	ppm	400	1000
Zone temperature set point [4]	K	283.15	303.15
AHU supply air temperature set point	K	288.15	313.15
Pump differential pressure	Pa	0	15000
Supply temperature set point for heating	K	283.15	333.15
Shading position for the different façades [4]	–	0	1

1. For each zone, the internal gains from lighting and equipment and occupancy have the same relative noise. An important difference compared to emulator 1 is that the electricity consumption of equipment and lighting is measured and added to the total electricity consumption in the score function.

3.3.3. Inputs and outputs

Table 5 shows the available control inputs for emulator 2. The zone temperature set points, supply temperature set points and zone CO_2 set points, are set point inputs to internal PID controllers, while the rest are direct control signals. However, all internal actuators are modelled with a second-order delay filter, to avoid unrealistic behaviour. Table 6 shows the available measurement points. In addition, measurements from an internal weather station were available.

4. Results

This section describes the results of the competition. First, an overview of the submissions is given, followed by a description of the top three controllers, and lastly, the performance of the top three controllers in the final competition stage is described.

Table 6. Measured outputs from emulator 2.

Description	Unit
AHU supply air volume flow rate	m^3/s
AHU extract air volume flow rate	m^3/s
AHU heating coil supply water temperature	K
AHU air temperature exiting heat recovery unit	K
AHU return air temperature	K
AHU supply air temperature	K
Zone CO_2 concentration [5]	ppm
AHU electrical power consumption	W
Pump electrical power consumption	W
Auxilary electrical power consumption	W
Radiator heating power consumption	W
Radiator heating energy consumption (accumulated)	kWh
Ventilation heating power consumption	W
Ventilation heating energy consumption (accumulated)	kWh
DH thermal power consumption	W
DH supply temperature	K
DH return temperature	K
AHU heating coil return water temperature	K
Zone air temperature [5]	K

4.1. Submission overview

In total, 138 participants/teams signed up for the competition. During the training stage, running from July 8th to September 15th, 94 submissions from a total of 22 active teams were received. Only 12 of these achieved a total score higher than 0, and were found eligible for participation in the competition stage.

In the competition stage, running from September 23th to September 30th, 5 of the 12 participants made a submission, but only three of them submitted valid attempts for all scenarios.

Of the 5 final submissions, two (2^{nd} and 5th) where heuristic based, while the three others where data-driven. The 1st and 3rd place use data-driven predictors to optimize the energy use by the heating system (see below for details on the models), but use heuristic rules to control the actuators so that the correct amount of energy is used. The 4th place controller, on the other hand, uses a direct mapping of the current state and future forecast to the room temperature setpoint, using a sigmoid function on a linear parametric model.

As documentation of the controllers was only required in the final stage, the authors have limited knowledge about the control methods applied by those who did not submit in the final stage. However, based on questions asked in the forum, it is known that at least 4 participants tried to apply RL based algorithms, but none of them made a final submission. However, in the training phase, they scored significantly lower than the top three submissions.

4.2. Top three controllers

4.2.1. 1st place

Controller 1 employs a bi-level data-enabled predictive control (DeePC) method (Shi et al. 2025) with efficient adaptive updates (Shi, Lian, and Jones 2024), built and updated directly from past data. Let y, u, w denote system outputs, HVAC inputs and external inputs, respectively. For any signal s at time t, let s_t and $s_{t_1|t}$ respectively represent the measurement and the prediction for time t_1. Define s_t^{Δ} and $s_{t|t}^{\Delta}$ as the measurement and prediction sequence from time t over the next Δ steps.

The controller is summarized in Algorithm 1. The prediction function $f(\cdot)$ in (5) is a non-parametric model built directly from the dataset $\mathcal{D}_t$. It is linear in $u_{t|t}^{t_p}$, making the DeePC (5) a linear programming problem. The adaptive updates in Equation (4) capture the recent building dynamics, which are generally nonlinear and time-varying (Afroz et al. 2018). Although the dataset grows over time, Shi, Lian, and Jones (2024) provides a method to maintain fixed computational complexity. Compared to linear state-space methods such as system identification, the bi-level DeePC avoids explicit state estimation and the need for an observer, reducing commissioning effort. Its effectiveness has been validated in real-world building control experiments. Further details are available in Shi et al. (2025) and Shi, Lian, and Jones (2024).

The above controller was applied to both emulators with similar configurations. Here, u was the total heating power, y contained the indoor temperatures for each zone, and w included outdoor temperature (TDryBul) and solar radiation (HGloHor). BOPTEST provided the forecast of $w, E, \underline{Y}, \overline{Y}$ for Step 2) in Algorithm 1. For each scenario (peak/typical heat days), $\mathcal{D}_0$ was collected during the corresponding two weeks using a rule-based controller in the training stage. Due to the importance of the peak score in (1), Q_{max} was first selected via coarse grid search during training and then fine-tuned within a narrower range in the competition stage. The other parameters were set as: $t_p = 96, t_i = 12, R_1 = 10, R_2 = 0.1, \eta = 0.999$, $T_s = 900$ s. They were not heavily tuned, as (Shi et al. 2025) indicates they are limited impact on prediction performance.

The execution of $u_{t|t}^*$ in Step 3) of Algorithm 1 differed slightly between the two emulators. In Emulator 1, the AHU coil valve was set to 0, as the heat recovery

Algorithm 1: Online operation of Controller 1

Given: initial data $\mathcal{D}_0$ containing T-step past sequence of y, u, w, prediction steps t_p, initial measurement steps t_i, depth $L := t_p + t_i$, regularization weights R_1, R_2, forgetting factor $\eta \in (0, 1]$, power limit Q_{max}, sampling time T_s

1) Retrieve the most recent L-step y, u, w data, $\{y_{t-L}^L, u_{t-L}^L, w_{t-L}^L\}$. Update the dataset $\mathcal{D}_t$ as:

$$\mathcal{D}_t = \eta \mathcal{D}_{t-1} \cup \{y_{t-L}^L, u_{t-L}^L, w_{t-L}^L\} \tag{4}$$

2) Obtain forecasts for $w_{t|t}^{t_p}$, energy prices $E_{t|t}^{t_p}$, and output bounds $\underline{Y}_{t|t}^{t_p}, \overline{Y}_{t|t}^{t_p}$. Construct the bi-level DeePC as (5) and solve the optimal input $u_{t|t}^*$.

$$\begin{aligned}
\min_{y_{t|t}^{t_p}, u_{t|t}^{t_p}, \delta_i} \quad & \sum_{i=t}^{t+t_p-1} E_{i|t} u_{i|t} + R_1 \delta_i \\
\text{s.t.} \quad & u_{i|t} \in [0, Q_{max}], y_{i|t} \in [\underline{Y}_{i|t} - \delta_i, \overline{Y}_{i|t} + \delta_i], \quad \forall\, i \in t, \ldots t + t_p - 1. \\
& y_{t|t}^{t_p} = f(\mathcal{D}_t, u_{t|t}^{t_p}, w_{t|t}^{t_p}, R_2)
\end{aligned} \tag{5}$$

3) Apply the first control input $u_{t|t}^*$.
4) Wait until the next sampling time and repeat from Step 1).

unit could not fully compensate for heat losses. During each sampling period, the radiator valve was toggled on and off such that the average heating power approximated $u^*_{t|t}$. In Emulator 2, the AHU coil and radiator valves were not directly controllable. Instead, the pump differential pressure was toggled on/off to track $u^*_{t|t}$. The radiator setpoints were set to 1 °C above the corresponding zone temperatures, and the supply water temperature was held constant to maintain stable heating power. The default CO_2 controller was left unchanged, as its electricity usage was negligible compared to heating costs.

Lastly, limitations and potential improvements are noted. First, the toggling used to execute control inputs may increase hardware wear. This could be mitigated by using local controllers (e.g. new PID loops) to track the desired heating power more smoothly. Second, the default shading controller in Emulator 2 reduces solar heat gains. A custom shading controller could improve energy efficiency, but occupant comfort constraints related to light levels must be defined.

4.2.2. 2nd place

The controller developed for the second-place solution employs a heuristic and rule-based approach to optimize energy efficiency in building HVAC systems while maintaining occupant comfort. This method leverages historical and current electricity and DH price data to make informed heating decisions, focussing on preheating the building during low-price periods and maintaining comfortable temperatures during high-price periods. The approach minimizes the objective function, which balances operational cost, peak heating demand, and thermal discomfort, as defined in the competition's scoring criteria.

