POLITECNICO DI TORINO
Repository ISTITUZIONALE

CityLearn v2: energy-flexible, resilient, occupant-centric, and carbon-aware management of grid-interactive communities

Original
CityLearn v2: energy-flexible, resilient, occupant-centric, and carbon-aware management of grid-interactive communities / Nweye, K., Kaspar, K., Buscemi, G., Fonseca, T., Pinto, G., Ghose, D., Duddukuru, S., Pratapa, P., Li, H., Mohammadi, J., Lino Ferreira, L., Hong, T., Ouf, M., Capozzoli, A., Nagy, Z.. - In: JOURNAL OF BUILDING PERFORMANCE SIMULATION. - ISSN 1940-1493. - ELETTRONICO. - (2025). [10.1080/19401493.2024.2418813]

Availability:
This version is available at: 11583/2995379 since: 2024-12-16T17:03:35Z

Publisher:
Taylor and Francis Ltd.

Published
DOI:10.1080/19401493.2024.2418813

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

17 September 2026

CityLearn v2: Energy-flexible, resilient, occupant-centric, and carbon-aware management of grid-interactive communities

Kingsley Nweye[a], Kathryn Kaspar[b], Giacomo Buscemi[c], Tiago Fonseca[d], Giuseppe Pinto[e], Dipanjan Ghose[a], Satvik Duddukuru[a], Pavani Pratapa[a], Han Li[f], Javad Mohammadi[a], Luis Lino Ferreira[d], Tianzhen Hong[f], Mohamed Ouf[b], Alfonso Capozzoli[c] and Zoltan Nagy[a,*]

[a]*The University of Texas at Austin, 301 E. Dean Keeton St., ECJ 4.200, Austin, 78712, Texas, USA*
[b]*Concordia University, Montreal, Quebec, Canada*
[c]*Politecnico di Torino, Torino, Italy*
[d]*INESC-TEC/Polytechnic of Porto - School of Engineering, Porto, Portugal*
[e]*PassiveLogic, Amsterdam-Centrum, Noord-Holland, Nederland*
[f]*Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley, 94720, California, USA*

ARTICLE INFO

Keywords:
distributed energy resources
occupant-centric control
reinforcement learning
grid resilience
sustainability

ABSTRACT

As more distributed energy resources become part of the demand-side infrastructure, quantifying their energy flexibility on a community scale is crucial. CityLearn v1 provided an environment for benchmarking control algorithms. However, there is no standardized environment utilizing realistic building-stock datasets for distributed energy resource control benchmarking without co-simulation or third-party frameworks. CityLearn v2 extends CityLearn v1 by providing a stand-alone simulation environment that leverages the End-Use Load Profiles for the U.S. Building Stock dataset to create grid-interactive communities for resilient, multi-agent distributed energy resources and objective control with dynamic occupant feedback. While the v1 environment used pre-simulated building thermal loads, the v2 environment uses data-driven thermal dynamics and eliminates the need for co-simulation with building energy performance software. This work details the v2 environment and provides application examples that use reinforcement learning control to manage battery energy storage system, vehicle-to-grid control, and thermal comfort during heat pump power modulation.

Acronyms

ADR automated demand response.

AI artificial intelligence.

BAU business-as-usual.

BEPS building energy performance simulation.

BES building energy system.

BESS battery energy storage system.

CAIDI Customer Average Interruption Duration Index.

COP coefficient of performance.

DER distributed energy resource.

DHW domestic hot water.

DoD depth-of-discharge.

DOE Department of Energy.

DR demand response.

ERCOT Electric Reliability Council of Texas.

ESS energy storage system.

EUI energy use intensity.

EULP End-Use Load Profiles.

EV electric vehicle.

G2V grid-to-vehicle.

GEB grid-interactive efficient building.

GHG greenhouse gas.

HVAC heating ventilation and air conditioning.

IEQ indoor environmental quality.

KPI key performance indicator.

LoD level of detail.

*Corresponding author
nweye@utexas.edu (K. Nweye); kathryn.kaspar@mail.concordia.ca (K. Kaspar); giacomo.buscemi@polito.it (G. Buscemi); calof@isep.ipp.pt (T. Fonseca); giuseppe-pinto@polito.it (G. Pinto); dipanjan02@utexas.edu (D. Ghose); satvik.duddukuru@utexas.edu (S. Duddukuru); pavani1404@utexas.edu (P. Pratapa); hanli@lbl.gov (H. Li); javadm@utexas.edu (J. Mohammadi); llf@isep.ipp.pt (L.L. Ferreira); thong@lbl.gov (T. Hong); mohamed.ouf@concordia.ca (M. Ouf); alfonso.capozzoli@polito.it (A. Capozzoli); nagy@utexas.edu (Z. Nagy)
ORCID(s): 0000-0003-1239-5540 (K. Nweye); 0000-0002-9781-534X (K. Kaspar); 0000-0002-5592-3107 (T. Fonseca); 0000-0002-2967-0812 (D. Ghose); 0000-0002-5976-8853 (L.L. Ferreira); 0000-0003-1886-9137 (T. Hong); 0000-0002-0083-4983 (A. Capozzoli); 0000-0002-6014-3228 (Z. Nagy)

LSTM long short-term memory.

MAPE mean absolute percentage error.

MPC model predictive control.

OCC occupant-centric control.

PV photovoltaic.

QOI quantities of interest.

RBC rule-based control.

RC resistance-capacitance.

RES renewable energy source.

RL reinforcement learning.

RLC reinforcement learning control.

RMSE root mean square error.

SAC soft actor-critic.

SAIFI System Average Interruption Frequency Index.

SoC state-of-charge.

TES thermal energy storage.

ToU time-of-use.

U.S. United States.

V2G vehicle-to-grid.

WWR window-to-wall ratio.

ZNE zero-net energy.

1. Introduction

The electricity grid is undergoing system-wide changes due to the adoption of renewable energy sources (RESs), and electrification of buildings as well as transport systems, with the goal of reducing the carbon footprint from power generation and consumption [90]. However, the intermittency of RESs introduces additional challenges of grid instability due to the mismatch between electricity generation and demand [88] and thus, risks reducing the economic value of such sources [32]. At the urban scale, this intermittency causes periods of sudden and rapid increase in demand for fossil-fueled generation that threaten grid resiliency. The 'duck curve' is an example of such a period where depleted solar generation resulting from loss of daylight requires a steep ramp up in generation by fossil-fueled power plants to meet demand and is exacerbated by high photovoltaic (PV) penetration [15]. In extreme cases where power plants are unable to ramp up at the rate needed to meet demand, rotating blackouts are imposed to prevent grid imbalance.

1.1. Energy flexibility in grid-interactive efficient buildings

The Department of Energy (DOE), introduced the grid-interactive efficient building (GEB) initiative to promote the integration of distributed energy resources (DERs) such as PV systems, electric vehicles (EVs), active energy storage systems e.g. battery energy storage systems (BESSs), thermal energy storages (TESs), and passive energy storage systems e.g. thermal mass (walls) in buildings that can provide the grid with energy flexibility. To this end, a GEB is defined by Neukomm et al. as an energy-efficient building that uses smart technologies and on-site DERs to provide demand flexibility [also referred to as energy flexibility] while co-optimizing for energy cost, grid services, and occupant needs and preferences, in a continuous and integrated way. This differs from a building's participation in traditional demand response events that are unidirectional, reactive to specific events of high generation costs and peak, or incentivized consumption increase or decrease, and dependent on customer participation [87]. Temporary load shedding during on-peak periods, load shifting to off-peak periods, load modulation, and dispatch of generated renewable power for on-site consumption or grid export are several ways in which buildings can use these DERs to activate their flexibility and reshape their load profile to form a community of grid-interactive buildings [65].

As more DERs become part of the demand-side infrastructure, it is important to quantify the energy flexibility they provide on the urban scale as well as identify best control strategies to accelerate the design and adoption of energy flexibility programs. Particularly, understanding the impact of geographic, climatic, and occupant behavioral differences on the effectiveness of DERs in providing flexibility could shape future building design choices and provide a reference for policymakers. In their review on residential building energy flexibility quantification, Li et al. found four key performance indicator (KPI) classifications including energy, power, cost, duration, emission, and comfort [53]. Studies utilize KPIs from one or more classifications to define the flexibility provided by a proposed control solution, and the choice of KPI may be stakeholder-driven e.g. comfort and cost related KPIs are relevant to a building's occupant while a utility provider will be more concerned about energy and power related KPIs. Energy flexibility KPIs range from low complexity metrics e.g., peak power reduction [1] and relative CO_2 emissions reduction [94] to high complexity metrics such as flexibility indicator [102] and price responsiveness [46]. It is noteworthy that some of the frequent energy flexibility KPI found in the literature e.g. flexibility factor [49], energy efficiency of demand response action [7], and efficiency of automated demand response (ADR) [82] need a defined event of high peak or energy price in their calculation thus, are not well-suited for GEBs as a GEB's operation, by definition, is not event-driven but continuous. Furthermore, energy flexibility KPIs are classified as baseline-dependent or baseline-agnostic, where the former refers to metrics that are relative quantities and

require a reference performance profile from a business-as-usual (BAU) scenario but the latter is calculated as an absolute metric [52].

Also, to carry out an urban scale energy flexibility analysis, an inventory of the current building stock is crucial. The End-Use Load Profiles (EULP) for the United States (U.S.) Building Stock dataset [99], generated with ResStock [98] and Comstock [40] engines, uses physics-based energy models to provide over 900,000 synthetic building models that are calibrated with real-world data and represent the residential and commercial building stock in the U.S. The buildings differ in end-use fuel source, occupant behavior that affect energy profiles, archetype, construction material, climatic region and other distinguishable building characteristics. With the availability of such a large and diverse energy model repository, residential, commercial, and mixed-use grid-interactive communities to study the impact of DERs and energy flexibility programs are made possible.

It is, however, challenging to coordinate multiple DERs in a single building or multiple buildings while ensuring efficient and flexible operation that does not increase the risk of occupant discomfort. These DERs could also cause a deviation from the BAU load profile and could introduce new peaks [51, 72]. Also, EV loads, if not properly managed, have the potential to destabilize the grid with up to 20% increase in peak electricity demand [9] and 35% increase in electricity consumption [29].

Aside the risk of power outages caused by supply-demand mismatch, wildfires [103] and extreme weather such as heat waves [26], and winter storms [14] also exacerbate the likelihood of power outages leading to unserved energy and occupant discomfort. Thus, a control solution for DERs must be resilient during periods of curtailed supply or extreme cases of blackouts to mitigate the disaster vulnerability of occupants. To this end, we formally, define resiliency to refer to how a building performs during extreme weather or system failure that lead to power outages [38] and limit the scope to only thermal and electric energy aspects.

Furthermore, the stochasticity of occupant behavior may result in different energy profiles, thus indicates the influence that appliance usage, thermostat setpoints, and other occupant-driven end-uses have on energy consumption [2]. However, typical thermal comfort models such as Fanger's Predicted Mean Vote-Predicted Percentage Dissatisfied model [25] do not capture occupant preferences and behaviours that are observed in residential buildings. Smart thermostat data, such as ecobee's Donate Your Data dataset [58] can be leveraged to develop more realistic representations of both thermostat setpoint profiles and thermostat override behavior by occupants in the residential setting. In this way, thermostat setpoints can be automatically controlled for the implementation of demand response (DR) or energy flexibility, while occupants can choose to override the setpoint changes due to discomfort, thus balancing both energy efficiency and thermal comfort.

Compared to simpler rule-based control (RBC) [57], advanced control algorithms such as model predictive control (MPC) [20] and reinforcement learning control (RLC) [63] could provide solutions to these control challenges by adapting their control policy to unique building characteristics, occupant behaviors, and pricing or demand response signals from the power grid, while cooperating towards improving multi-objective KPIs. When improved, these KPI such as minimizing the peak load in a building district and maintaining building-specific comfort needs, ensure energy efficiency and high indoor environmental quality (IEQ) in buildings.

