Abstract
Building upon prior research that highlighted the need for standardizing environments for building control research, and inspired by recently introduced benchmarks for real life reinforcement learning control, here we propose a non-exhaustive nine real world challenges for reinforcement learning building controller. We argue that building control research should be expressed in this framework in addition to providing a standardized environment for repeatability. Advanced controllers such as model predictive control and reinforcement learning control have both advantages and disadvantages that prevent them from being implemented in real world buildings. Comparisons between the two are seldom, and often biased. By focusing on the benchmark problems and challenges, we can investigate the performance of the controllers under a variety of situations and generate a fair comparison. Lastly, we call for a more interdisciplinary effort of the research community to address the real world challenges, and unlock the potentials of advanced building controllers.

Keywords
Building Control, Benchmarking, , Reinforcement Learning

1 Introduction
Buildings account for ∼ 40% of the global energy consumption and ∼ 30% of the associated greenhouse gas emissions, while also offering a 50—90% CO₂ mitigation potential [12]. Optimal decarbonization requires electrification of end-uses and concomitant decarbonization of electricity supply, efficient use of electricity for lighting, space heating, cooling and ventilation (HVAC), and domestic hot water generation, and upgrade of the thermal properties of buildings [11]. A major driver for grid decarbonization is integration of renewable energy systems (RES) into the grid (supply), and photovoltaics (PV) and solar-thermal collectors into residential and commercial buildings (demand). Electric vehicles (EVs), with their storage capacity and inherent connectivity, hold a great potential for integration with buildings [13]. However, this grid building integration must be carefully managed during operation to ensure reliability and stability of the grid [8, 20].

Demand response (DR) as an energy-management strategy enables end-consumers to provide the grid with more flexibility by reducing their energy consumption through load curtailment, shifting their energy consumption over time, or generating and storing energy at certain times. In exchange, consumers typically receive a reduction of their energy bill [18]. HVAC can contribute to load curtailment events by modifying the temperature set points, participating in load shifting by pre-heating or pre-cooling the buildings [2] (passive energy storage), or by directly storing thermal energy in an energy storage system (active energy storage). Thermostats with DR functionality can provide energy savings to residential customers by allowing electricity retailing companies to adjust set-points during peak-demand events. Widespread integration of communication technologies allows all involved systems (PV, HVAC, storage, EVs, thermostats, etc) to exchange information on their operation, leading to the concept of smart cities, allowing cities to achieve energy savings, and become more efficient, livable, and sustainable [3].

Advanced control systems can be a major driver for DR by automating the operation of energy systems, while adapting to individual characteristics of occupants and buildings. However, for DR to be effective, loads must be controlled in a responsive, adaptive and intelligent way. When all the electrical loads react simultaneously to the same price signals, aggregated electricity peaks could be shifted rather than shaved. Therefore, there is a need for more efficient and effective ways of coordinating the response of all the technologies described above.

Advanced control algorithms such as model predictive control (MPC) [5] and deep reinforcement learning (RL) [21] have been proposed for a variety of building control applications. While both methods have their disadvantages, e.g., MPC requiring a model while RL being data intensive, spectacular applications and results have been presented in the past several years. In addition, recently, hybrid methods, based on physics constrained neural networks for models have begun to emerge [6].

A major challenge for this research area is the ability to compare and benchmark results. As argued in [22], a
Building on [22], and inspired by [7], the purpose of this paper is to introduce and discuss specific real world benchmark problems for building control that our community should be focusing on, to advance the field in a meaningful way.

2 Real-world challenges

Dulac-Arnold et al. provide nine real-world challenges for control systems [7]. The benchmarks they present are not suitable to evaluate building control systems as they are based on small scale environments without the necessary domain knowledge or context. Also while [7] argue the real-world challenges in the context of reinforcement learning, they hold true for any advanced control algorithm, and so we proceed without the assumption for a specific control paradigm.

In the following, we present a first set of challenges, we provide the description of [7] in italics.