The core strategy exploits the building's thermal inertia. During low electricity price periods, the controller preheats the building to a target temperature, allowing it to coast through high-price periods with minimal heating. It uses only past and current electricity prices, maintaining a running average or minimum over a window to classify prices as low or high. When the current price falls below a threshold relative to this reference, the controller preheats; during high-price periods, it minimizes heating while keeping temperatures within comfort bounds.

To manage peak heating and district demand – key components of the objective function – the controller uses two strategies. For scenarios where peak demand is less critical, heating cycles on and off every 5 min, smoothing demand over 15 min windows. When peak demand heavily impacts the score, radiator valves are set to a low, constant level (e.g. 0.05 for Emulator 1's typical heat day), ensuring finer control. Fan operations follow the baseline controller, but with fans turned off when CO_2 levels drop below a threshold (e.g. 500 ppm). In Emulator 2, shades stay at their minimum position to maximize heating efficiency.

The control logic is formalized in the algorithm below, generalizing the approach across emulators (Adrenalin1 and Adrenalin2) and scenarios (peak and typical heat days), with parameters tuned for each:

Algorithm 2: Heuristic HVAC Control Algorithm

```
Initialize parameters based on scenario

for each control step do
    Fetch current price and forecasts (price,
      temperature, CO2)
    Update price history, compute
      reference_price:
    if reference_type == 'min' then
        reference_price ← min(price_history)
    end
    else
        reference_price ← avg(price_history)
    end
    threshold ← reference_price × multiplier
    Set heating setpoint:
    if current price ≤ threshold then
        Set temperature setpoint to preheat_temp
    end
    else
        if building is occupied then
            Set temperature setpoint to
              comfort_temp
        end
        else
            Set temperature setpoint to low_temp
              or deactivate heating
        end
    end
    Manage peak demand:
    if finer control then
        set valve to constant value
    end
    else
        if step index mod 3 ≠ 0 then
            Deactivate heating
        end
    end
end
```

Parameters were tuned iteratively using emulator runs. For Emulator 1's peak heat day, the threshold was 1.05 times the minimum price, with a preheat temperature of 297 K. For its typical heat day, thresholds used 1.6 times the average, with heating at 0.05. For Emulator 2, the peak heat day used a 1.05 multiplier and 299 K preheat, while the typical day used 294.5 K.

This lightweight, deterministic approach requires minimal data and computation compared to machine learning methods. It offers predictability and reliability, with potential scalability via online learning for real-time adaptation to varying conditions.

4.2.3. 3rd place

The controller adopted by this team is an explicit Model Predictive Control (MPC) controller, whose parameters are tuned using Bayesian optimization to maximize the overall performance score. For each scenario, an MPC controller is deployed in the simulated environment for a two-week period. The performance score is then evaluated and provided to a Bayesian optimization algorithm, which updates the controller parameters to improve overall performance. An illustration can be found in Figure 6.

The MPC controller regulates indoor temperature within the comfort range while minimizing energy cost. It employs a model-based approach that uses an Autoregressive with Exogenous Inputs (ARX) model, trained using data collected from a bang-bang controller for each scenario. The ARX model identifies the system dynamics through linear difference equations, making it well-suited for linear time-invariant (LTI) systems. Specifically, for a single-input single-output system (SISO) of order three, it solves the following problem

$$\begin{aligned} \min_{\boldsymbol{a},\boldsymbol{b}} \quad & \|\boldsymbol{e}\| \\ \text{s.t.} \quad & y_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + b_1 u_{t-1} \\ & \quad + b_2 u_{t-2} + b_3 u_{t-3} + e_t, \forall t \in \{3, \dots, T\}, \end{aligned}$$

where u is the system input, y is the system output, $\boldsymbol{a} = [a_1, a_2, a_3]^\top$, $\boldsymbol{b} = [b_1, b_2, b_3]^\top$, $\boldsymbol{e} = [e_3, \dots, e_T]^\top$. For the model used in this competition, the averaged heating power over a 15 min interval from the DH system, the weather temperature, and the solar irradiation

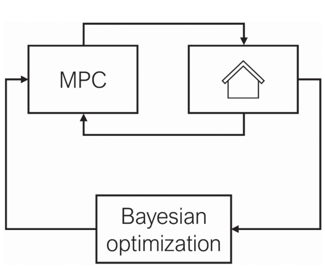


Figure 6. Illustration of the control scheme.

are chosen as inputs, while the indoor temperature is chosen as the output. Once identified, the model parameters remain fixed throughout the operation.

Since the peak heating demand from the DH system, primarily driven by the radiator, contributes to the total performance score, the radiator heating power is treated as the control input in the MPC formulation to regulate peak demand. To do so, an upper bound is imposed on the control input. However, since only the valve (Adrenalin 1) or the set point (Adrenalin 2) can be directly actuated, a modulation technique is employed to ensure that the radiator delivers the desired heating power, which is illustrated in Figure 7.

In Adrenalin 1, the maximum heating power of the radiator is estimated by fully opening the valve for 30 s and measuring the resulting heat output. Based on this, the valve opening duration is calculated such that the average heating power over a 15 min interval matches the value prescribed by the MPC controller. Once this duration is reached within the 15 min window, the valve is closed. In Adrenalin 2, the system is actuated every 30 s, and the radiator heat output is continuously monitored. When the accumulated energy reaches the point where the average heating power aligns with the MPC target, the radiator is turned off.

The MPC controller is parameterized for each scenario, e.g. the upper bound on heating power and the weights in the objective function. During the training phase, the controller is evaluated with different parameters, and the resulting performance scores form the training dataset for Bayesian optimization. By actively and efficiently exploring the parameter space, Bayesian optimization identifies the parameter that maximizes the performance for each scenario. In practical applications where experiments cannot be repeatedly reset, an iterative learning controller tuning approach can be adopted. Controller performance can be assessed at a chosen temporal resolution, such as on a daily or weekly basis. Bayesian optimization can then leverage historical data to continuously refine the controller in real time.

4.3. Comparison of the controller performances

This section presents and discusses the results of the competition. First, the overall scores are discussed, followed by a detailed analysis of the controllers' behaviour.

4.3.1. Scores and KPIs

Figure 8 shows a bar plot of the total scores of the final submissions to the competition stage. The colours of the bars divide the score into the different emulators and scenarios. The higher bars for the Adrenalin1 emulator indicate that the improvements over the baseline were larger

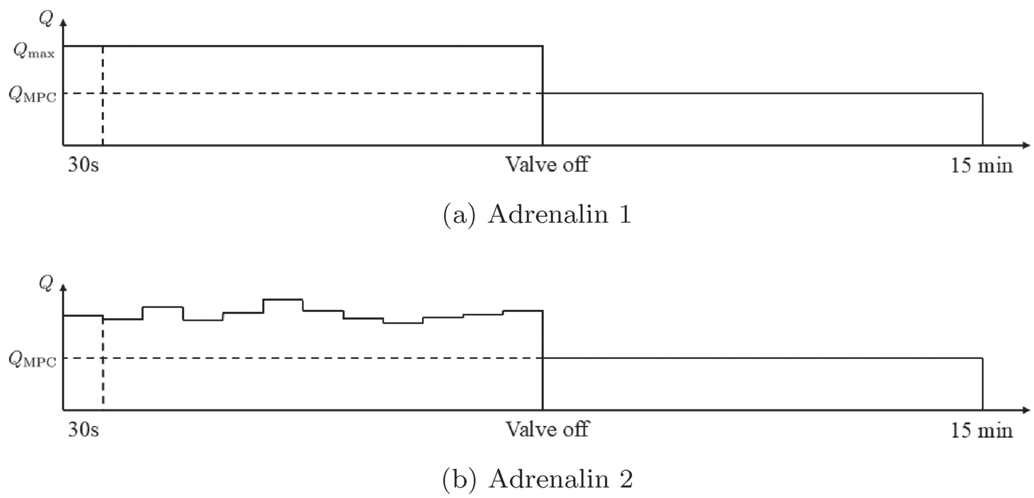


Figure 7. Illustration for modulation. (a) Adrenalin 1. (b) Adrenalin 2.