1.2. Research gaps

A research gap in the adoption of advanced control for DER control is the ability to benchmark algorithm performance [97]. Control benchmarking requires a simulation environment with two main components: (1) a building thermal dynamics model and (2) a control interface. The thermal dynamics model provides an output of indoor dry-bulb temperature given external weather boundary conditions, internal heat gains and losses, heating ventilation and air conditioning (HVAC) control actions as input. There are three thermal dynamics modeling techniques including physics-based modeling (white-box), reduced-order modeling (grey-box), and data-driven modeling (black-box) [4]. White-box modeling is deterministic and based on thermodynamics and heat transfer equations however, requires extensive domain knowledge on these topics to develop an accurate model. Grey-box modeling simplifies the heat transfer interactions in the building by reducing them to a resistor-capacitor model similar to an electric circuit, but model-fitting and determining the required circuit's complexity can prove difficult. Black-box modeling uses machine learning algorithms to predict the relationship between temperature and observations in the building thus overcomes the domain knowledge barrier. However, the models are difficult to explain and their applications are only valid for predictions with similarities to what was seen during training.

The control interface on the other hand, provides a standardized protocol of exchanging control actions and sensor observations between a control algorithm and building model. Examples of such control interfaces are OpenAI Gym [13] and Farama Foundation Gymanasium [89] and are compatible with standardized control algorithm libraries that provide varieties of ready-to-use algorithms e.g. Stable-Baselines3 [81] and RLlib [55].

A building emulator then provides an environment with DER abstraction and thermal dynamics in the form of a white-box, grey-box, or black box model, which enables the designer to focus on the control implementation and benchmarking within a defined control interface. Emulators such as Energym [86] and BOPTEST-Gym [5] are based on the OpenAI Gym interface. They provide high-fidelity thermal dynamics models through physics-based co-simulation engines including EnergyPlus and the Modelica-based Buildings Library. However, their dependency on such

engines as well as required extensive domain knowledge and input data from users raise their entry level. Moreover, they are designed for limited DER and building scenarios, do not consider a variety of DER loads e.g., EVs, do not model grid-resiliency nor occupant behavior, and do not allow for district-level control nor energy flexibility KPI evaluation.

Some of the limitations in BOPTEST-Gym and Energym have been addressed in other frameworks. The DOPTEST framework [6] that builds upon BOPTEST-Gym to support district-level and multi-agent control but like BOPTEST-Gym, requires compute-intensive co-simulation and is limited to building and district environments provided within the framework. OCHRE [11] is a Python-based emulator for both single and multi building control that makes use of an RC thermal dynamics model with user-configurable envelope parameters and DERs. Its DER models are defined similarly to EnergyPlus DERs thus, provides a blend of high fidelity modeling and custom buildings without the need for co-simulation with EnergyPlus. However, its interface does not follow any standardized control environment protocol. Thus, there is a need for an emulator that provides a thermal dynamics model and standard control interface with a low entry barrier for users to benchmark DER control algorithms.

1.3. CityLearn v1

CityLearn is an open-source Gymnasium environment for the easy implementation and benchmarking of RBC, RLC and MPC algorithms for DERs in a DR or grid-interactive community setting. CityLearn is used to reshape the aggregated electricity load profile by controlling DERs in a district of diverse buildings, and allows for multi-agent control and district-level KPIs evaluation. With CityLearn v1 introduced in [95], some of its applications summarized in Table 1 include control benchmarking problems comparing multi-agent and single-agent RLC [93], RBC and RLC [68], MPC and RLC [104], and MPC forecasting models [48]. Other applications include transfer learning [72], meta-learning [105], and voltage regulation [79]..

1.4. Contributions

The CityLearn v2 release that we present here provides a framework that incorporates thermal dynamics modeling in [80] that allows for heat pump capacity/power control to provide partial load satisfaction, precooling and preheating services in buildings. Additionally, this new release integrates an EV module to provide grid-to-vehicle (G2V) and vehicle-to-grid (V2G) control, as well as occupant modeling for thermostat setpoint override to simulate realistic occupant-centric control (OCC) and thermal comfort assessment. CityLearn v2 also provides an integration with the EULP for the U.S. Building Stock dataset based on the work by Nweye et al. [67] as well as the functionality to assess the resiliency of control algorithms during power outage events through a stochastic power outage model based on distribution system reliability metrics [41] such that it is generalizable by providing custom or location-specific metrics. For compatibility with standard control interfaces, the v2 environment follows the latest Gymansium control protocol and provides wrappers for interfacing with Stable-Baselines3 and RLlib single and multi-agent algorithms. Thus, we summarize the contributions of our work as providing:

1. a simulation environment that leverages realistic building-stock datasets for resilient multi-agent, DER and objective control with dynamic occupant feedback in grid-interactive communities;
2. an all-in-one environment without the need for co-simulation that is designed according to the standardized Gymnasium interface; and
3. extensible interface for standardized control algorithm libraries.

The remainder of our paper is organized as follows: Section 2 explains the environment and its internal models in detail, while Section 3 describes the environment's control interface and configurations. We provide example applications of the CityLearn v2 environment in Section 4 and discuss the implications of the new environment in building controls research as well as its limitations in Section 5. Finally, we conclude in Section 6. An Appendix section provides supporting environment documentation.

2. Environment

CityLearn[1] models a district of buildings with similar or different loads, electric devices, energy storage systems (ESSs) and electricity sources that satisfy the loads as shown in Figure 1. The environment can consist of only one building or a district of buildings with no upper limit on the number of buildings. The electric devices and ESSs are one of many DERs described in Appendix A.1. We further describe how these DERs are used in a building to satisfy loads, store energy, or provide electricity supply in Section 2.1. The power outage model is described in Section 2.2, while Sections 2.3 and 2.4 define the in-built KPIs and datasets for control benchmarking. We end this section in Section 2.5 by describing a framework that leverages the EULP dataset to create virtual grid-interactive communities for CityLearn.

2.1. Building model

The building model is illustrated in Figure 1 where we show the transfer of actions from the control agent and its reception of measurable observations that describe the building's states. This building-control interface is explained in Section 3.

A building is a combination of electricity sources that power controllable DERs, including electric devices and ESSs used to satisfy thermal and electrical loads as well as provide the grid with energy flexibility (Sections 2.1.1 to 2.1.4). The building is modeled as a single thermal zone where space thermal loads affect its indoor dry-bulb temperature. We use a data-driven thermal dynamics model to quantify the effect of the thermal load on temperature

[1] https://www.citylearn.net

Table 1: CityLearn v1 applications

Ref.	Description	Control	Buildings	Duration	DERs	KPIs	Key findings
[93]	Benchmarking	RBC, RLC	9 commercial and multi-family	5 years	TES, PV	consumption, peak, ramping, load factor	Multi-agent RL performs superior to RBC and single-agent RLC.
[68]	Benchmarking	RBC, RLC	9 commercial and multi-family	5 years	TES, BESS, PV	consumption, peak, ramping, load factor	Offline learning from good RBC improves RL performance.
[104]	Benchmarking	MPC, RLC	5 single-family	1 year	BESS, PV	cost, emissions	MPC more explainable; RLC more computationally tractable for multi-system control.
[48]	Benchmarking	MPC	15 commercial	10 years	BESS, PV	consumption, cost, emissions	Diminishing return as training data approaches 2 years; change-point analysis improves accuracy and data efficiency.
[72]	Transfer learning	RBC, RLC	17 single-family	1 year	BESS, PV	consumption, peak, ramping, load factor, cost, emissions	Transferred RLC better than RBC and comparable to non-transferred RLC, while reducing training cost.
[105]	Meta-learning	RBC, MPC, RLC	9 commercial and multi-family	1 year	TES, PV	consumption, peak, ramping, load factor	Meta-RLC converges fasterand stably while and performing better in unseen buildings compared to alternatives including randomly initialized RLC.
[79]	Voltage regulation	RBC, RLC	192 commercial and multi-family	1 year	TES, BESS, PV	voltage	RLC reduces instances of undervoltage and overvoltage.

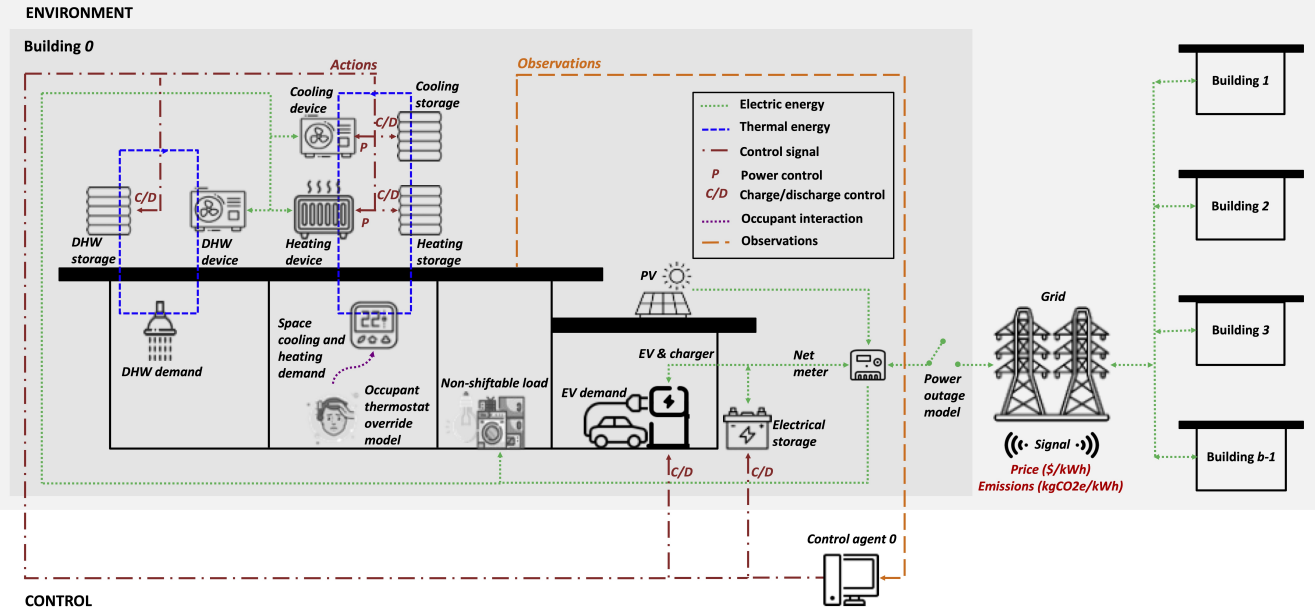


Figure 1: CityLearn building model including electricity sources that power controllable DERs including electric devices and ESSs, used to satisfy thermal and electrical loads as well as provide the grid with energy flexibility. A distinction is made between environment and control aspects of a building to show the transfer of actions from the control agent and reception of measurable observations by the control agent that quantifies the building's states.

(Section 2.1.5) and an occupant model that has the ability to override the temperature setpoint is described in Section 2.1.6.

2.1.1. Loads

There are five loads in a building including space cooling, space heating, domestic hot water (DHW) heating, electric equipment, and EV loads. The space cooling and heating loads refer to the energy needed to maintain the indoor dry-bulb temperature at its setpoint. We make no distinction between cooling and heating setpoints and CityLearn does not allow for simultaneous space cooling and heating in the same time step for a building. The DHW heating load is the total heating energy needed to satisfy hot water end-uses such as shower, bathroom, and kitchen sinks, and other water end uses requiring water heating that are not space heating-related. Electric equipment refer to non-shiftable plug loads such as lighting, entertainment and kitchen appliances. The EV load is the energy required to charge an EV to a scheduled departure state-of-charge (SoC). Within CityLearn, the space cooling, space heating, DHW heating, electric equipment, and EV loads are called `cooling_demand`, `heating_demand`, `dhw_demand`, `non_shiftable_load`, and `ev_demand`.

Not all loads need to exist in a building e.g., a building situated in a heating dominant climate may not have cooling loads year-round. Also, any one or all of these loads are either known a priori from building energy performance simulation (BEPS) [93] or real-world measurement [72]. In these instances, the ideal load must be satisfied. Alternatively, they are controlled loads and are inferred at runtime e.g., heat pump power control driving space cooling or heating loads. To satisfy these loads in either the ideal or control-action case, we make use of HVAC systems directly (Section 2.1.2), or ESSs (Section 2.1.3) through load shifting, where these systems are powered by the electricity sources described in Section 2.1.4.

2.1.2. Electric devices

The `cooling_device`, `heating_device`, and `dhw_device` are HVAC electric device objects in CityLearn that are used to satisfy the space cooling, space heating and DHW heating loads respectively. The `cooling_device` is a heat pump while the `heating_device` and `dhw_device` are either heat pump or electric heater type. The CityLearn version at the time of writing allows for `cooling_device` and `heating_device` control subsequently, driving `cooling_demand` and `heating_demand`.

These HVAC systems are powered directly by one or more of the electricity sources in Section 2.1.4 and if available, may be used to charge TES systems in the building.