C1: Being able to learn on live systems from limited samples: In this benchmark, the controller is initialized randomly and has to learn to perform only based on the samples it observes. The sample size can be artificially reduced by presenting the controller only with a subset of the data, e.g., every 3 hours instead of every 15min. The algorithms can be evaluated on how quickly in terms of time or sample number they converge, and how stable their exploration is. Conversely, we can evaluate the trade-off between data requirement and controller performance.

C2: Dealing with unknown and potentially large delays in the system actuators, sensors, or feedback: The thermal dynamics of buildings are such that the effects of controller actions to adjust the HVAC systems are observed in delays. This has implications for, e.g., pre-cooling/heating of buildings to take advantage of the thermal mass of buildings. The controller need to implicitly and automatically learn the dynamics of the building. Benchmark models with different thermal mass from light to heavy should be created, and the converged controller should be compared to understand the relationship between longer delays in feedback (higher thermal mass) and controller performance.

C3: Learning and acting in high-dimensional state and action spaces: This benchmark addresses the scalability of a proposed controller. As buildings can inherently have a large state-action space, controller can be evaluated on specific subsets of them to understand how the performance changes. In the case of controlling multiple buildings (or multiple zones within a building), scalability refers to essentially increasing the number of buildings (or zones) and observe the control performance.

C4: Reasoning about system constraints that should never or rarely be violated: This is a central benchmark as building control problems are indeed often presented as balancing between reducing energy use while maintaining comfortable conditions. Other constraints in the energy system are operational, such as ensuring a minimum state of charge, maintaining operational temperatures within limits, etc. The algorithms should be evaluated on both the number of violations during the learning process as well as for the converged policy. Integration of constraint violation into the objective function is addressed in B6 below.

C5: Interacting with systems that are partially observable, which can alternatively be viewed as systems that are non-stationary or stochastic: This benchmark has two parts. In the first part, increasing levels of Gaussian noise should be added to the observations and actions of the agent to simulate sensor failure, which can be common in any real life systems, like buildings and HVAC systems. We then observe the performance of the algorithms for the various levels of noise. In the second part, we can observe how a controller performs on a perturbed system. Perturbations can consist of retrofit measures on buildings (improving envelop or windows), improving equipment, changed occupant behavior or different climate. We can then judge the algorithms on their ability to perform their previously learned policy on the perturbed system.

C6: Learning from multi- or poorly specified objective functions: Energy management in buildings in inherently multi-objective, especially when considering multiple zones or multiple buildings. Another example is when there is a global objective (overall building energy use) as well as multiple local objectives (equipment operation). As mentioned in B4, constraints can be incorporated into the objective function directly. When evaluating the controller performance, the individual objectives should be separated to allow for a fair comparison.

C7: Being able to provide actions quickly, especially for systems requiring low latencies: Latency is a delay in executing a control action after acquiring a measurement due to long computational time. Latencies in real life systems can occur if the system dynamics are fast or the computational times long. A practical example for smart buildings and microgrids is if the computation is taking place in the cloud adding also data transfer to the execution time, which can be exacerbated by connectivity issues. To observe the impact of latency, time-step delays of various lengths should be included into the control execution and the impact on their performance should be evaluated.

C8: Training off-line from the fixed logs of an external behavior policy: The benchmark here is to learn a control law from data generated by a sub-optimal reference controller, e.g., a rule based controller, which is often available, essentially a system log. In addition to the control environment, datasets of various size, e.g., two weeks, one months, 6 months should be provided that are generated with a known reference rule based controller. Then, the controllers can be evaluated on the ability to improve these baselines.

C9: Providing system operators with explainable policies: Here we deviate from the description in [7] who propose to generate figures to improve the interpretability of the results. Rather, for the building context, what is needed is that the control actions can be explained simply to building managers. Advances in explainable AI are needed, and algorithms that
might perform suboptimally, yet are easier to explain are favored as they are more likely to get accepted, and thus implemented. A consensus between modelers and system operators on the standards and outcomes of a control law could be established to facilitate effective communication amongst invested parties.

Challenges for advanced building controllers, except for the most trivial of cases, are almost always a combination of the challenges C1—C9 above compounding the complexity. We describe several in the following.