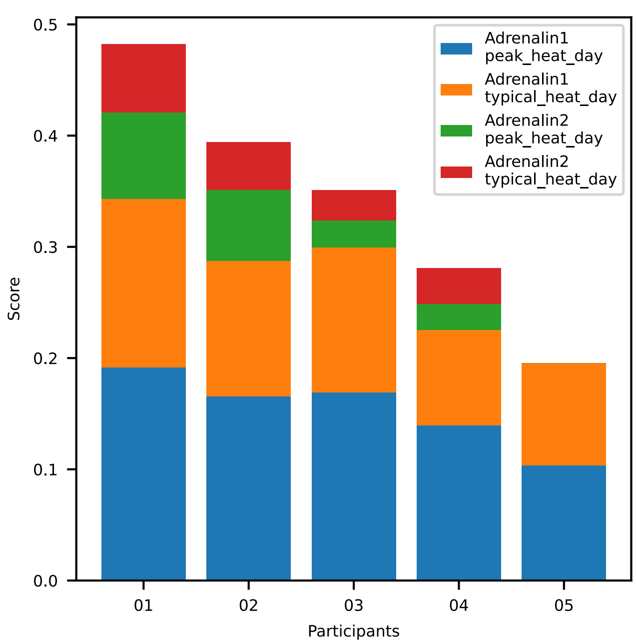


Figure 8. Total score in the competition stage.

than for the Adrenalin2 emulator. This could be due to the larger potential for savings in Adrenalin1 and the greater complexity of controlling the multizone Adrenalin2 emulator. One can also see that the third-ranked solution performed better than the second-ranked on Adrenalin1, but worse on Adrenalin2. The reason for this is at least partially related to operation of the solar blinds in Adranlin2, and is further discussed below.

Figure 9 shows the value of the score function (Equation (1)) for each of the emulators and scenarios, disaggregated into each of the linear components. For each of the scenario, the scores are normalized by the baseline score. This shows that the relative score reduction is much higher for Adrenalin1 than for Adrenalin2. However, it is important to remember that for Adrenalin2, the auxiliary electricity consumption (lighting, plug loads, etc.) is included in the cost function. This reduces the relative saving potential, as these loads cannot be reduced. This is further detailed below.

One can observe that a key part of the cost reduction in the Adrenalin1 typical_heat_day scenario comes from reducing the peak heating demand, while for the rest of the scenarios, the reduction is mainly due to energy cost reductions. This is due to the Adrenalin1 baseline having a high morning peak (see Figure 11) as a consequence of the night setback. In the peak_heat_day scenario, on the other hand, the same high levels of energy use are sustained also during daytime, with the heating system operating close to its maximum capacity most of the time – being it the coldest period – thus leaving little room for shifting consumption.

We can also observe that most of the controllers introduce a slight penalty due to thermal discomfort (*tdis*). However, since thermal discomfort is heavily penalized in the score function, the absolute deviations are small and likely not noticeable by the occupants. In contrary, none of the controllers violates the indoor air quality constraint (*idis*). Actually, none of the controllers try to exploit the energy saving and load shifting potential of the ventilation systems, but let the baseline DCV controller work as normal.

Table 7 shows the absolute energy and peak costs for all test cases and scenarios for the baseline and top three controllers. As commented above, the peak cost reduction is significant only in the Adrenalin1 typical_heat_day scenario, where the high morning peak can be avoided by re-modulating the night setback baseline control. Thus, Figure 10 focuses only on the energy cost, breaking it down further into its two components: energy use and

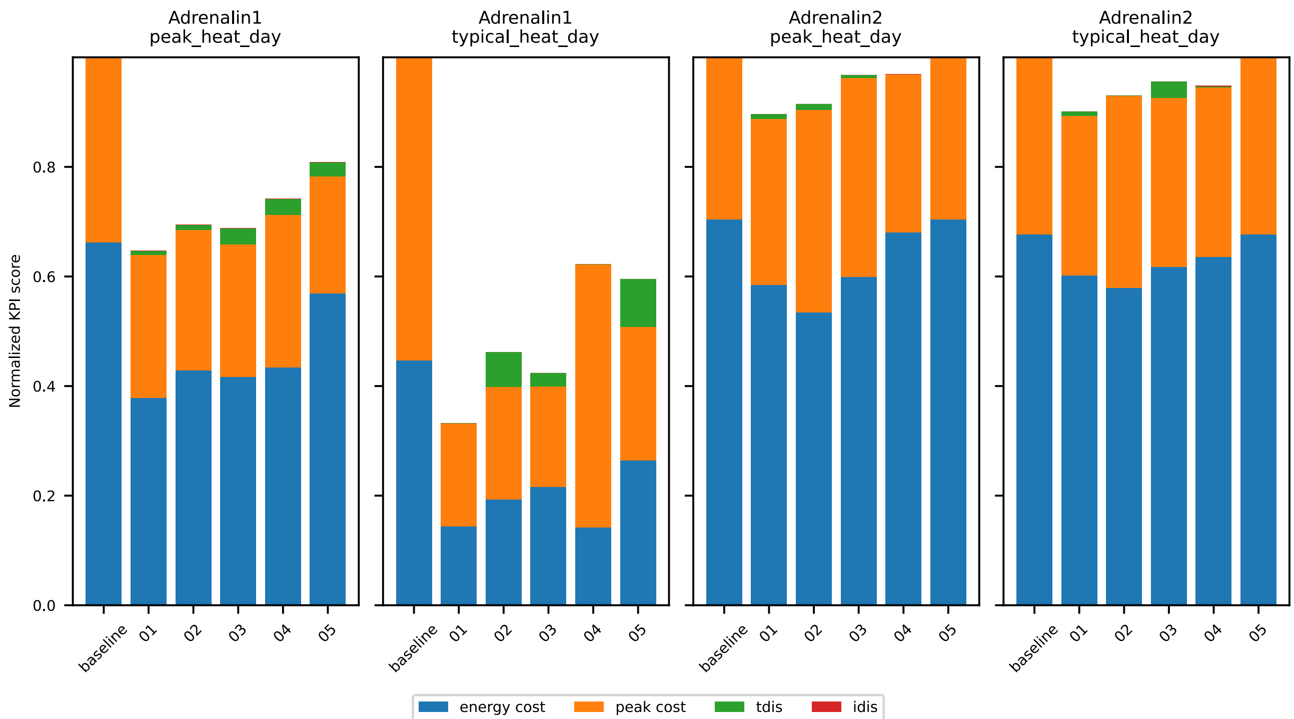


Figure 9. Disaggregation of KPI scores for the scenarios, normalized by the baseline scores.

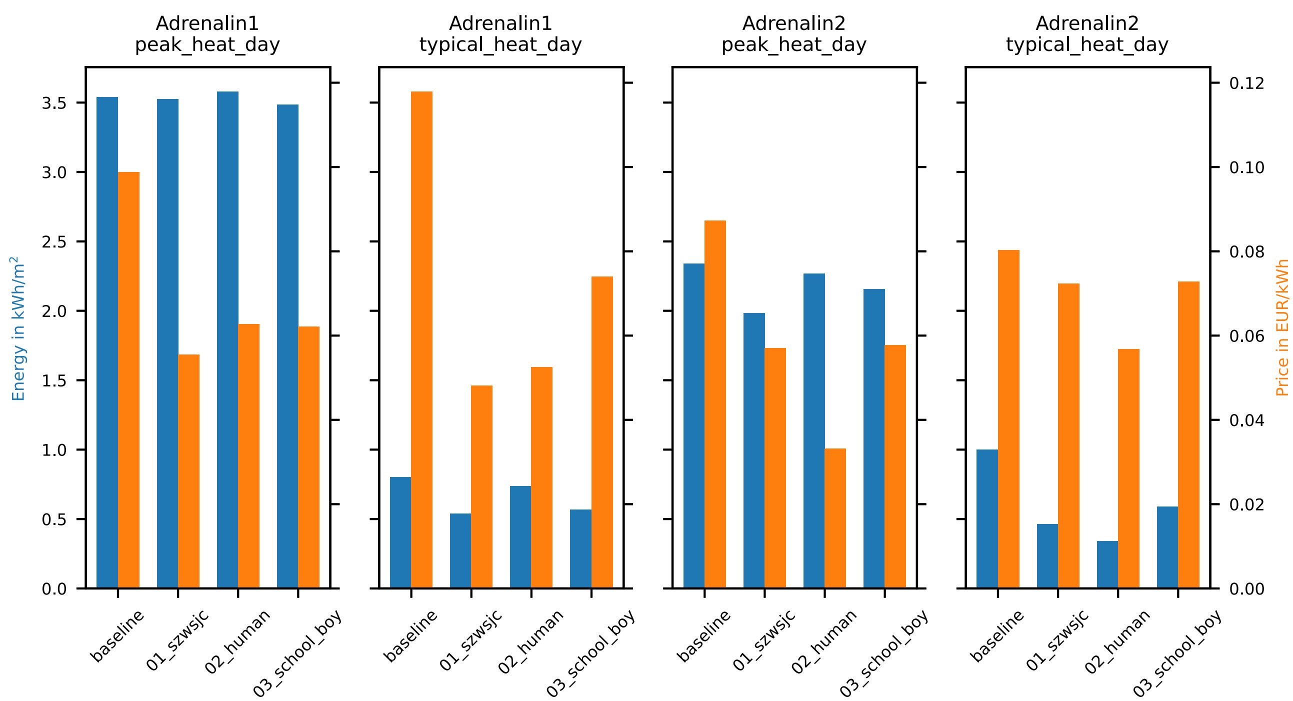


Figure 10. District heating energy use and average energy price for the different controller on all the scenarios.