The `electric_vehicle_charger` is used to satisfy EV load or acts an intermediate link between the EV and grid for V2G applications.

2.1.3. Energy storage systems

There are up to five optional and controlled ESSs in a building including `cooling_storage`, `heating_storage`, `dhw_storage`, `electrical_storage`, and `electric_vehicle` ESS objects. The `cooling_storage`, `heating_storage`, and `dhw_storage` are TES DER type and provide space cooling, space heating and DHW heating load shifting flexibility respectively. They are charged by the HVAC device used to meet the thermal load which they service e.g. is the `cooling_device` charging the `cooling_storage`. However, the building's thermal load is

first satisfied and what is left of the HVAC device's nominal power is used to charge the TES. When discharging, the energy from the TES is first used to meet the building's thermal load before the HVAC device.

The `electrical_storage` is a BESS DER type that powers any of the aforementioned electric devices when in discharge mode or is powered by one or more of the electricity sources in Section 2.1.4. If the `electrical_storage` discharges more energy than needed to meet the building loads, the excess is sent to the grid as part of the building's net export.

The `electric_vehicle` is an EV DER type and performs similar function as the `electrical_storage`, however, the EV is available on a schedule defined by its arrival and departure times. The EV can be used in three modes: G2V, V2G, and no control (i.e., where the EV acts as a load without any possible control over its charging) [27].

2.1.4. Electricity sources

The electric devices are primarily powered by the electric grid. CityLearn at the time of writing, does not have a grid model so the power a building is able to draw from the grid at a given time step is unconstrained, except in the case of a power outage (Section 2.2). Optionally, a building may have a PV system (`pv`) that provides self-generation as a first source of electricity before the grid. Optional `electrical_storage` and `electric_vehicle` are charged by the grid and `pv` but also augment the `pv` and grid when in discharge mode to supply the building with electricity. Excess self-generation, `electrical_storage` and `electric_vehicle` discharge are sent to the grid as apart of the building's net export.

$$E_t^{\text{building, net}} = E_t^{\text{cooling_device}} + E_t^{\text{heating_device}} + E_t^{\text{dhw_device}} + E_t^{\text{non_shiftable_load}} + E_t^{\text{electrical_storage}} + E_t^{\text{electric_vehicle_charger}} + E_t^{\text{pv}} \quad (1)$$

Given the aforementioned electric devices, ESSs and electricity sources, the net electricity consumption of the building (Equation (1)) is thus, the sum of the positive electricity consumption by the `cooling_device` ($E_t^{\text{cooling_device}}$), `heating_device` ($E_t^{\text{heating_device}}$), `dhw_device` ($E_t^{\text{dhw_device}}$), and `non_shiftable_load` ($E_t^{\text{non_shiftable_load}}$), mixed polarity electricity consumption of the `electrical_storage` ($E_t^{\text{electrical_storage}}$), and `electric_vehicle_charger` ($E_t^{\text{electric_vehicle_charger}}$), and negative polarity electricity generation by the `pv` (E_t^{pv}). $E_t^{\text{cooling_device}}$, $E_t^{\text{heating_device}}$, and $E_t^{\text{dhw_device}}$ are the electricity to be used to satisfy the `cooling_demand`, `heating_demand`, and `dhw_demand` after accounting for energy discharge from the `cooling_storage`, `heating_storage`, and `dhw_storage`. Therefore, a case where there the control action for these three TESs is adequate to fully satisfy the loads, $E_t^{\text{cooling_device}}$, $E_t^{\text{heating_device}}$, and $E_t^{\text{dhw_device}}$ will equal zero.

The district-level net electricity consumption is the sum of $E_t^{\text{building, net}}$ for all buildings as defined in Equation (2).

$$E_t^{\text{district, net}} = \sum_{i}^{b-1} E_t^{\text{building } i\text{,net}} \quad (2)$$

2.1.5. Thermal dynamics model

The building's thermal dynamics is modeled as a long short-term memory (LSTM) network that predicts the indoor dry-bulb temperature, T_t^{in}, given a lookback, l, of previous observations in the building [80]. The model function is defined in Equation (3) where its independent variables include indoor dry-bulb temperature (T^{in}), outdoor dry-bulb temperature (T^{out}), cooling or heating load depending on `hvac_mode` defined in Appendix A.3 ($Q^{\text{cooling | heating}}$), direct solar irradiance ($Q^{\text{solar, direct}}$), diffuse solar irradiance ($Q^{\text{solar, diffuse}}$), occupant count ($U$), month ($m$), day-of-week ($d$), and hour ($h$). While the T^{in} input to the model is the previous l time steps, other input are their observations between the current time step, t, and previous $l-1$ time steps. Appendix A.2 explains the methodology for generating the model fitting data, describes how these data are used in training and testing the LSTM network, and shows the testing results for three datasets.

$$T_t^{\text{in}} = f\left(T_{t-1\ldots t-l}^{\text{in}}, \left(T^{\text{out}}, Q^{\text{cooling | heating}}, Q^{\text{solar, direct}}, Q^{\text{solar, diffuse}}, U, m, d, h\right)_{t\ldots t-(l-1)}\right) \quad (3)$$

2.1.6. Occupant-thermostat override model

Local or regional thermostat data including indoor air temperature, date/time, HVAC system state (e.g. heat, cool) and thermostat events (e.g. override occurrence) can be used to develop occupant-thermostat override models. Smart thermostats are programmable and allow for the user to set various program settings during specified hours and duration, such as 'Home', 'Away', and 'Sleep'. When a user decides to change from the set program, an override occurs, and generally this behavior can be interpreted as discomfort-driven. However, not all occupants have the same thermostat setpoint preferences and subsequent override behavior. As such, to develop the occupant-thermostat override models, we firstly (a) cluster the thermostat setpoint profiles to identify distinct thermostat setpoint preferences in the region; (b) develop a model to predict the probability of a setpoint override for each occupant type; (c) estimate the magnitude of the thermostat setpoint override.

Firstly, we develop models for three periods (summer, winter, shoulders) separately, as HVAC system state and occupant override behavior are different during each period. Next, we query the setpoint data for one period (e.g. winter) and extract discomfort-driven behavior from the dataset. In the case study presented in [56], seven out of 20 occupants showed no correlation between the occurrence of a setpoint increase and any environmental variable, e.g. indoor or outdoor air temperature, thus showing the difficulties in interpreting and using thermostat data. In a winter heating

scenario, we expect an inverse relationship between the indoor air temperature and a setpoint override (the data was organized such that '0' represented no override and '1' represented an override). We thus firstly used Pearson Correlation to extract all homes/occupant data with a correlation less than -0.20 (some correlation) between the indoor air temperature and an increase to the temperature setpoint, thus focusing the modeling on this subset of the data.

Using the remaining data, we then used k-means correlation to identify clusters based on the average indoor air temperature during a setpoint override, separating the overrides by whether they increased or decreased the setpoint. The Silhouette score [85] was used to determine the number of clusters, and we visualized the average daily setpoint profile as well as a histogram of the indoor air tempterature to ensure unique clustering. It would also be possible to do k-shape clustering on the average daily setpoint profiles as well to create the unique clusters, as was done in [76].

We next model the probability of a setpoint override for each occupant type, again separating setpoint increase and setpoint decrease behavior for the overrides. If occupancy information is available, extract only the time steps in which the home is occupied for this modeling; otherwise, assume an acceptable occupancy rate for the dataset. We then replicate the methodology used in [56, 33] and use discrete-time Markov logistic regression models to predict the probability of a thermostat setpoint override, an industry standard as referenced in [19]. Equation 4 shows the form of the logistic regression equation, with the indoor air temperature, T_{in}, representing the input variable and p representing the probability of a setpoint increase or decrease (override = 1, in either case, but the scenarios are evaluated separately). Coefficients a and b are determined in the training, and the fit of the curve is evaluated using the p-value.

$$p(override = 1) = \frac{1}{1 + e^{-(a+bT_{\text{in}})}} \tag{4}$$

Once the probability of an override is modeled, the next step is determining by how much the setpoint is changed in the event that an override occurs. Depending on the data available, it is possible to randomly sample from the distribution of the setpoint overrides to estimate the change in setpoint during the override. We further developed random forest models to classify the change to the setpoint during an override as either small (less than 0.5 °C) or large (greater than 0.5 °C).

2.2. Power outage model

A power outage is defined as a time series, O, of binary signals where $O_t = 0$ indicates no outage at time step, t, and $O_t = 1$ indicates an outage. A sequence of consecutive values of $O_t = 1$ is regarded as a power outage event and overlapping events are merged into one event. These signals are either arbitrarily defined or generated using some stochastic model.

We provide a stochastic power outage model based on distribution system reliability metrics [41] including the number of non-momentary electric interruptions, per year, an average customer experiences (System Average Interruption Frequency Index (SAIFI)), and the average number of minutes it takes to restore non-momentary electric interruptions (Customer Average Interruption Duration Index (CAIDI)). These metrics are reported annually and location-specific [91] allowing for model adaptation to real-world locations and scenarios. A given day of the year having an outage event or not is determined by randomly sampling from the Bernoulli distribution in Equation (5), where the probability of an event, p, is the ratio of SAIFI to number of days in a non-leap year (365). The start of an event is randomly sampled from a uniform distribution, where Equation (6) is an example of such distribution for an hourly time step. Finally, the duration of each power outage event is sampled from an exponential distribution with rate, λ, set to CAIDI (Equation (7)). Sampling from these distributions for a fixed SAIFI and CAIDI is controlled by the model's random seed.

$$\text{Outage event} \sim \text{Bernoulli}\left(p = \frac{\text{SAIFI}}{365}\right) \tag{5}$$

$$\text{Outage start hour} \sim U(0, 23) \tag{6}$$

$$\text{Outage duration (minutes)} \sim Exp(\lambda = \text{CAIDI}) \tag{7}$$

$$E_t^{\text{available}} = \begin{cases} E_t^{\text{pv}} \\ -\left(E_t^{\text{cooling_device}} + E_t^{\text{heating_device}}\right. \\ +E_t^{\text{dhw_device}} + E_t^{\text{non_shiftable_load}} \\ +E_t^{\text{electrical_storage}} \\ \left.+E_t^{\text{electric_vehicle_charger}}\right) & \text{if } O_t > 0 \\ \infty & \text{otherwise} \end{cases} \tag{8}$$

Equation (8) defines the available electricity supply during a power outage event and normal grid operation. In an outage event, the grid is unable to provide buildings with electricity and buildings can only make use of the flexibility provided by ESSs and PV system after subtracting electricity use by (1) the `cooling_device`, `heating_device`, and `dhw_device` to meet thermal loads and charge TESs, (2) `non_shiftable_load`, as well as (3) `electrical_storage` and `electric_vehicle_charger` if in charging mode. Whereas, during normal operation, there is unlimited supply from the grid.