**Building Managers** Most of commercial buildings are operated by a dedicated maintenance group, e.g., facility managers. Because the operation cost of building systems takes a significant portion in large commercial buildings, it is essential to include facility managers when we develop and deploy advanced building controllers in commercial buildings, and provide explanations of the algorithms outputs and decisions (C9). In addition, there are numerous safety constraints in commercial building systems (e.g., ventilation rate, cooling and heating capacity), often required by building standard. Especially RL agent sometimes may explore the action space and decide on control actions that may lead to a long term cumulative reward, however, it cannot be accepted by facility managers and building standards (C4, B6). Facility managers are also intimately familiar with their building, therefore control algorithms (and their designers) should strive to include such domain knowledge, which brings us to:

**Sequence of Operation** Typical building HVAC systems have hundreds if not thousands of datapoints from which the state-action space are derived (C9). Therefore some dimensionality reduction without compromising on understanding the system is necessary. The HVAC systems are operated using rule based controllers, which are programmed in the building automation system (BAS) or building energy system (BES). One objective of advanced controllers should be to learn this control strategy from a combination of live data (C1) and logged data (C7), to be able to improve upon it and/or have it as a fall back option. In addition, the rate at which decisions need ideally be taken by the controller might differ from the rate of data storage and communication capabilities of the BAS/BAS (C7).

**Occupant-Centric Control (OCC)** The main purpose of HVAC control whether commercial or residential is to provide a healthy and comfortable environment for the occupants. However, most advanced controllers are focused solely on energy savings and reduce the occupant to some imaginary comfort level, often modeled as a target temperature set-point or temperature band to be maintained by the controllers [15]. This may lead to discomfort and more often than not, occupants deactivate or override controllers leading in the end of energy waste. In recent years, this situation has started to shift with the focus on OCC attempting to quantify or adapt to the preferences of the occupants without overburdening them [16]. However, it is challenging to design good OCC controllers as the comfort level of the occupants, or their goals must be inferred from few interactions (C2), can vary or even shift over time (C5, B9). In addition, because it is challenging to model occupant behavior and occupant-building interaction, experimental investigations are the preferred way, ideally in long-term, large, and double-blind or randomized experiments.

**Thermal/Electric Microgrids** The next generation building systems will be interacting with the electric grid, provide energy flexibility services to the grid, and incorporate multiple types of renewable resources, often in coordination with several buildings in the neighborhood, i.e., grid-interactive energy efficient buildings [4]. Designing advanced controllers for this environment requires defining and organizing the multiple objectives (C6) with multiple stakeholders (building owners, operators, grid operators), and handling events, on substantially different timescales, for example charging of EVs vs thermostat control (C9). Each individual building has also its very distinct demand for electricity driven mainly by the occupants behavior, and so orchestrating several tens to hundreds of buildings with potentially shifting occupant preferences becomes a herculean task (C7, C3, C4, C6, C9).

### 3 Discussion

From the descriptions in the previous section it is apparent that building systems, especially when considered as part of a larger view incorporating the occupants as well as grid operators are a complex system-of-systems. The extreme version being linking occupant behavior of various buildings (mixed commercial/residential) through their building systems (storage, renewable generation) to grid benefits (peaks shift, shaving, etc).

Advanced building controllers are needed to improve upon the industry standard of pre-determined set-points, that do not take into account predictions or allow to optimize the operational sequence [21]. Model predictive control (MPC) has been developed in the petrochemical industry in the 1970s and applied across many industries since then [14]. MPC requires the development of a mathematical model for the plant to be controlled, which works well for replicable systems (cars, planes). The uniqueness of buildings and their energy systems, and the engineering costs incurred when for developing and calibrating a model, however, made it such that despite all advances, MPCs have not been adopted in the building industry [10, 17]. Reinforcement learning algorithms have been considered to address the shortcomings of MPC by potentially being model-free. However, RL approaches can be more data intensive and more time-consuming compared to MPCs. Comparisons, if even performed, are often biased toward one type of algorithm, and therefore relatively meaningless. The benchmarks introduced here specifically focus on the breadth of applications rather than on one specific problem. This will allow for a fair comparison.