average energy price; where a lower average energy price is achieved by shifting energy use to hours with lower prices. It only shows the DH related costs, as this is mainly what the controllers influence.

Looking at the typical heat day scenarios, it can be observed that for the Adrenalin1 emulator, which have a baseline controller with night setback, the energy cost savings of 55–73% is leveraged on modest energy use savings of 9–33% and a more significant average price reduction of 37–59%. Conversely, for the Adrenalin2 emulator, where the baseline did not have night setback, the energy cost savings of 47–76% is leveraged on significant energy use savings 41–66% but more modest energy price reduction of 9–29%.

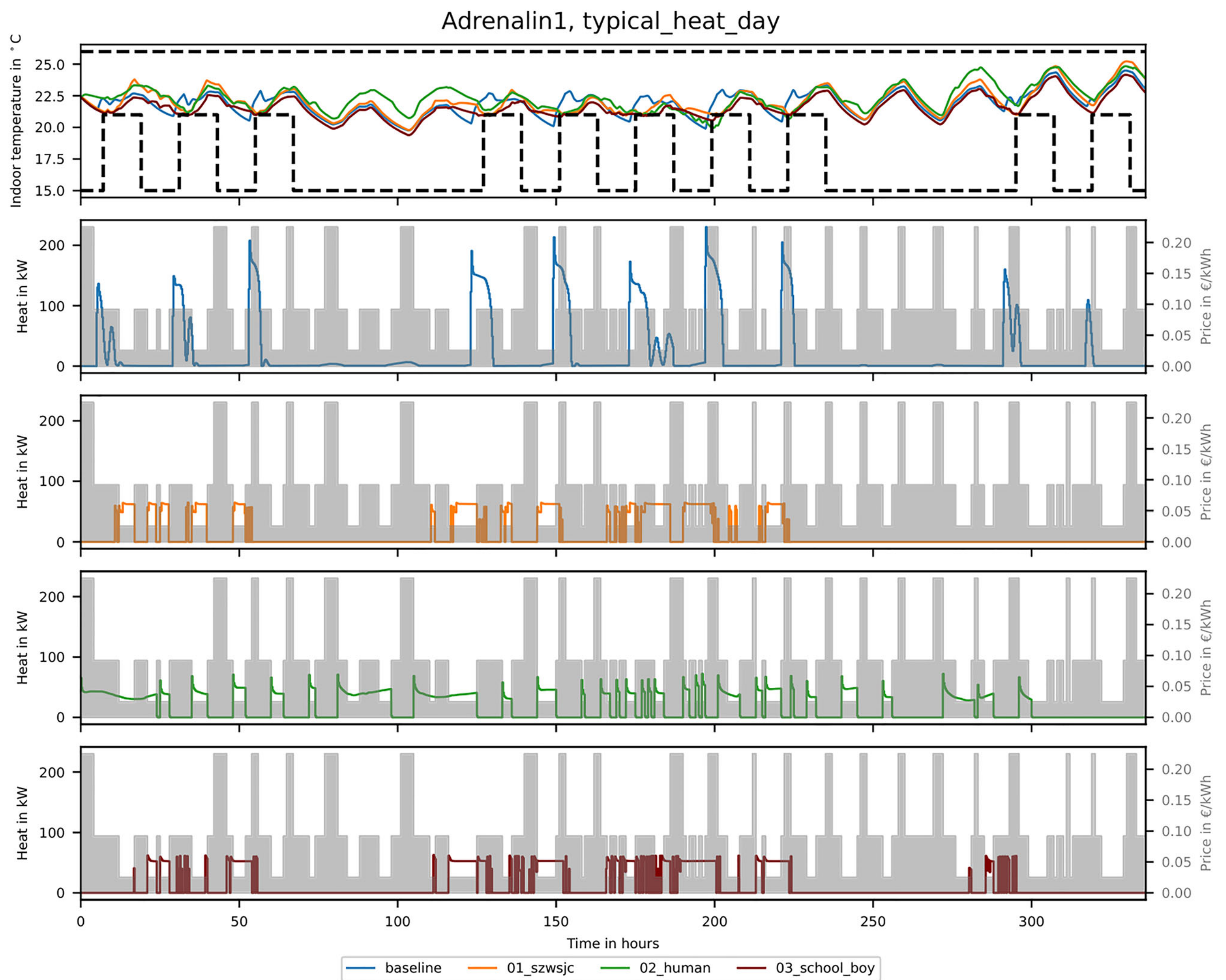


Figure 11. Comparison of indoor temperature (top plot) and heat use between the baseline (second plot) and the top 3 controllers (plot 3-5) on the Adrenalin1 emulator and typical heat day scenario. Measured heat use is resampled to 15 min mean. For the heat use plots, the price signal is shown in the background linked to the right axis.

Looking at the peak heat day scenarios, the energy savings from the smart controllers are considerably less, with ≈0 and 3–15% for Adrenalin1 and Adrenalin2, respectively, given that the heating system is already operating close to its full capacity in these cold periods. Nevertheless, their ability to shift the energy use to hours with lower prices is still significant with 36–44% and 34–62%, respectively (thus achieving 36–44% and 39–63% energy cost savings, respectively). This is especially interesting because it shows energy flexibility's potential to improve the utilization of energy distribution grids.

4.3.2. Control behaviour analysis

Figure 11 presents the measured indoor temperature (top plot) and heat demand (separately in the four lower plots) for Adreanlin1, typical_heat_day scenario, for the baseline and the top three submissions. It can be observed that the controllers shift energy to hours with lower prices and reduce the peaks. However, one should be aware that the heat demand is shown as a 15 min average, to match the input to Equation (1), and that the controllers create significantly higher peaks at higher frequencies. The reason for this can be observed in Figure 12, which shows the main control inputs used to control the heat, zoomed in on one day. All the top three controllers use a 15 min interval to create on/off cycles, both to ensure that the right amount of energy is emitted into the envelope and to reduce the 15 min peak. This creates valve cycling, which can result in increased wear and tear.

For Adrenalin1, the main control input applied in the solutions is the radiator valve position. In addition, solution 01 and 03 forces the ventilation heating coil valve to be closed, which in principle means turning off the heating battery in the ventilation. In cold periods, this result in low supply air temperatures, which can be uncomfortable for the occupants.

Table 7. Energy cost and peak cost for DH only. Values are in c€/m^2.