2.3. Energy flexibility key performance indicators

Table A.5 summarizes the energy flexibility KPIs within CityLearn that are used for thermal-comfort, energy cost, environmental, energy, and resilience evaluation. These KPIs

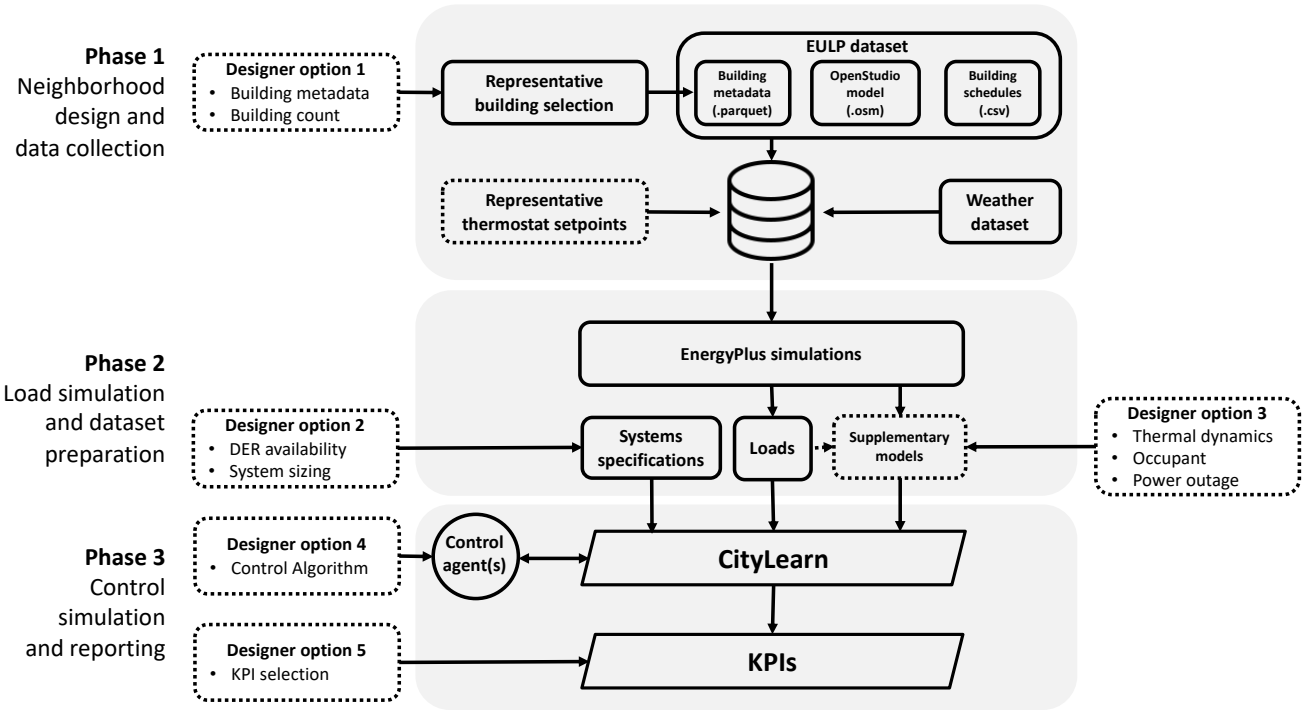


Figure 2: Framework for designing virtual grid-interactive communities in CityLearn from the End-Use Load Profiles for the United States Building Stock dataset (adapted from [67]).

are calculated at either the building or district level by default and KPIs defined at the building level are evaluated at the district level by calculating the building-level average. The KPIs are defined such that the control objective is to minimize their value.

2.4. Datasets

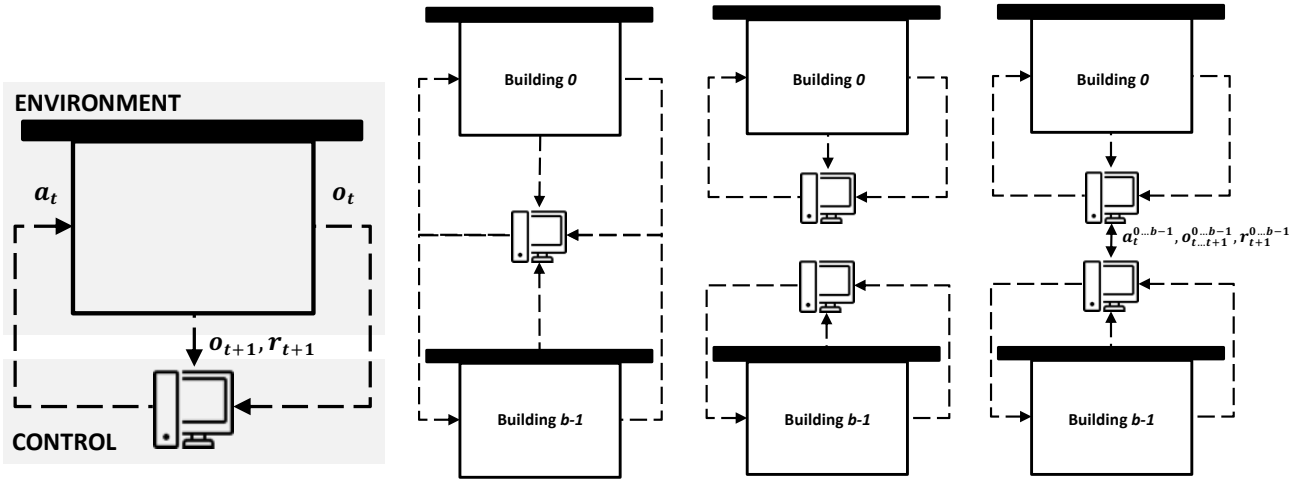
Table A.6 is a summary of the in-built datasets in CityLearn that were designed for a distribution of control problems including those specific to The CityLearn Challenge [96, 64, 71]. Each dataset is a unique environment configuration where environments differ by their location and climate zone that influence their weather file and building characteristics, time series date range, building count, model availability (thermal, occupant and power outage models) that affect the complexity of the control problem, control signal availability including electricity pricing and carbon emissions time series, building loads, controllable ESSs, and PV system availability for self-generation.

2.5. Environment design workflow

Figure 2 depicts a framework for designing virtual grid-interactive communities in CityLearn by leveraging the EULP dataset as well as other data sources [67]. This framework is split into three phases: (1) neighborhood design and data collection, (2) load simulation and dataset preparation and (3) control simulation and reporting.

In the neighborhood design and data collection phase, the neighborhood designer pre-filters the EULP dataset for buildings whose metadata match certain criteria such as archetype, vintage, location, and HVAC equipment type and the designer may choose to define the building count in the neighborhood. Random selection or data-driven methods are used to select buildings up to the specified count. A data-driven approach is provided in [67] that makes use of cluster-frequency-based sampling from buildings clustered by metadata similarities including orientation, decade of construction, number of occupants, infiltration rate, ceiling, slab and wall insulation, energy use intensity (EUI) and window-to-wall ratio (WWR). The building metadata, occupancy and load schedules, weather data and energy models for the selected buildings are then stored in a central database for easy retrieval and manipulation. Optionally, the variance in building loads and indoor environmental conditions across buildings is improved by replacing the default thermostat setpoint schedules with real-world thermostat setpoints from the ecobee Donate Your Data dataset [58].

In the building load simulation and dataset preparation phase, the collected data are used to run BEPS simulations following the methodology for thermal dynamics modeling in Section 2.1.5. The LSTM thermal dynamics model is interchangeable with other thermal dynamics modeling methodology e.g. resistance-capacitance (RC) modeling. In addition to the data needed to train the thermal dynamics model other ideal loads data including DHW heating loads and plug loads are retrieved from the BEPS. The choice of occupant and power outage models are also decided in this second phase. Finally in the second phase, designer input for DER availability as well as their technical specifications are defined.



(a) Farama Foundation Gymnasium interface.

(b) Single-agent (left), independent multi-agent (middle), and coordinated multi-agent (right) control configurations.

Figure 3: Control summary in CityLearn.

The simulated loads, models and system specifications from the second phase are utilized to create a virtual representation of the intended neighborhood in the control simulation and reporting phase. The designer selects a control algorithm to manage the DERs. Post-simulation evaluation of control performance is achieved by the user-selected KPIs.

3. Control

Control in CityLearn is defined by its interface and configuration as shown in Figure 3. The CityLearn environment makes use of the Farama Foundation Gymnasium interface [89], built upon the OpenAI Gym interface [13], for standardized RLC environment design. Figure 3a illustrates this interface where there is an observation-action-reward exchange loop between the environment and control agent as the environment transitions from one time step to another. In the current time step, t, the control agent receives the environment's observations, o_t and prescribes actions a_t. The actions are applied to the environment to affect the observations at the next time step, o_{t+1}. o_{t+1} and a reward, r_{t+1} (from reward function, R) that quantifies the quality of a_t in optimizing the outcome of a control objective or KPI are returned to the control agent to teach it to learn a control policy, π. π maps actions to observations that maximize the cumulative reward over an episode i.e., the terminal state of the environment, after initialization ($t = 0$), beyond which there are new observations. Appendices A.3 and A.4 describe the observation and action spaces in CityLearn while Appendix A.5 outlines the internally defined reward functions in the environment.

Figure 3b shows three possible control configurations in CityLearn namely; single-agent, independent multi-agent, and coordinated multi-agent. In the single-agent configuration, there is a one-to-many relationship between the control agent and buildings where a centralized agent collects observations and prescribes actions for all DERs in the district and, receives a single reward value each time step to learn a generalized control policy. This is akin to an energy aggregator controlling flexible resources in a distributed manner. The independent multi-agent configuration has a one-to-one agent-building relationship thus, there are as many rewards as buildings each time step and a unique control policy is learned for each building. The coordinated multi-agent configuration is similar to the independent multi-agent configuration except that agents can share information to achieve cooperative objectives e.g. district peak reduction or competitive objectives e.g. price bidding in the energy flexibility market. Appendix A.6 discusses some of the in-built control agent algorithms in CityLearn that use one or more of these configurations as well as CityLearn's provisions for interfacing with third-party standardized control algorithm libraries.

We emphasize that CityLearn is not limited to RLC algorithms alone despite its Gymnasium interface as it works with other simple control theory algorithms e.g., RBC [68] as well as advanced control theory algorithms e.g., MPC [104] as shown in Figure 4. In RBC, the reward is not utilized while in MPC, the reward is akin to the control objective. Also, RBC policy is static and does not consider the observations in the following time step to update its rules.

4. Examples

This section provides example applications of the CityLearn v2 environment. Section 4.1 showcases 17 different control tasks of differing complexity, Section 4.2 applies the EULP dataset to create three representative neighborhoods for ESS control, Section 4.3 provides a commercial building district V2G application for a European EV dataset, and Section 4.4 demonstrates the use of the occupant thermostat

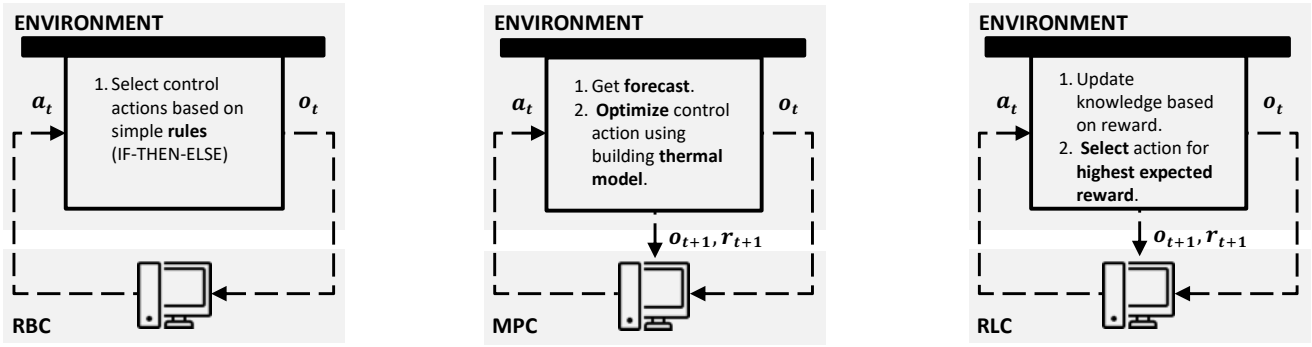


Figure 4: Integration of rule-based control (RBC), model predictive control (MPC), and reinforcement learning control (RLC) with CityLearn.

Table 2
Summary of CityLearn v2 control tasks in [70] describing the number of buildings in controlled district, KPI to minimize, control agent and controlled DERs.

ID	Buildings	KPI	Control	DHW TES	BESS	PV	Heat pump
x-b1_b2-x-x	2	-	Baseline				
x-b1_b2-x-pv	2	-	Baseline			✓	
rbc-b1-c-dhw	1	Cost	RBC	✓			
rbc-b1-e-dhw	1	Emissions	RBC	✓			
rbc-b1-c-bess_pv	1	Cost	RBC		✓	✓	
rbc-b1-e-bess_pv	1	Emissions	RBC		✓	✓	
rbc-b1-c-dhw_bess_pv	1	Cost	RBC	✓	✓	✓	
rbc-b1-e-dhw_bess_pv	1	Emissions	RBC	✓	✓	✓	
rbc-b1_b2-p-bess_pv	2	Peak	RBC		✓	✓	
rlc-b1-c-dhw	1	Cost	SAC	✓			
rlc-b1-e-dhw	1	Emissions	SAC	✓			
rlc-b1-c-bess_pv	1	Cost	SAC		✓	✓	
rlc-b1-e-bess_pv	1	Emissions	SAC		✓	✓	
rlc-b1-c-dhw_bess_pv	1	Cost	SAC	✓	✓	✓	
rlc-b1-e-dhw_bess_pv	1	Emissions	SAC	✓	✓	✓	
rlc-b1-d_o-hp	1	Discomfort & electricity consumption	SAC				✓
rlc-b1_b2-p-bess_pv	2	Peak	SAC		✓	✓	

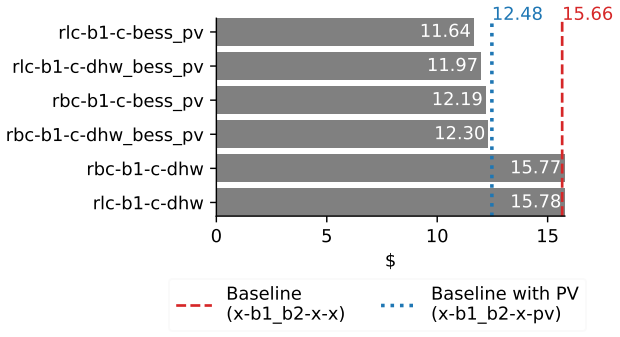
setpoint override model in conjunction with heat pump power modulation.