The most commonly solved problem in building control is energy demand optimization under comfort constraints which relates to C4,C6. As noted, many of the challenges appear simultaneously in real life, and given the complexity it is likely that not a single algorithm will be performing best at all of them. Comparing them across a variety of benchmarks will help solidify knowledge and create progress.
In particular, we emphasize the need for standardizing computational environments, such as the COmprehensive Building simulator (COBS) [23] or CityLearn [19] using a common interface, e.g., OpenAI Gym [1], and releasing datasets and implementations open source. This can help spark a development rush similar to the one that the ImageNet dataset sparked for the deep learning community [9].

However, in contrast to ImageNet’s development, a more in-depth collaboration and exchange between researchers in the built environment and computer science would be beneficial to transfer domain knowledge from buildings to controller design on the one hand and facilitating transitioning theoretical findings of algorithms into practice on the other. Common venues or guest invitation to each others venues could be established: ACM’s BuildSys/energy and the ASHRAE/IBPSA communities should explore common pathways for knowledge exchange to ultimately unlock the built environments potential to reduce greenhouse gas emissions.

4 Conclusion

We have introduced a set of benchmark problems for real world building control problem. While there are many research challenges that remain in this realm, this short article attempted to highlight the need for an organized move forward of the community in addressing both fundamental computational challenges, but in a way that applies to the larger problems in the built environment. It is not our intention to imply that the list above is an exhaustive list of benchmarks. Rather, by highlighting typical real world problems, our aim is to inspire researchers to define their environments and the problems they are addressing with these benchmarks as a standard framework.