Testcase	Scenario	Participant	Energy cost	Peak cost	Total
Adrenalin1	peak_heat_day	baseline	34.9	16.9	51.9
		01_szwsjc	19.6	12.9	32.4
		02_human	22.4	12.6	35.0
		03_school_boy	21.7	11.8	33.4
	typical_heat_day	baseline	9.4	11.5	20.9
		01_szwsjc	2.6	3.2	5.8
		02_human	3.9	3.6	7.5
		03_school_boy	4.2	3.1	7.3
Adrenalin2	peak_heat_day	baseline	20.4	6.3	26.7
		01_szwsjc	11.3	6.9	18.2
		02_human	7.5	11.9	19.4
		03_school_boy	12.4	11.4	23.8
	typical_heat_day	baseline	8.0	4.0	12.1
		01_szwsjc	3.4	2.1	5.4
		02_human	1.9	5.8	7.7
		03_school_boy	4.3	3.1	7.4

Figure 13 shows the same measurements as Figure 11 but for Adrenalin2. The indoor temperature is shown as the mean of all the zones. As the baseline controller of Adrenalin2 does not include a night set-back, the controllers can exploit this to save energy. On the other side, the controllers' ability to shift energy use to hours with low prices is not as obvious as for Adrenalin1. This is especially true for the data-driven approaches (01 and 03). Although for Adrenalin2, the baseline also contains significant fluctuations in the heat demand, driven by the hysteresis control in the radiator valve models, the fluctuations are still more severe in the submitted controllers. One can also observe that submission 02 obtains slightly higher indoor temperatures, while Figure 10 shows lower energy consumption. This is because it exploits the possibility to control the solar shading, which none of the other two controllers do. However, it is questionable how this would affect occupant comfort.

Figure 14 shows also the main control inputs for Adrenalin2, zoomed in on one day of the peak_heat_scenario. Here, one can observe that the data-driven controllers (01 and 03) mainly use the main circulation pump as the control parameter. In Adrenalin2, the radiator valve position was not available as a control input, and controlling the main pump is a way to control the heating similarly as in Adrenalin1. However, it also means that the 5-zone building is fully controlled as a single zone. All the controllers use the same indoor temperature set point for all zones. While Figure 9 shows that the thermal discomfort violations are limited, there are significant variations in the indoor temperatures between the different zones. This indicates that there is potential for further energy savings by better control of the zone temperatures.

One can also observe that controller 03 operates the main pump pressure between 0 and 15,000 Pa, while the other controllers have a maximum pressure difference of 35,000 Pa. This is due to a bug in the emulator design, limiting the external control input to 15,000 Pa, while the baseline controller is fixed to 35 000 Pa. Controller 01 solves this by deactivating the control input to reset the pump pressure to baseline. The impact is expected to be limited on the control performance. One can see that the maximum heat output from controller 03 is higher or equal to the two other controllers.

5. Discussion on the competition hosting and outcome

5.1. Emulator design and scoring

Analysing the time-series plots above, it is clear that all the winning controllers create much more fluctuations both in set points and actuator positions. These fluctuations are not optimal for the system and its components, as they can increase wear and tear, thereby raising maintenance costs. For the competition, these are still considered good solutions, as there is no penalty in the score function for this behaviour. A KPI for actuator travel in BOPTEST is currently under development (Blum et al. 2025). This would allow for adding a penalty in the score function. However, it is not straightforward how to weight such wear and tear in a cost function. That all the top three solutions use heuristics in the generation of the control signal shows that the emulator design was able to highlight that modelling actuator response is a key challenge in data-driven control of buildings.

Another factor exploited by the controllers was that the evaluation of thermal comfort was based solely on dry bulb temperature. In Adrenalin1, this was exploited by turning off the ventilation heating and allowing low supply air temperatures, which can be perceived as uncomfortable by the occupants. In Adrenalin2, controller 02 fixed the solar shading in the open position to reduce heating demand, which can lead to both local thermal discomfort and blinding discomfort. Solar shading control is an interesting part of building HVAC control, but it is challenging to include the perceived comfort in the emulator models.

While the white box emulators used for the competition are able to model the known physical processes occurring in the building and the HVAC systems, the behaviour of a building is largely influenced by unknown and stochastic processes. The only stochastic elements that were included in the emulators were the occupancy profiles. This yields some stochasticity in the internal heat gains and CO_2 generation, but does not include any stochastic behaviour from the occupants. The weather forecast is perfect, and there is no sensor uncertainty. The stochastic nature of buildings does influence predictive

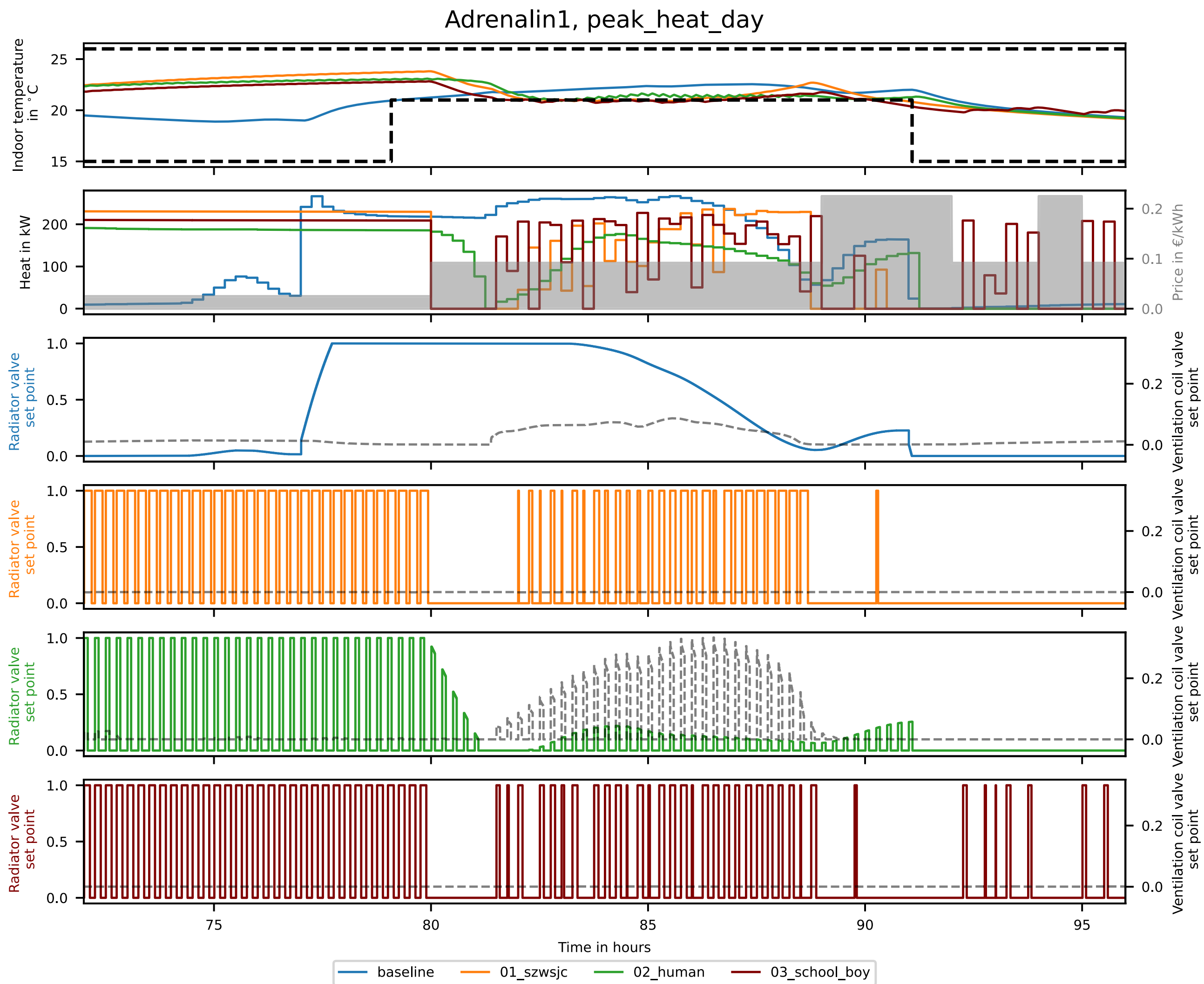


Figure 12. Comparison of performance between the top 3 controllers on the Adrenalin1 emulator and peak heat day scenario. Zoomed in on one day. The top plot shows the indoor temperature while the second shows the measured heat use resampled to 15 min mean. Plot 3–6 shows the radiator valve set point (left axis) and the ventilation heating coil valve set point for the baseline and top 3 controllers separately.

controller performance, and it is expected that some control approaches are more sensitive to noise than others. However, it is uncertain how a more realistic representation of stochasticity would influence the final results of the competition. There is an ongoing effort to include forecast uncertainty in the BOPTEST framework (Blum et al. 2025).