4.1. Control task distribution in CityLearn v2

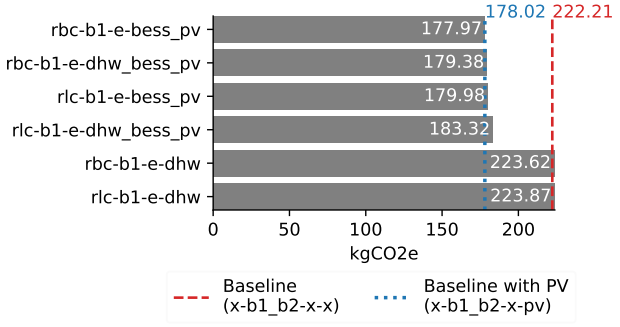
The work in [70] showcases 17 different control tasks (Table 2) of differing complexity that are solved in CityLearn v2 using a subset of The CityLearn Challenge 2023 datasets. Complexity refers to (1) the number of controllable DERs present in a building including DHW TES (`dhw_storage`), BESS (`electrical_storage`) when paired with a PV system and heat pump (`cooling_device`), (2) the simplicity of the control algorithm i.e., explainable RBC or adaptive but black-box RLC, (3) the control objective (reward function for RLC) designed for reduction in electricity consumption, energy cost, greenhouse gas (GHG) emission, discomfort, peak demand, or a combination of the aforementioned objectives, and (4) the size of the district i.e., number of buildings.

4.1.1. Cost or emission reduction

Figure 5 shows cost ($) and emissions ($kgCO_2e$) in a building from RBC or RLC of DHW TES, BESS-PV system or both when the control objective is energy cost (Figure 5a) or emission (Figure 5b) reduction. The results are compared to baseline configurations without PV (*x-b1_b2-x-x*) and with PV (*x-b1_b2-x-pv*) in a 17-day test period. The PV system advantage in terms of cost and emissions reduction is approximately 20.0%, and is reflective of the 21.5% zero-net energy (ZNE) that the PV system was sized for. The results show that RLC provides further cost reduction of up to 6.7% compared to RBC while, neither algorithm provides substantial reduction in emissions irrespective of what DERs are available. These result from using emissions as control signal highlighted the importance of simultaneously decarbonizing the supply-side as demand-side end uses are electrified or risk an adverse effect of increased emissions. The variance in carbon emissions were less than 1.0% and renewable energy sources in the Electric Reliability Council of Texas (ERCOT) grid for the time period was estimated at only 20.1% with 5.5% standard deviation [24] making it a weak control signal.



(a) Cost ($).



(b) Emissions ($kgCO_2e$).

Figure 5: Cost ($) and emissions ($kgCO_2e$) results in [70] from either RBC or RLC of DHW TES, BESS-PV system or both when the control objective is to reduce cost or emission. The dashed red line shows the cost for a baseline scenario and the dotted blue line shows the cost for a baseline scenario but with solar generation to augment electricity from the grid.

Figure 6 shows the RLC actions in the initial seven days of the test period for the *rlc-b1-c-dhw* and *rlc-b1-c-dhw_bess_pv* configurations that have a cost reduction objective. The trend in the ESS SoC as a consequence of the actions are shown as well as shaded regions to indicate the time-of-use (ToU) pricing. The SoC trend shows that the RLC agent learns to charge the ESSs during the off-peak period while discharging mainly during the on-peak period. Compared to the BESS, the DHW TES rarely discharges its full capacity. This observation is attributed to the infrequent occurrence of DHW loads where only 7% of the total building load is attributed to DHW load, as there is no DHW load during 330/407 control time steps.

4.1.2. Peak reduction

The district-level daily peak load is shown in Figure 7 for the baseline with PV generation (*x-b1_b2-x-pv*) and the two configurations with peak reduction objective but differing control algorithms. The RBC in *rbc-b1_b2-p-bess_pv* that has been fine-tuned to target energy discharge during peak periods provides a 2.6% advantage in average daily peak reduction over the *rlc-b1_b2-p-bess_pv* alternative that uses electricity consumption as a reward signal.

4.1.3. Discomfort and electricity consumption reduction

In the *rlc-b1-d_o-hp* simulation that modulates the `cooling_device` heat pump power, a modified version of the comfort reward (Equation (A.34)) is used where only cooling mode is considered. Exponents $a = 1$ and $b = 1$, and a penalty coefficient, m, is multiplied with the reward when there is over-cooling i.e., $T_t^{\text{in}} < T_t^{\text{spt}}$. Thus, comfort and energy efficiency can be learned by way of this penalty in the reward function for increased electricity consumption when over-cooling. Table 3 shows the effects on discomfort and electricity consumption from tuning m. The percent changes in KPIs by setting $m = 3$, $m = 6$, and $m = 12$ are compared to the baseline case, $m = 1$ (no penalty for over-cooling). By setting $m = 3$, consumption and discomfort are reduced by 3.0% by 50.0% decrease in over-cooling. In contrast, $m = 3$ increases under-cooling by only 20.0%. Larger increments in m show an increase in discomfort of almost 100.0% from under-cooling compared to the baseline but up to 12.0% decrease in consumption. However, the average temperature difference between the indoor dry-bulb temperature and setpoint for the selected m values, irrespective of over-cooling or under-cooling, is less than 1.5C.

4.2. Energy storage system control in representative single-family neighborhoods

In [67], the framework for designing virtual grid-interactive communities is applied to create three representative single-family neighborhoods in Alameda County, California (CA); Chittenden County, Vermont (VT); and Travis County, Texas (TX). For each location, six clusters of building metadata were found from which a frequency-based sampling is used to select up to 100 buildings. The CA, VT, and TX neighborhoods have 73, 43 and, 100 buildings respectively and each building is equipped with BESS-PV system and DHW TES in an independent multi-agent RLC architecture. The solar penalty reward, Equation (A.33), that encourages net-zero energy use by penalizing load satisfaction from the grid when there is stored energy in ESSs as well as penalizing net export when ESSs are not fully charged is used to train the independent agents.

Figure 8 compares the district-level net electricity consumption for scenarios with and without storage for a winter week in CA and VT neighborhoods, and a summer week in the TX neighborhood. The shaded regions show the average $\text{SoC}^{\text{BESS}} \pm$ one standard deviation. In CA and TX, the control agents learn to take advantage of the daytime solar generation to charge the BESSs and release the stored energy in the evening thus, reducing the daily peaks by 8.0% and 42.0% on average in either neighborhood. The VT profiles for the cases with and without storage are similar as the agents do not learn the load shifting task as in the case of the other two neighborhoods. The SoC^{BESS} distribution in VT shows irregular charge-discharge cycles and underutilized battery capacity for load shifting. This disparity in the VT results indicates the need for location-specific control strategies.

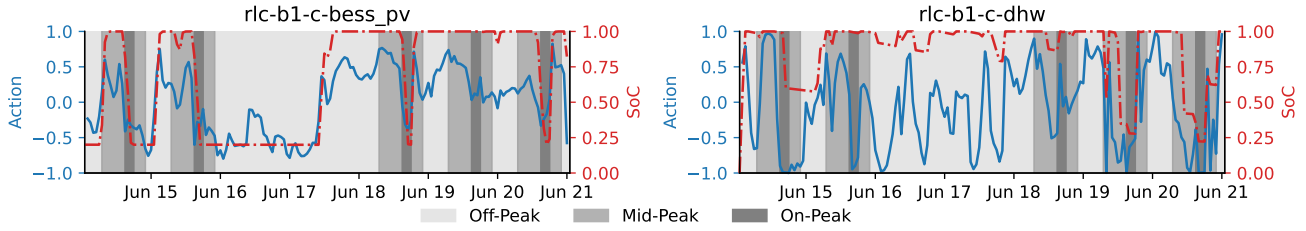


Figure 6: RLC action for BESS (left) and DHW TES (right), and consequent SoC trend in the initial seven days of the two-week evaluation period when the control objective is to minimize cost in [70].

Table 3
Effect of coefficient, m on discomfort, average discomfort when over-cooling and under-cooling , and electricity consumption objectives when used in heat pump control simulation, *rlc-b1-d_o-hp* ([70]). The percent change in KPI, by varying m is compared to the baseline case, $m = 1$, where there is no penalty for increased consumption from over-cooling. Improvement in a KPI is highlighted in blue while deterioration is highlighted in red.

m	Discomfort (°Ch)	Over-cool (°C)	Under-cool (°C)	Electricity consumption (kWh)
1	201.8	0.4 ± 0.5	0.5 ± 0.5	384.8
3	195.7 (-3.0%)	0.2 ± 0.2 (-50.0%)	0.6 ± 0.5 (20.0%)	373.0 (-3.0%)
6	315.9 (56.5%)	0.3 ± 0.2 (-25.0%)	0.9 ± 0.5 (80.0%)	352.5 (-8.3%)
12	399.6 (98.0%)	0.1 ± 0.4 (-75.0%)	1.0 ± 0.6 (100.0%)	338.1 (-12.0%)

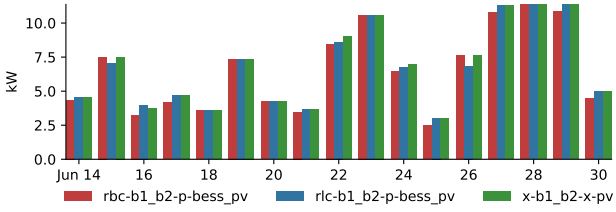


Figure 7: District-level daily peak load for two-building district where each building has BESS-PV system in [70].

4.3. Vehicle-to-grid control

The work in [27] demonstrates a V2G scenario in CityLearn v2 for a nine-building residential and commercial building district including a medium-sized office, fast-food restaurant, standalone retail store, strip mall, and five medium-scale multifamily residences. The control signal is the real-time electricity pricing from the Iberian wholesale energy market (OMIE) in Portugal. The control algorithm, EnergAIze [28] is a decentralized multi-agent reinforcement learning and is benchmarked against a baseline where there are no EV loads. A synthetic 12-EV schedule is used to train the reinforcement learning (RL) agent to prioritize.

Figure 9 shows an EV's controlled SoC with respect to expected departure SoC and real-time pricing for a day, where the green region is when the EV is connected to a charger. At 8 am, the EV leaves with an SoC close to what is expected at departure. On arrival (6 pm onwards), the RL agent learns to charge when pricing is low and to discharge to the building when pricing is high.

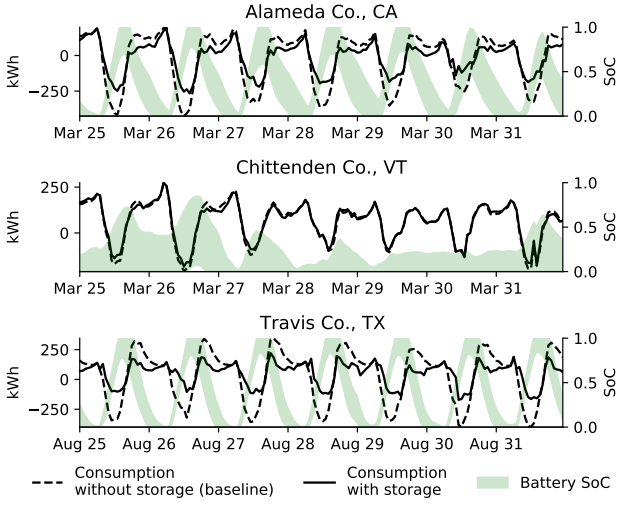


Figure 8: Results from [67] comparing district-level net electricity consumption for scenarios with and without storage. The shaded regions show the average SoC^{BESS} ± one standard deviation.