References

[1] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Iie Tang, and Wojciech Zaremba. 2016. OpenAI Gym. arXiv preprint: 1606.01540 (6 2016). http://arxiv.org/abs/1606.01540
[2] Kenneth Bruninx, Dieter Patteeuw, Erik Delarue, Lieve Helsen, and William D’Haeseleer. 2013. Short-term demand response of flexible electric heating systems: The need for integrated simulations. International Conference on the European Energy Market, EEM May (2013), 28–30. https://doi.org/10.1109/EEM.2013.6607333
[3] Hadeh Choureshi, Taewoo Nam, Shawn Walker, J. Ramon Gil-Garcia, Sehl Meliouli, Karine Nahon, Theresa A. Pardo, and Hans Jochen Scholl. 2011. Understanding smart cities: An integrative framework. Proceedings of the Annual Hawaii International Conference on System Sciences (2011), 2289–2297. https://doi.org/10.1109/HICSS.2012.615
[4] DOE. 2021. A National Roadmap for Grid-Interactive Efficient Buildings. (2021), 166.
[5] Ján Drgoňa, Javier Arroyo, Iago Cepero Figueroa, David Blum, Krysztof Arendi, Donghun Kim, Enric Perarnau Ollé, Juraj Oravecz, Michael Wetter, Draguna L. Vrabie, and Lieve Helsen. 2020. All you need to know about model predictive control for buildings. Annual Reviews in Control 50, May (2020), 190–232. https://doi.org/10.1016/j.arcont.2020.09.001
[6] Ján Drgoňa, Aaron R. Tuor, Vikas Chandan, and Draguna L. Vrabie. 2020. Physics-constrained deep learning of multi-zone building thermal dynamics. Energy and Buildings 243 (2021). https://doi.org/10.1016/j.enbuild.2021.110992
[7] Gabriel Dulac-Arnold, Nir Levine, Daniel J. Mankowitz, Jerry Li, Cosmin Paduraru, Sven Gowal, and Todd Hester. 2021. Challenges of real-world reinforcement learning: definitions, benchmarks, and analysis. Number 0123456789. Springer US. https://doi.org/10.1007/s10994-021-05961-4
[8] B. Dupont, K. Dietrich, C. De Jonghe, A. Ramos, and R. Belmans. 2014. Impact of residential demand response on power system operation: A Belgian case study. Applied Energy 122 (2014), 1–10. https://doi.org/10.1016/j.apenergy.2014.02.022
[9] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 248–255. https://doi.org/10.1109/CVPRW.2009.5268489
[10] Georgios D. Kontes, Georgios I. Giannakis, Victor Sánchez, Pablo de Agustín-Camacho, Ander Romero-Amororuto, Natalia Panagiotidou, Dimitrios V. Rovas, Simone Steiger, Christopher Mutschler, and Gunnar Greus. 2018. Simulation-based evaluation and optimization of control strategies in buildings. Energies 11, 12 (2018), 1–23. https://doi.org/10.3390/en11123576
[11] Benjamin D. Leibowitz, Christopher M. Landau, Max T. Brozynski, Jose R. Vazquez-Canteli, Nicolas Castillo Castejon, and Zoltan Nagy. 2018. Optimal decarbonization pathways for urban residential building energy services. Applied Energy 230, May (2018), 1311–1325. https://doi.org/10.1016/j.apenergy.2018.09.046
[12] O. Lacon and D. Urry-Vorsatz. 2014. Fifth Assessment Report, Mitigation of Climate Change. Intergovernmental Panel on Climate Change (2014), 674–738.
[13] S Mohagheghi, J Stoupis, Z. Wang, and Z. Li. 2012. Demand Response Architecture-Integration into the Distribution Management System. SmartGridComm (2010), 501–506.
[14] Manfred Morari and Jay H. Lee. 1999. Model predictive control: Past, present and future. Computers and Chemical Engineering 23, 4-5 (1999), 667–682. https://doi.org/10.1016/S0098-1354(99)00301-9
[15] Jose R. Vázquez-Canteli, Jérôme Kämpf, Gregor Henze, and Zoltan Nagy. 2019. Reinforcement learning for building predictive control. Building and Environment 165, August (2019), 106351. https://doi.org/10.1016/j.buildenv.2019.106351
[16] Samuel Prívara, Jiří Cigler, Zdeněk Váňa, Frauke Oldewurtel, Carina Sager-Schep, and Eva Záčeková. 2013. Building modeling as a crucial part of future climate change. Intergovernmental Panel on Climate Change (2013), 674–738.
[17] Samuel Prívara, Jiří Cigler, Zdeněk Váňa, Frauke Oldewurtel, Carina Sager-Schep, and Eva Záčeková. 2013. Building modeling as a crucial part of building predictive control. Energy and Buildings 56 (2013), 8–22. https://doi.org/10.1016/j.enbuild.2012.10.024
[18] Pierluigi Siano. 2014. Demand response and smart grids - A survey. Renewable and Sustainable Energy Reviews 30 (2014), 461–478. https://doi.org/10.1016/j.rser.2013.10.022
[19] Jose R. Vazquez-Canteli, Jérôme Kämpf, Gregor Henze, and Zoltan Nagy. 2019. CityLearn v1.0: An OpenAI gym environment for demand response with deep reinforcement learning. BuildSys 2019 - Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (2019), 356–357. https://doi.org/10.1145/3363032.3364998
[20] Jose R. Vazquez-Canteli and Zoltan Nagy. 2019. Reinforcement learning for demand response: A review of algorithms and modeling techniques. Applied Energy 235 (2 2019), 1072–1089. https://doi.org/10.1016/j.apenergy.2018.11.002
[21] Zhe Wang and Tianzhen Hong. 2020. Reinforcement learning for building control: The opportunities and challenges. Applied Energy 269, April (2020), 115036. https://doi.org/10.1016/j.apenergy.2020.115036
[22] David Wölfle, Arun Vishwanath, and Hartmut Schmeck. 2020. A Guide for the Design of Benchmark Environments for Building Energy Optimization. BuildSys 2020 - Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (2020), 220–229. https://doi.org/10.1145/3408308.3427614
[23] Tianyu Zhang and Ornud Ardakian. 2020. COBS: Comprehensive Building Simulator. In Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation. ACM, New York, NY, USA, 314–315. https://doi.org/10.1145/3408308.3431119