5.2. The participation and winning solutions

Based on the number of participants signing up for the competition (138), the number of valid submissions in the final stage of the competition (≈9%) is somewhat disappointing. There could be several reasons for the high dropout rate. It is assumed that a significant portion of the teams signing up were mainly attracted by the prize money and abandoned the competition when understanding the complexity of participating. Scalability of control approaches for building HVAC systems is a known challenge, meaning that for most developers, a significant effort would be required to submit a valid solution. Another reason could be the requirement to make the code used for the competition publicly available. With the necessary development effort and potential commercial value, this could be a decisive factor for some. The dropout between the training and final stage, might also be explained by the fact that participants could calculate their scores locally before submitting. Some participants might have been reluctant to submit their solution when seeing that they would not be able to compete for the top three positions. With the limited amount of valid submissions in the final stage of the competition, it is natural that only a limited share of the different modelling paradigms applied in advanced building control

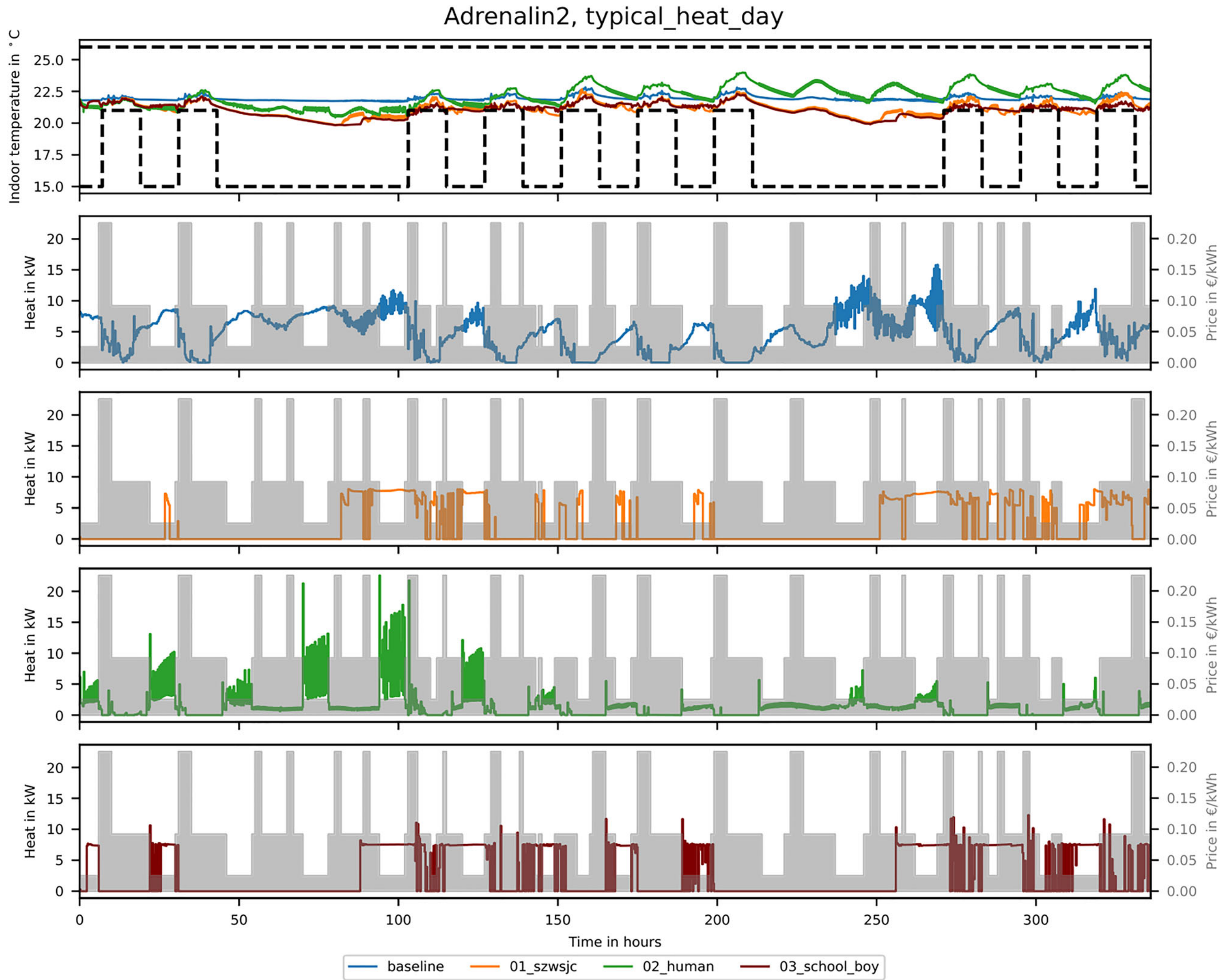


Figure 13. Comparison of mean indoor temperature (top plot) and heat consumption between the baseline (second plot) and the top 3 controllers (plot 3–5) on the Adrenalin2 emulator and typical heat day scenario. Measured energy is resampled to 15 min mean. For the heat consumption plots, the price signal is shown in the background linked to the right axis.

are represented. Since competitors were only required to document their solution when submitting in the final stage, it is not known what other approaches were tried by participants. However, it is known to the competition organizers that some RL based control attempts were made. The BOPTEST-Gym environment was made available for offline training of RL based algorithms in the training stage, without any limitations on the number of training data generation. Even with the extensive training possibilities achievable with the BOPTEST-gym environment and the offline emulator, the applied RL algorithms did not seem to be able to compete with the winning solutions.

5.3. Application of the results

One of the goals of the competition was to promote scalable solutions suitable for implementation in real buildings. However, ensuring this through competition design is challenging. One key design challenge was how to have a fair competition and mitigate the risk of cheating. It was not desirable to have rules that could not be enforced. Enabling the participants to run the emulators offline in the training phase allowed them to train or tune their controllers on a much richer data set than normally available for real buildings. Both the data-driven controllers in the top three solutions generate training data by applying bang-bang control signals to excite the building. This could be challenging with occupied real buildings. In addition, all the participants tuned their controllers by rerunning the scenarios multiple times, which is not feasible in reality. For the competition stage, the participants could not perform offline training, but there were no limitations on how many times they could run the scenarios in the time frame of the competition stage. For future development of the BOPTEST framework, for