4.4. Occupant comfort feedback during automated demand response

A case study is presented in [45], where the heat pumps of ten single-family homes are controlled for DR during a three-month winter heating period in Montreal, Quebec. The study included two occupant types, Average and Tolerant, based on different thermostat temperature setpoint preferences, and the integration of the occupants into the control scheme was modeled using three level of details (LoDs). LoD 1 represents the baseline scenario where fixed setpoint

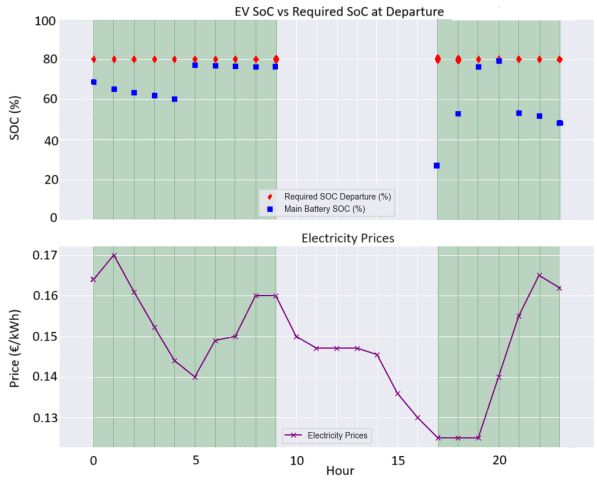


Figure 9: Controlled EV SoC in V2G control scenario (top) to reduce energy cost for a real-time pricing scheme (bottom). The green region is when the EV is connected to a charger.

schedules are used for each occupant type with no control of the HVAC system. LoD 2 also uses a fixed setpoint schedule but assumes occupant thermal comfort within a band of 2°C from the setpoint schedule following the methodology used in [80], thus allowing for flexibility of the HVAC system. In the most detailed LoD, LoD 3 incorporates occupant-thermostat override models and DR events. The setpoint is reduced by 1.1°C (2°F) during the three or four-hour DR events and the occupants can adjust the setpoint at each hourly time step due to thermal discomfort, which is dictated by the occupant-thermostat override models. We refer the reader to the source code repository for reproducibility[2].

The EULP dataset provides input data for building energy models in CityLearn, and the objective is to determine the flexibility services provided through the automated control of residential heating systems in a ten-home neighborhood through setpoint adjustments during DR events. In LoD 1 there is no controlled resource, while for LoD 2 and LoD 3 the SAC RL algorithm [34] is used to control the heat pump power to deliver adequate heating energy to each building to maintain indoor temperature in the comfort range for the occupant while providing flexibility. The control is evaluated using grid-level KPIs including cost, total electricity consumption, and average daily peak, while also quantifying occupant discomfort through the number of thermostat overrides made. Figure 10 shows the electricity consumption by household during the DR events, during which LoD 3 implemented thermostat setbacks to specifically target electricity reductions during these hours while LoD 1 and LoD 2 use the normal setpoint schedules. The results show that during the DR events, the electricity consumption was reduced by approximately 17% on average during these peak hours for LoD 3, even with some occupants overriding the setpoint changes during the DR events.

[2] https://github.com/kkaspar10/Occupant_Thermostat_Int/tree/master

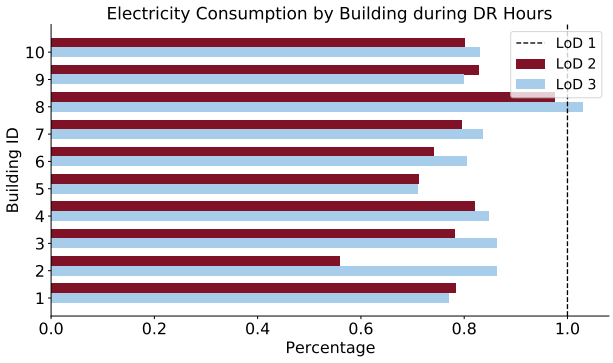


Figure 10: Total electricity consumption during DR hours for LoDs 2 and 3 with respect to LoD 1.

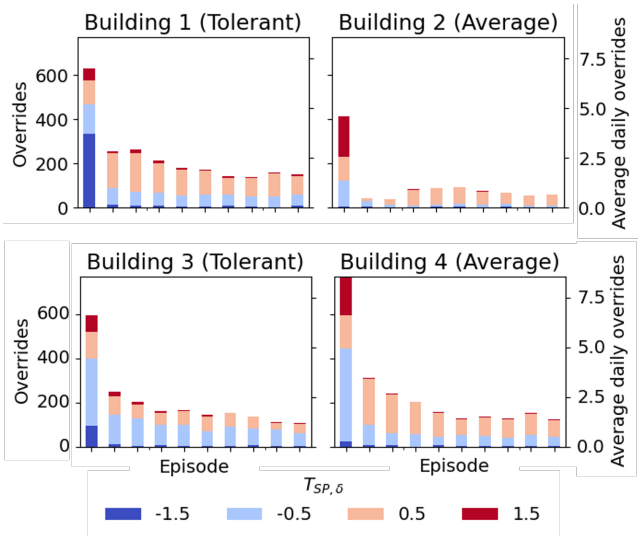


Figure 11: Number of occupant overrides and magnitude of the override for the 10 training episodes (adapted from [45]).

In addition to energy efficiency metrics, the agents are able to also reduce the number of thermostat overrides over time. Figure 11 shows the number of thermostat overrides throughout the ten training episodes for Buildings 1 and 3 with Tolerant occupants and Buildings 2 and 4 with Average occupants, where the darker red and blue represent larger changes to the setpoint and the light red and blue represent smaller changes to the setpoint. The figure shows that the agents were thus able to learn the best actions while considering the unique preferences and override behaviors of the two occupant types, and by using a multi-agent approach, the agents are able to learn optimal decision-making that achieves both occupant thermal comfort and DR goals.

5. Discussion

5.1. Energy flexibility quantification

The heat pump power control example that balances comfort and electricity use (see Table 3) shows that one KPI may be improved at the expense of the other. Thus, it is important to have a comprehensive quantification of flexibility

caused by a control solution, including the adverse effects it may have. Methods for energy flexibility quantification is an open research question [54]. The work by Li et al. has focused on bridging this research gap by identifying 48 data-driven energy flexibility KPIs in the literature [52]. Ongoing work in [44] is to provide these KPIs in a Python package for integration into simulation environments such as CityLearn. Thus, a future CityLearn release will integrate this package to provide users with a plethora of energy flexibility KPIs that better inform their control design.

5.2. Energy flexibility potential of the U.S. residential building stock

The example in Section 4.2 assessed energy flexible neighborhoods in only three climate zones with a limited control period of three months in the winter or summer. While the work demonstrates the efficacy of the framework, it is limited in providing conclusive results on energy flexibility of the U.S. residential building stock for the current and future climatic conditions. Also, the considered controlled DERs are limited to only ESS for load shifting without consideration of flexibility provided by other DERs e.g. heat pumps and EVs nor consideration for control resilience and occupant model in the control loop. Future work will conduct a comprehensive analysis of various DER availability configurations in representative neighborhoods of at least, the 16 climate zones classifications by ASHRAE that differ in temperature and humidity conditions, to provide a holistic view of the energy flexibility potential of the residential building stock.

The authors of the EULP dataset recommend using at least 1000 unique building energy models to smoothen out load profiles on aggregate as fewer models may be biased towards the characteristics of the sampled models [99]. They reference electricity quantities of interest (QOI) including peak load, peak timing, and average annual consumption, to determine the uncertainty brought about by different model sample sizes. At 1000 models, the uncertainty is kept at 15% or less.

Nevertheless, the EULP dataset authors propose making assertions based on individual building EULP when studying a distribution of possible outcomes e.g. the impact of PV-BESS sizing on electricity bill and peak demand [78], which align with our application of the dataset to study the outcomes from different control algorithms on DER-provided energy flexibility. This is similar to the work in Earle et al. that uses only 30 models to represent a synthetic single-family home community in Denver, Colorado, for MPC of DERs as well as the work in [62] where only five models per location are used to represent typical residential buildings in Sacramento, Boston, and Dallas, for the investigation of resilience-based RBC and MPC of DERs. Additionally, computational overhead that comes with large sample sizes may render them prohibitive leading to fewer than 1000 models being used to represent geographical areas that are as small as county-level [31] or large as country-wide [59].

Moreover, our framework for designing virtual grid-communities in CityLearn that leverages the EULP dataset (Section 2.5) does not impose any upper limit on the number of buildings that can be included in a community. The three datasets that we release using this framework represent hypothetical single-family neighborhoods in Alameda County, California; Chittenden County, Vermont; and Travis County; Texas. Our synthetic Alameda and Travis County neighborhoods are made up of 100 building models each, which will result in only 10% or less uncertainty in eight out of eleven QOI while the remaining QOI have less than 30% uncertainty with reference to the results of the sampling size experiment carried our in [99]. The Chittenden County neighborhood has 47 building models hence, will have higher uncertainty. Irrespective of these datasets having less than the recommended 1000 sample size limit, we make an effort to ensure our sampled buildings are representative of the overall building characteristics through cluster-frequency-based sampling. Users may also, apply our framework with larger sample sizes or more robust clustering techniques that may include other building metadata but also occupant behavior classification based on equipment schedules.

5.3. Electric vehicle impact on energy flexibility

The V2G control example shows the potential benefits of increased EV adoption when properly managed. The observation across the different chargers in the district for the entire four-year simulation period is an 11% reduction in energy cost and 21% reduction in average daily peak (see [27]). With policies aimed at replacing internal combustion engine vehicle fleets with EVs, the EV share of controllable DERs will increase. There has been a rise of EV sales from 4% of total worldwide car sales in 2020 to over 14% in 2022 [42]. These sales are incentivized by tax credits [43], access to carpool lanes, reduced car registration fees, and subsidies for constructing residential charging stations [30].

5.4. Occupant impact on energy flexibility

The work on integrating the occupant thermal comfort feedback presented in Section 4.4, it is important to note that all occupant modeling was done using ecobee's Donate Your Data dataset [58] for homes in Quebec, Canada during the winter heating season. As such, future work would be required to determine the occupant models' applicability in other climates, and additional location-specific override models could be developed using the methodology presented in the study. However, one of the key contributions of this work is showing the impact that the occupant behavior LoD has on the load shifting potential and occupant thermal comfort. We provide several examples via the LoDs of how one may incorporate occupants in the control scheme that reflect a more realistic scenario of a DR program in which occupants are allowed to opt out of thermostat setbacks during the DR events.

5.5. Resilience model

Our power outage model assumes a constant blackout probability on any given day of the year. This aligns with assumptions that have been made in existing literature [77]. Although, weather-related and other natural events events have a large share on occurrences and number of affected customers on aggregate, historically, other man-made events cause power outages including, equipment failure, operator error, voltage reduction, supply shortage, intentional attack, and fire [35]. To incorporate the correlation of extreme weather events into the power outage model, periods of cold snaps and heat waves need to be identified. The work by Ouzeau et al. provides a methodology to identify cold snap and heat wave events that is based on location-specific quantiles of daily temperature distributions over a 30-year period [75]. The Bernoulli probability in Equation (5) may then be reduced or set to zero for days that do not experience a cold snap nor heat wave.

5.6. Heat pump model simplification

We note that the heat pump model in CityLearn assumes an ideal refrigeration cycle and the coefficient of performance (COP) is temperature-driven alone. Thus, the partial load behaviour may slightly underestimate or overestimate the COP. Also, the control signal for the heat pump at this time does not allow for thermostatic control of the heat pump through setpoints which could limit its suitability for true demand response applications. Thus, CityLearn is best applied to simple and quick comparative analysis between control algorithms with minimal design cost. Whereas, an environment that utilizes high-fidelity energy models e.g., OCHRE [11] is better suited for analyses that are purposed for scrutiny of absolute energy and cost values. Improving the heat pump model will be considered in future work but users may implement their model of choice while using the current model definition as a reference.

5.7. District energy systems

One limitation of CityLearn is that there are no district energy system models such as centralized PV system, ESS, and thermal energy networks. However, from a control perspective, the buildings can share information about their states to achieve district-level energy flexibility. Future work could consider the integration of such systems in the environment with reference to simulation environments like CityBES [37] and URBANopt [23]. Future development of CityLearn may also include adding multi-zone building thermal models that allow simultaneous cooling and heating as well as consideration of indoor humidity loads, and adding indoor air quality models and consideration of IEQ metrics in the KPIs. Field validation of CityLearn in real building districts is also needed to demonstrate its usability and performance, as well as to inform the development of best practice guides.