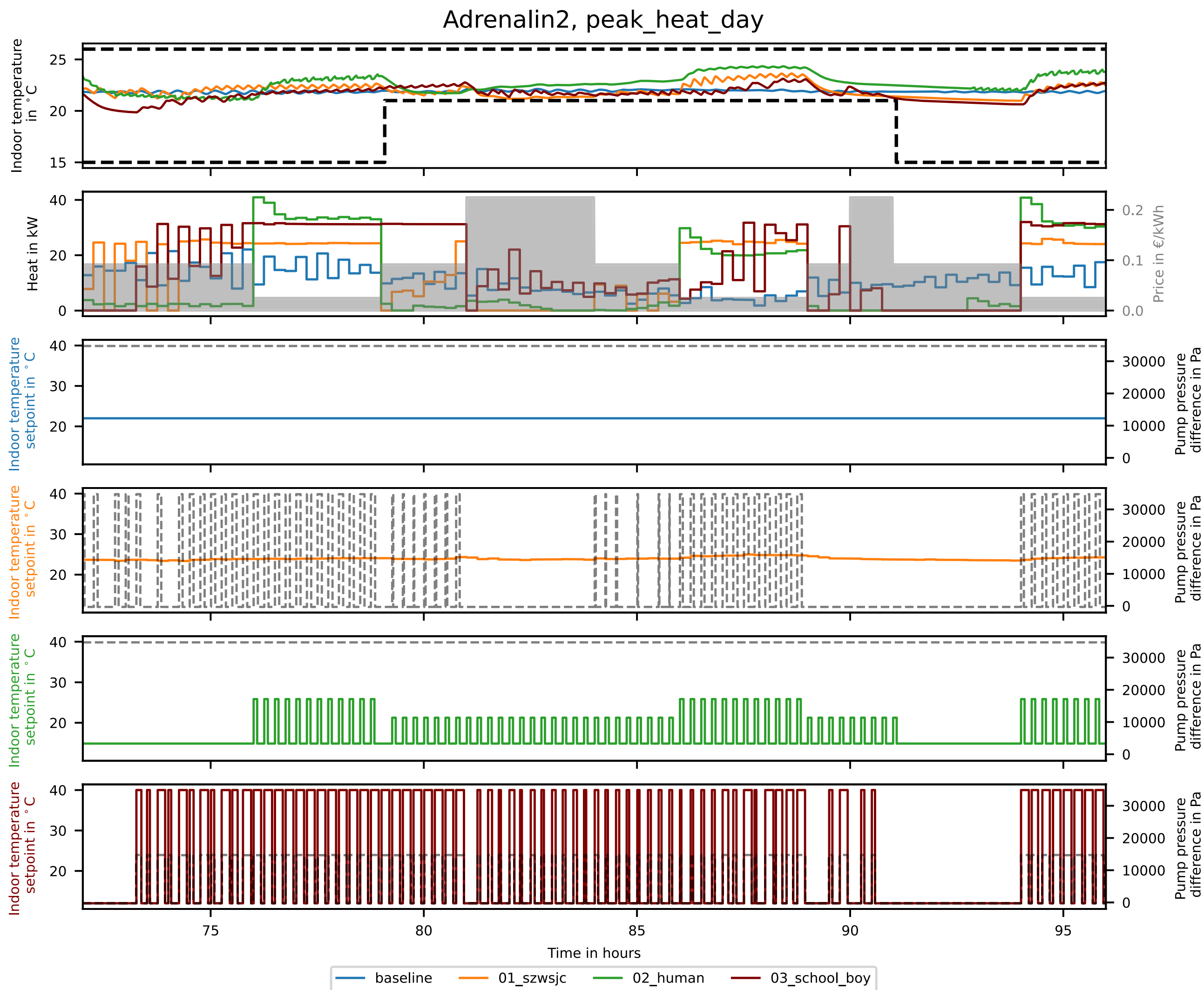


Figure 14. Comparison of control inputs from the top 3 controllers on the Adrenalin2 emulator and peak heat day scenario. Zoomed in on one day. The top plot shows the indoor temperature while the second shows the measured heat use resampled to 15 min mean. Plot 3–6 shows the indoor temperature set point (left axis) and the main pump pressure difference set point for the baseline and top 3 controllers separately.

hosting competitions, a method for limiting the number of scenario simulations for each user could be relevant. In principle, such a limitation could be enforced through the hosting platform, but it would require that the participants submits the code to run the control to the platform instead of running it directly and submitting the test_ids. That means that the participant would need to submit their full codebase in an executable environment, such as a docker container (Docker 2025), which would further increase the development demand. A key benefit of the BOPTEST framework is that the control development is platform and programming language independent. Keeping this property makes creation of Docker templates for the participants challenging.

6. Conclusion and future work

This study presented the design, execution, and outcomes of the Adrenalin BOPTEST Challenge, a crowd-sourcing effort aimed at benchmarking smart HVAC control algorithms under fair and transparent conditions. The competition attracted 138 participants. However, that only 9% submitted valid solutions for the final stage highlights the complexity of the task and the substantial development effort required to produce scalable control algorithms suitable for real-world deployment. The successful completion of the competition demonstrated that the BOPTEST framework, together with available competition hosting platforms, is well suited for the purpose, enabling comparison between fundamentally different control approaches. At the same time, the results highlighted some limitation in the emulator design and evaluation, including the lack of stochastic elements, perfect weather forecasts, simplified comfort models, and the absence of penalties for actuator wear.

The analysis of the submitted controllers (Section 4) demonstrates that data-driven and heuristic approaches can deliver significant improvements in energy cost and flexibility. Across scenarios, thermal energy cost reductions of 36–76% relative to a baseline, were achieved. These gains were obtained either through substantial energy savings (up to 66% for Adrenalin2, typical heat day) or by effectively shifting heating load away from high-price periods (up to 62% reductions in average energy price for Adrenalin2, peak heat day). At the same time, all top-ranked solutions relied on on/off cycling or rapidly varying setpoints, underscoring a current gap between competition-optimal behaviour and physically desirable operation.

Future work should focus on developing the competition framework to improve the usefulness of the results of future competitions. This includes developing more realistic emulators that include stochastic elements and uncertainties. Additionally, incorporating actuator wear and tear into the score function could provide a more comprehensive evaluation of controller performance. Finally, future competitions could benefit from stricter rules on the number of scenario simulations allowed, to better mimic real-world constraints in available training data.

Acknowledgements

This paper is part of the ADRENALIN (Data-driven smart buildings: data sandbox and competition) project, funded by the Joint Call 2020 (MICall20) for transnational projects on digital transformation for green energy transition. The funding partners participating in the Joint Call are from ERA-Net Smart Energy Systems (ERA-Net SES) and the Mission Innovation (MI) Initiative, coordinated by the Nordic Energy Research with prj.nr. 111406. This research was also supported by the Research Council of Norway, through Harald Taxt Walnum's Institute PhD grant (342617) at SINTEF. This research was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Building Technologies of the U.S. Department of Energy, under Contract No. DE-AC02-05CH11231. Jicheng Shi and Colin Jones received support from the Swiss National Science Foundation (SNSF) under the NCCR Automation project, grant agreement 51NF40_180545. The work of Wenbin Wang was supported by the Swiss Federal Office of Energy SFOE as part of the SWEET consortium SWICE, grant agreement SI/502495-01. The work of Wenjie Wu was supported by the Swiss National Science Foundation under NCCR Automation, grant agreement 51NF40_180545.

This work emerged from the IBPSA Project 2, an international project conducted under the umbrella of the International Building Performance Simulation Association (IBPSA) to develop and demonstrate the Building Optimization Testing Framework (BOPTEST) for the testing, evaluating, and benchmarking of building and community energy system controls.

Author contributions

CRediT: **Harald Taxt Walnum:** Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Balázs András Tolnai:** Methodology, Software, Writing – original draft, Writing – review & editing; **David Blum:** Methodology, Software, Writing – original draft, Writing – review & editing; **Jicheng Shi:** Investigation, Software, Writing – original draft, Writing – review & editing; **Colin N. Jones:** Supervision, Writing – review & editing; **Deep Dessai:** Investigation, Software, Writing – original draft, Writing – review & editing; **Wenbin Wang:** Investigation, Software, Writing – original draft, Writing – review & editing; **Wenjie Xu:** Investigation, Software, Writing – original draft, Writing – review & editing; **Sebastien Gros:** Conceptualization, Supervision, Writing – review & editing; **Igor Sartori:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing

Copyright notice

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This paper is part of the ADRENALIN (Data-driven smart buildings: data sandbox and competition) project, funded by the Joint Call 2020 (MICall20) for transnational projects on digital transformation for green energy transition. The funding partners participating in the Joint Call are from ERA-Net Smart Energy Systems (ERA-Net SES) and the Mission Innovation (MI) Initiative, coordinated by the Nordic Energy Research with prj.nr. 111406. This research was also supported by the Research Council of Norway, through Harald Taxt Walnum's Institute PhD grant (342617) at SINTEF. This research was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Building Technologies of the U.S. Department of Energy, under Contract No. DE-AC02-05CH11231. Jicheng Shi and Colin Jones received support from the Swiss National Science Foundation (SNSF) under the NCCR Automation project, grant agreement 51NF40_180545. The work of Wenbin Wang was supported by the Swiss Federal Office of Energy SFOE as part of the SWEET consortium SWICE, grant agreement SI/502495-01. The work of Wenjie Wu was supported by the Swiss National Science Foundation under NCCR Automation, grant agreement 51NF40_180545.

ORCID

Harald Taxt Walnum https://orcid.org/0000-0002-8718-6797
David Blum http://orcid.org/0000-0003-3231-7937
Sebastien Gros https://orcid.org/0000-0001-6054-2133
Igor Sartori https://orcid.org/0000-0002-5537-0020

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Adrenalin. 2026. "Adrenalin – BOPTEST Challenge: Smart Building HVAC Control." Accessed: March 2, 2026. https://data.sintef.no/feature/fe-7d9861ec-ae6e-4947-ae43-85f27fee3cd3.

ADRENALIN project. 2026. "ADRENALIN Project.' Accessed: March 2, 2026. https://adrenalin.energy/.

Afroz, Z., G. Shafiullah, T. Urmee, and G. Higgins. 2018. "Modeling Techniques Used in Building Hvac Control Systems: A Review." *Renewable and Sustainable Energy Reviews* 83:64–84. https://doi.org/10.1016/j.rser.2017.10.044.

Arroyo, J., C. Manna, F. Spiessens, and L. Helsen. 2021. "an OpenAI-Gym Environment for the Building Optimization Testing (BOPTEST) Framework." In *Proceedings of the 17th IBPSA Conference*, September. Belgium: Bruges.