5.8. Building archetype limitations

We build CityLearn purposefully not as a co-simulation of different environments, but as a single environment framework to simplify the process of control benchmarking and adoption in existing workflows. To this end, there are typically two use-cases in CityLearn. In the first, load shifting and shedding flexibility are provided through heat pump control thus, affecting the temperature and comfort in a thermal zone. This use-case is limited to single-family residential building archetype because of the LSTM thermal dynamics model that assumes a single thermal zone that typically applies to this archetype. Our examples in Sections 4.1 and 4.4 reflect this use-case. In a multi-family or commercial building scenario where there are multiple thermal zones maintained at different temperature setpoints and serviced by different HVAC systems, the heat transfer between thermal zones will not be captured in the current CityLearn setup. Also, the occupant thermostat override model methodology is based on data from residential buildings and will not reflect occupant behavior in a commercial building. Thus, users should consider these limitations when setting up their environment. Future work will expand the thermal dynamics modeling approach to include techniques such as grey-box RC models to offer more options for users. CityLearn being open-source, users can also integrate their own dynamic models.

Nevertheless, there is the second use-case in CityLearn that permits the inclusion of multi-family and commercial buildings where load shifting flexibility is provided from active ESSs. In this use-case, the building's thermal demand and other electrical loads are not altered and are provided to CityLearn in time series form from BEPS software e.g., EnergyPlus or smart metering. The EV example (Section 4.3) is one of such use-case as well as applications in existing literature [93, 72].

5.9. Environment observability

The observations listed in Table A.1, though exhaustive and indeed observable, may be cost prohibitive to sense or calculate in practice [12]. An example of such costly observation is solar irradiance where few sources provide real-time irradiance data and in most cases are behind a paywall. Similarly, the real time cooling, heating and DHW loads in a building without an accompanying digital twin are challenging to measure compared to the electricity consumption that these loads depend on. Also, EV departure time is not always accurately known apriori without deliberate insight from the occupant. Yet, the simulation and real-world environments should have similar control spaces for reproducibility of simulation results in practice. Thus, the control designer must consider the practical implications of their observation space on real world adaptation.

5.10. Scalability of reinforcement learning control

While existing literature has demonstrated control scalability in CityLearn [3], a challenge of large scale analysis especially using RLC is that it may be computationally and data cost prohibitive to train a unique control policy for each building. Transfer learning can alleviate these costs by leveraging the trained policy of a source building to initialize those of the remaining target buildings with outcomes that

are better or comparable to baseline and reference control policies [69, 16]. On the other hand, with the diversity of control tasks i.e., controllable DERs, meta-learning can provide a generalized algorithm-agnostic control policy that can quickly adapt to any single control task [8]. Yet, there are few applications of meta-learning in building control and demand response [60]. A meta-policy will allow for decentralized heterogeneous control that depicts real world large scale deployment of advanced controllers where different buildings, and in special cases, different DERs, use unique control algorithms, observation and action spaces [100].

5.11. The CityLearn Challenge

Despite its potential, there are still open questions regarding the plug-and-play capabilities, performance, safety of operation, and learning speed of RLC [21]. The CityLearn Challenge is an avenue for crowd-sourcing solutions to building control problems, using any control algorithm of choice. It attracts a multidisciplinary participation including researchers, industry experts, sustainability enthusiasts and artificial intelligence (AI) hobbyists with a focus on DER control algorithm benchmarking for demand response and energy flexibility settings. Previous editions of The CityLearn Challenge have investigated control policy transferrability across different climate zones [96], a realistic implementation of model-free RLC in buildings where training and evaluation were done on a single four-year episode [64], and BESS control using a real-world ZNE community dataset for cost and emissions optimization [71]. Other competitions in the energy domain focus on building load predictions [61], grid power flow optimization [36] and pathways to building electrification and decarbonization [74]. Thus, we envision opportunities for collaboration amongst these different competitions where their unique objectives and experience, can provide holistic solutions to the problems in the building energy and control communities.

6. Conclusion

In this paper, we introduce CityLearn v2, an environment based on the Farama Foundation Gymnasium interface, that leverages the End-Use Load Profiles for the U.S. Building Stock dataset to create virtual grid-interactive communities for benchmarking resilient multi-agent, distributed energy resource and objective control with dynamic occupant feedback. CityLearn allows for custom building districts as users can select which buildings to include, distributed energy resources to control, and observations to sense. Users can also design custom reward functions and choose between three control configurations: central single-agent, independent multi-agent and coordinated multi-agent with provision of built-in control algorithms and support for third-party control libraries. These modelling options improve the level of realism in CityLearn, which is an important guideline when designing these kinds of environments [101].

Rolnick et al. investigate how the machine learning community can help tackle climate change [84]. The authors identify electrification of urban energy systems as a field where machine learning can provide high leverage. The review of use cases of machine learning in the building's life cycle by Hong et al. indicates promising potential of machine learning based smart controls for grid-interactive efficient buildings [39]. CityLearn is aimed precisely at this objective, and a standardized environment, distributed as educational material [73], will allow more researchers from the computer science and machine learning communities to participate in the search for solutions to tackle climate change. Also, by simplifying the building energy system and providing different environment datasets, CityLearn can help building energy engineers and control engineers to focus on control algorithm development with less time spent setting up the environment.

Lastly, the application examples presented in this work are by no means an exhaustive demonstration of CityLearn's functionalities. Instead, we envision that our contribution provides environment and control examples as well as their implementation source code for actors in the building control community who are interested in using CityLearn to solve their distributed energy resource control problems.

CRediT authorship contribution statement

Kingsley Nweye: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization . **Kathryn Kaspar:** Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft . **Giacomo Buscemi:** Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft . **Tiago Fonseca:** Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization . **Giuseppe Pinto:** Methodology, Software, Investigation, Data curation, Writing - review & editing . **Dipanjan Ghose:** Validation, Writing - original draft, Visualization . **Satvik Duddukuru:** Software, Data curation . **Pavani Pratapa:** Software, Data curation . **Han Li:** Methodology, Writing - review & editing . **Javad Mohammadi:** Writing - review & editing, Supervision . **Luis Lino Ferreira:** Methodology, Resources, Writing - review & editing, Supervision . **Tianzhen Hong:** Methodology, Writing - review & editing, Supervision . **Mohamed Ouf:** Methodology, Resources, Writing - review & editing, Supervision . **Alfonso Capozzoli:** Methodology, Resources, Writing - review & editing, Supervision . **Zoltan Nagy:** Conceptualization, Methodology, Resources, Writing - review & editing, Supervision, Project administration .

7. Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

- Kingsley Nweye, Dipanjan Ghose, Javad Mohammadi, and Zoltan Nagy received funding from the

Climate Change AI Innovation Grants program (Contract IG-2023-32). Javad Mohammad would like to acknowledge the support from NSF (# 2313768) and AFOSR (# FA9550-23-1-0203).

- Kathryn Kaspar is supported by the Fonds de recherche du Québec – Nature et technologies (FRQNT) Doctoral Research Scholarship (B2X). Mohamed Ouf was supported by the FRQNT Research Support for New Academics (Grant #315109) and the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (RGPIN-2020-06804).

- The work of Alfonso Capozzoli and Giacomo Buscemi was done in the framework of the PRIN 2020 project: "OPTIMISM—Optimal refurbishment design and management of small energy micro-grids", funded by the Italian Ministry of University and Research (MUR).

- The work by Tiago Fonseca and Luis Lino Ferreira was supported by the OPEVA project that has received funding within the Chips Joint Undertaking (Chips JU) from the European Union's Horizon Europe Programme and the National Authorities (France, Czechia, Italy, Portugal, Turkey, Switzerland), under grant agreement 101097267. The paper is also supported by Arrowhead PVN, proposal number 101097257. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Chips JU. Neither the European Union nor the granting authority can be held responsible for them.

- The work by Han Li and Tianzhen Hong was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Building Technologies of the United States Department of Energy, under Contract No. DE-AC02-05CH11231.

8. Data availability statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

[1] Afzalan, M., Jazizadeh, F., 2019. Residential loads flexibility potential for demand response using energy consumption patterns and user segments. Applied Energy 254, 113693. URL: https://www.sciencedirect.com/science/article/pii/S0306261919313807, doi:10.1016/j.apenergy.2019.113693.

[2] Akbari, S., Haghighat, F., 2021. Occupancy and occupant activity drivers of energy consumption in residential buildings. Energy and Buildings 250, 111303. URL: https://www.sciencedirect.com/science/article/pii/S0378778821005879, doi:10.1016/j.enbuild.2021.111303.

[3] Almilaify, Y., Nweye, K., Nagy, Z., 2023. SCALEX: SCALability EXploration of Multi-Agent Reinforcement Learning Agents in Grid-Interactive Efficient Buildings, in: Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, Association for Computing Machinery, New York, NY, USA. pp. 261–264. URL: https://doi.org/10.1145/3600100.3623749, doi:10.1145/3600100.3623749.

[4] Arendt, K., Jradi, M., Shaker, H., Veje, C., 2018. Comparative Analysis of White-, Gray- and Black-Box Models for Thermal Simulation of Indoor Environment: Teaching Building Case Study, ASHRAE/IBPSA-USA. pp. 173–180. URL: https://publications.ibpsa.org/conference/paper/?id=simbuild2018_C025.

[5] Arroyo, J., Manna, C., Spiessens, F., Helsen, L., 2021. An Open-AI gym environment for the Building Optimization Testing (BOPTEST) framework, IBPSA. pp. 175–182. URL: https://publications.ibpsa.org/conference/paper/?id=bs2021_30380, doi:10.26868/25222708.2021.30380.

[6] Arroyo, J., Verleyen, L., Bex, L., Hermans, L., Saeed, M., Lu, Y., Depoortere, J., Bálint, A., Delarue, E., Driesen, J., Deconinck, G., Blum, D., Helsen, L., 2023. Prototyping the DOPTEST Framework for Simulation-Based Testing of System Integration Strategies in Districts.

[7] Bampoulas, A., Saffari, M., Pallonetto, F., Mangina, E., Finn, D.P., 2021. A fundamental unified framework to quantify and characterise energy flexibility of residential buildings with multiple electrical and thermal energy systems. Applied Energy 282, 116096. URL: https://www.sciencedirect.com/science/article/pii/S0306261920315191, doi:10.1016/j.apenergy.2020.116096.

[8] Beck, J., Vuorio, R., Liu, E.Z., Xiong, Z., Zintgraf, L., Finn, C., Whiteson, S., 2023. A Survey of Meta-Reinforcement Learning URL: https://arxiv.org/abs/2301.08028, doi:10.48550/ARXIV.2301.08028.

[9] Birk Jones, C., Vining, W., Lave, M., Haines, T., Neuman, C., Bennett, J., Scoffield, D.R., 2022. Impact of Electric Vehicle customer response to Time-of-Use rates on distribution power grids. Energy Reports 8, 8225–8235. URL: https://www.sciencedirect.com/science/article/pii/S235248472201188X, doi:10.1016/j.egyr.2022.06.048.

[10] Blair, N., Dobos, A.P., Freeman, J., Neises, T., Wagner, M., Ferguson, T., Gilman, P., Janzou, S., 2014. System advisor model, sam 2014.1. 14: General description. Technical Report. National Renewable Energy Lab.(NREL), Golden, CO (United States).

[11] Blonsky, M., Maguire, J., McKenna, K., Cutler, D., Balamurugan, S.P., Jin, X., 2021. OCHRE: The Object-oriented, Controllable, High-resolution Residential Energy Model for Dynamic Integration Studies. Applied Energy 290, 116732. URL: https://www.sciencedirect.com/science/article/pii/S0306261921002464, doi:10.1016/j.apenergy.2021.116732.

[12] Blum, D., Arroyo, J., Huang, S., Drgoňa, J., Jorissen, F., Walnum, H.T., Chen, Y., Benne, K., Vrabie, D., Wetter, M., Helsen, L., 2021. Building optimization testing framework (BOPTEST) for simulation-based benchmarking of control strategies in buildings. Journal of Building Performance Simulation 14, 586–610. URL: https://www.tandfonline.com/doi/full/10.1080/19401493.2021.1986574, doi:10.1080/19401493.2021.1986574.

[13] Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., Zaremba, W., 2016. OpenAI Gym. URL: http://arxiv.org/abs/1606.01540, doi:10.48550/arXiv.1606.01540. arXiv:1606.01540 [cs].