Bird, M., C. Daveau, E. O'Dwyer, S. Acha, and N. Shah. 2022. "Real-World Implementation and Cost of a Cloud-Based MPC Retrofit for HVAC Control Systems in Commercial Buildings." *Energy and Buildings* 270:112269. https://doi.org/10.1016/j.enbuild.2022.112269.

Blum, D., J. Arroyo, S. Huang, J. Drgoňa, F. Jorissen, H. T. Walnum, Y. Chen, et al. 2021. *Building Optimization Testing Framework (BOPTEST) for Simulation-Based Benchmarking of Control Strategies in Buildings*. https://www.tandfonline.com/doi/abs/10.1080/19401493.2021.1986574.

Blum, D., J. Arroyo, S. Huang, J. Drgoňa, F. Jorissen, H. T. Walnum, Y. Chen, et al. 2021. "Building Optimization Testing Framework (boptest) for Simulation-Based Benchmarking of Control Strategies in Buildings." *Journal of Building Performance Simulation* 14:586–610. https://doi.org/10.1080/19401493.2021.1986574.

Blum, D., X. Lu, L. Z. Urrutia, W. Zheng, H. T. Walnum, J. Arroyo, E. Borkowski, et al. 2025. "IBPSA Project 2 Building Optimization Testing Framework (BOPTEST): Overview, New Software Features, and Example Use Cases." In *Proceedings of Building Simulation 2025: 19th Conference of IBPSA*, Aug., Brisbane, Australia: IBPSA, (to Appear).

Blum, D., Z. Wang, C. Weyandt, D. Kim, M. Wetter, T. Hong, and M. A. Piette. 2022. "Field Demonstration and Implementation Analysis of Model Predictive Control in an Office HVAC System." *Applied Energy* 318:119104. https://doi.org/10.1016/j.apenergy.2022.119104.

COM(2020) 299 final. 2020. "Powering a Climate-Neutral Economy: An EU Strategy for Energy System Integration. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri = celex:52020DC0299.

Deru, M., K. Field, D. Studer, K. Benne, B. Griffith, P. Torcellini, B. Liu, et al. 2011. *U.S. Department of Energy Commercial Reference Building Models of the National Building Stock*, Tech. Rep. NREL/TP-5500-46861. Golden, CO (United States): National Renewable Energy Lab. (NREL). https://www.osti.gov/biblio/1009264.

Directive (EU). 2024. "Directive (EU) 2024/1275 of the European Parliament and of the Council of 24 April 2024 on the Energy Performance of Buildings (recast). https://eur-lex.europa.eu/eli/dir/2024/1275/oj.

Docker. 2025. Available at https://www.docker.com/.

Donoho, D. 2024. "Data Science at the Singularity." *Harvard Data Science Review* 6.

Drgoňa, J., J. Arroyo, I. Cupeiro Figueroa, D. Blum, K. Arendt, D. Kim, E. P. Ollé, et al. 2020. "All You Need to Know about Model Predictive Control for Buildings." *Annual Reviews in Control* 50:190–232. https://doi.org/10.1016/j.arcontrol.2020.09.001.

IEA, GlobalABC, and UN Environmental Programme. 2020. "2020 Global Status Report for Buildings and Construction: Towards a Zero-Emissions, Efficient and Resilient Buildings and Construction Sector – Executive Summary. Accepted: 2020-12-15T20:46:48Z. https://wedocs.unep.org/xmlui/handle/20.500.11822/34572.

Jradi, M., F. C. Sangogboye, C. G. Mattera, M. B. Kjærgaard, C. Veje, and B. N. Jørgensen. 2017. "A World Class Energy Efficient University Building by Danish 2020 Standards." *Energy Procedia* 132:21–26. https://doi.org/10.1016/j.egypro.2017.09.625.

Kathirgamanathan, A., M. De Rosa, E. Mangina, and D. Finn. 2021. "Data-Driven Predictive Control for Unlocking Building Energy Flexibility: A Review." *Renewable and Sustainable Energy Reviews* 135:110120https://doi.org/10.1016/j.rser.2020.110120.

Nweye, K., K. Kathryn, B. Giacomo, F. Tiago, P. Giuseppe, G. Dipanjan, D. Satvik, et al. 2025. "CityLearn V2: Energy-Flexible, Resilient, Occupant-Centric, and Carbon-Aware Management of Grid-Interactive Communities." *Journal of Building Performance Simulation* 18:17–38. https://doi.org/10.1080/19401493.2024.2418813.

Pavao, A., I. Guyon, A. C. Letournel, D. T. Tran, X. Baro, H. J. Escalante, S. Escalera, T. Thomas, and Z. Xu. 2023. "Codalab Competitions: An Open Source Platform to Organize Scientific Challenges." *Journal of Machine Learning Research* 24:1–6.

Sandberg, N. H., I. Sartori, O. Heidrich, R. Dawson, E. Dascalaki, S. Dimitriou, T. Vimm-r, et al. 2016. "Dynamic Building Stock Modelling: Application to 11 European Countries to Support the Energy Efficiency and Retrofit Ambitions of the EU." *Energy and Buildings* 132:26–38. https://doi.org/10.1016/j.enbuild.2016.05.100.

Satchwell, A., M. A. Piette, A. Khandekar, J. Granderson, N. M. Frick, R. Hledik, A. Faruqui, et al., 2021. "*A National Roadmap for Grid-Interactive Efficient Buildings*." https://gebroadmap.lbl.gov/.

Shi, J., Y. Lian, and C. N. Jones. 2024. "Efficient Recursive Data-Enabled Predictive Control." In *2024 European Control Conference (ECC)*, 880–887. IEEE.

Shi, J., Y. Lian, C. Salzmann, and C. N. Jones. 2025. "Adaptive Data-Driven Prediction in a Building Control Hierarchy: A Case Study of Demand Response in Switzerland." *Energy and Buildings* 333. https://doi.org/10.1016/j.enbuild.2025.115498.

Standard Norge. 2020. *SN-NSPEK 3031:2020*, 168. Oslo.

Togashi, E., T. Kobayashi, H. Ogata, H. Hatada, H. Ayame, K. Edahiro, M. Satoh, et al. 2024. "Development of a Variable Refrigerant Flow System Emulator to Host the Second World Championship in Cybernetic Building Optimization." *Journal of Building Performance Simulation* 17:480–509. https://doi.org/10.1080/19401493.2024.2346528.

Togashi, E., M. Masato, and Y. Yamamoto. 2020. "The First World Championship in Cybernetic Building Optimization." *Journal of Building Performance Simulation* 13:391–408. https://doi.org/10.1080/19401493.2020.1741685.

Togashi, E., and M. Miyata. 2019. "Development of Building Thermal Environment Emulator to Evaluate the Performance of the HVAC System Operation." *Journal of Building Performance Simulation* 12: 663–684.

Vázquez-Canteli, J. R., S. Dey, G. Henze, and Z. Nagy. 2020. "the CityLearn Challenge 2020." In *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, Nov., BuildSys '20, 2020, 320–321. New York, NY, USA: Association for Computing Machinery. https://dl.acm.org/doi/10.1145/3408308.3431122.

Walnum, H. T., I. Sartori, P. Ward, and S. Gros. 2025. "Demonstration of a Low-Cost Solution for Implementing MPC in Commercial Buildings with Legacy Equipment." *Applied Energy* 380:125012. https://doi.org/10.1016/j.apenergy.2024.125012.

Xin, X., Z. Zhang, Y. Zhou, Y. Liu, D. Wang, and S. Nan. 2024. "A Comprehensive Review of Predictive Control Strategies in Heating, Ventilation, and Air-Conditioning (HVAC): Model-Free VS Model." *Journal of Building Engineering* 94: 110013.

Yang, T., K. Filonenko, K. Arendt, and C. Veje. 2020. "Implementation and Performance Analysis of a Multi-energy Building Emulator." In *2020 6th IEEE International Energy Conference (ENERGYCon)*, Sep. 2020, 451–456. https://ieeexplore.ieee.org/document/9236623.

Yao, Y., and D. K. Shekhar. 2021. "State of the Art Review on Model Predictive Control (MPC) in Heating Ventilation and Air-Conditioning (HVAC) Field." *Building and Environment* 200:107952. https://doi.org/10.1016/j.buildenv.2021.107952.