[14] Busby, J.W., Baker, K., Bazilian, M.D., Gilbert, A.Q., Grubert, E., Rai, V., Rhodes, J.D., Shidore, S., Smith, C.A., Webber, M.E., 2021. Cascading risks: Understanding the 2021 winter blackout in Texas. Energy Research & Social Science 77, 102106. URL: https://www.sciencedirect.com/science/article/pii/S2214629621001997, doi:10.1016/j.erss.2021.102106.

[15] California Independent System Operator, 2016. What the duck curve tells us about managing a green grid. URL: https://www.caiso.com/documents/flexibleresourceshelprenewables_fastfacts.pdf.

[16] Coraci, D., Brandi, S., Hong, T., Capozzoli, A., 2023. Online transfer learning strategy for enhancing the scalability and deployment of deep reinforcement learning control in smart buildings. Applied Energy 333, 120598. URL: https://www.sciencedirect.com/science/article/pii/S0306261922018554, doi:10.

1016/j.apenergy.2022.120598.

[17] Crawley, D.B., Lawrie, L.K., Winkelmann, F.C., Buhl, W.F., Huang, Y.J., Pedersen, C.O., Strand, R.K., Liesen, R.J., Fisher, D.E., Witte, M.J., Glazer, J., 2001. EnergyPlus: creating a new-generation building energy simulation program. Energy and Buildings 33, 319–331.

[18] Deltetto, D., 2020. Data-driven coordinated building cluster energy management to enhance energy efficiency, comfort and grid stability. laurea. Politecnico di Torino. URL: https://webthesis.biblio.polito.it/16358/.

[19] D'Oca, S., Gunay, H.B., Gilani, S., O'Brien, W., 2019. Critical review and illustrative examples of office occupant modelling formalisms. Building Services Engineering Research and Technology 40, 732–757. URL: https://doi.org/10.1177/0143624419827468, doi:10.1177/0143624419827468. publisher: SAGE Publications Ltd STM.

[20] Drgoňa, J., Arroyo, J., Cupeiro Figueroa, I., Blum, D., Arendt, K., Kim, D., Ollé, E.P., Oravec, J., Wetter, M., Vrabie, D.L., Helsen, L., 2020. All you need to know about model predictive control for buildings. Annual Reviews in Control 50, 190–232. doi:10.1016/j.arcontrol.2020.09.001.

[21] Dulac-Arnold, G., Levine, N., Mankowitz, D.J., Li, J., Paduraru, C., Gowal, S., Hester, T., 2021. Challenges of real-world reinforcement learning: definitions, benchmarks and analysis. Machine Learning 110, 2419–2468. Publisher: Springer.

[22] Earle, L., Maguire, J., Munankarmi, P., Roberts, D., 2023. The impact of energy-efficiency upgrades and other distributed energy resources on a residential neighborhood-scale electrification retrofit. Applied Energy 329, 120256. URL: https://www.sciencedirect.com/science/article/pii/S0306261922015136, doi:10.1016/j.apenergy.2022.120256.

[23] El Kontar, R., Polly, B., Charan, T., Fleming, K., Moore, N., Long, N., Goldwasser, D., 2020. URBANopt: An Open-Source Software Development Kit for Community and Urban District Energy Modeling: Preprint. Technical Report NREL/CP-5500-76781. National Renewable Energy Lab. (NREL), Golden, CO (United States). URL: https://www.osti.gov/biblio/1677416.

[24] Electric Reliability Council of Texas, 2021. Fuel Mix Report: 2007 - 2020. URL: https://www.ercot.com/gridinfo/generation.

[25] Fanger, P.O., 1970. Thermal Comfort: Analysis and Applications in Environmental Engineering. Danish Technical Press. Google-Books-ID: S0FSAAAAMAAJ.

[26] Feng, K., Ouyang, M., Lin, N., 2022. Tropical cyclone-blackout-heatwave compound hazard resilience in a changing climate. Nature Communications 13, 4421. URL: https://www.nature.com/articles/s41467-022-32018-4, doi:10.1038/s41467-022-32018-4. publisher: Nature Publishing Group.

[27] Fonseca, T., Ferreira, L., Cabral, B., Severino, R., Nweye, K., Ghose, D., Nagy, Z., 2024a. EVLearn: Extending the CityLearn Framework with Electric Vehicle Simulation. URL: http://arxiv.org/abs/2404.06521, doi:10.48550/arXiv.2404.06521. arXiv:2404.06521 [cs, eess].

[28] Fonseca, T., Ferreira, L., Cabral, B., Severino, R., Praca, I., 2024b. EnergAIze: Multi Agent Deep Deterministic Policy Gradient for Vehicle to Grid Energy Management. URL: http://arxiv.org/abs/2404.02361, doi:10.48550/arXiv.2404.02361. arXiv:2404.02361 [cs].

[29] Galvin, R., 2022. Are electric vehicles getting too big and heavy? Modelling future vehicle journeying demand on a decarbonized US electricity grid. Energy Policy 161, 112746. URL: https://www.sciencedirect.com/science/article/pii/S0301421521006121, doi:10.1016/j.enpol.2021.112746.

[30] Ghose, D., 2023. Studying present and future electric vehicle impacts on the city of Austin's power grid. Ph.D. thesis. URL: https://hdl.handle.net/2152/119250.

[31] Gorman, W., Barbose, G., Pablo Carvallo, J., Baik, S., Miller, C.A., White, P., Praprost, M., 2023. County-level assessment of behind-the-meter solar and storage to mitigate long duration power interruptions for residential customers. Applied Energy 342, 121166. URL: https://www.sciencedirect.com/science/article/pii/S0306261923005305, doi:10.1016/j.apenergy.2023.121166.

[32] Gowrisankaran, G., Reynolds, S.S., Samano, M., 2016. Intermittency and the Value of Renewable Energy. Journal of Political Economy 124, 1187–1234. URL: https://doi.org/10.1086/686733, doi:10.1086/686733. _eprint: https://doi.org/10.1086/686733.

[33] Gunay, H.B., O'Brien, W., Beausoleil-Morrison, I., Bursill, J., 2018. Development and implementation of a thermostat learning algorithm. Science and Technology for the Built Environment 24, 43–56. URL: https://doi.org/10.1080/23744731.2017.1328956, doi:10.1080/23744731.2017.1328956. publisher: Taylor & Francis _eprint: https://doi.org/10.1080/23744731.2017.1328956.

[34] Haarnoja, T., Zhou, A., Abbeel, P., Levine, S., 2018. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor URL: https://arxiv.org/abs/1801.01290, doi:10.48550/ARXIV.1801.01290.

[35] Hines, P., Apt, J., Talukdar, S., 2008. Trends in the history of large blackouts in the United States, in: 2008 IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, pp. 1–8. URL: https://ieeexplore.ieee.org/document/4596715, doi:10.1109/PES.2008.4596715. iSSN: 1932-5517.

[36] Holzer, J., Coffrin, C., DeMarco, C., Duthu, R., Elbert, S., Greene, S., Kuchar, O., Lesieutre, B., Li, H., Mak, W.K., others, 2021. Grid Optimization Competition Challenge 2 Problem Formulation. Technical Report. Tech. rep. ARPA-E.

[37] Hong, T., Chen, Y., Lee, S.H., Piette, M., 2016. CityBES: A Web-based Platform to Support City-Scale Building Energy Efficiency.

[38] Hong, T., Malik, J., Krelling, A., O'Brien, W., Sun, K., Lamberts, R., Wei, M., 2023. Ten questions concerning thermal resilience of buildings and occupants for climate adaptation. Building and Environment 244, 110806. URL: https://www.sciencedirect.com/science/article/pii/S0360132323008338, doi:10.1016/j.buildenv.2023.110806.

[39] Hong, T., Wang, Z., Luo, X., Zhang, W., 2020. State-of-the-art on research and applications of machine learning in the building life cycle. Energy and Buildings 212, 109831. URL: https://www.sciencedirect.com/science/article/pii/S0378778819337879, doi:10.1016/j.enbuild.2020.109831.

[40] Horsey, H., Parker, A., Farthing, A., Dahlhausen, M., Praprost, M., Bianchi, C., Robertson, J., Horowitz, S., Zhang, L., 2020. ComStock™. Technical Report. National Renewable Energy Laboratory (NREL), Golden, CO (United States). URL: https://www.osti.gov/biblio/1817851, doi:10.11578/dc.20210830.5.

[41] IEEE, 2012. IEEE Guide for Electric Power Distribution Reliability Indices. doi:10.1109/IEEESTD.2012.6209381. conference Name: IEEE Std 1366-2012 (Revision of IEEE Std 1366-2003).

[42] International Energy Agency, . Trends in electric light-duty vehicles – Global EV Outlook 2023 – Analysis. URL: https://www.iea.org/reports/global-ev-outlook-2023/trends-in-electric-light-duty-vehicles.

[43] Jenn, A., Springel, K., Gopal, A.R., 2018. Effectiveness of electric vehicle incentives in the United States. Energy Policy 119, 349–356. URL: https://www.sciencedirect.com/science/article/pii/S0301421518302891, doi:10.1016/j.enpol.2018.04.065.

[44] Johra, H., Li, H., de Andrade Pereira, F., Nweye, K., Chamari, L., Nagy, Z., 2023. IEA EBC Annex 81 â€“ Data-driven smart buildings: Building-to-grid applications, in: Proceedings of building simulation 2023: 18th conference of IBPSA, IBPSA, Shanghai, China. pp. 2537–2544. URL: https://publications.ibpsa.org/conference/paper/?id=bs2023_1305, doi:https://doi.org/10.26868/25222708.2023.1305. iSSN: 2522-2708 tex.organisation: IBPSA.

[45] Kaspar, K., Nweye, K., Buscemi, G., Capozzoli, A., Nagy, Z., Pinto, G., Eicker, U., Ouf, M.M., 2024. Effects of occupant thermostat preferences and override behavior on residential demand response in CityLearn. Energy and Buildings , 114830URL: https://www.sciencedirect.com/science/article/pii/S0378778824009460, doi:10.1016/j.enbuild.2024.114830.

[46] Klaassen, E.A.M., Kobus, C.B.A., Frunt, J., Slootweg, J.G., 2016. Responsiveness of residential electricity demand to dynamic tariffs: Experiences from a large field test in the Netherlands. Applied Energy 183, 1065–1074. URL: https://www.sciencedirect.com/science/article/pii/S0306261916313538, doi:10.1016/j.apenergy.2016.09.051.

[47] Kostopoulos, E.D., Spyropoulos, G.C., Kaldellis, J.K., 2020. Real-world study for the optimal charging of electric vehicles. Energy Reports 6, 418–426. URL: https://www.sciencedirect.com/science/article/pii/S2352484719310911, doi:10.1016/j.egyr.2019.12.008.

[48] Langtry, M., Wichitwechkarn, V., Ward, R., Zhuang, C., Kreitmair, M.J., Makasis, N., Xuereb Conti, Z., Choudhary, R., 2024. Impact of data for forecasting on performance of model predictive control in buildings with smart energy storage. Energy and Buildings 320, 114605. URL: https://www.sciencedirect.com/science/article/pii/S0378778824007217, doi:10.1016/j.enbuild.2024.114605.

[49] Le Dréau, J., Heiselberg, P., 2016. Energy flexibility of residential buildings using short term heat storage in the thermal mass. Energy 111, 991–1002. URL: https://www.sciencedirect.com/science/article/pii/S0360544216306934, doi:10.1016/j.energy.2016.05.076.

[50] Lee, J., Kim, J.M., Yi, J., Won, C.Y., 2021. Battery Management System Algorithm for Energy Storage Systems Considering Battery Efficiency. Electronics 10, 1859. URL: https://www.mdpi.com/2079-9292/10/15/1859, doi:10.3390/electronics10151859. number: 15 Publisher: Multidisciplinary Digital Publishing Institute.

[51] Lee, Z.E., Max Zhang, K., 2022. Unintended consequences of smart thermostats in the transition to electrified heating. Applied Energy 322, 119384. URL: https://www.sciencedirect.com/science/article/pii/S0306261922007243, doi:10.1016/j.apenergy.2022.119384.

[52] Li, H., Johra, H., de Andrade Pereira, F., Hong, T., Le Dréau, J., Maturo, A., Wei, M., Liu, Y., Saberi-Derakhtenjani, A., Nagy, Z., Marszal-Pomianowska, A., Finn, D., Miyata, S., Kaspar, K., Nweye, K., O'Neill, Z., Pallonetto, F., Dong, B., 2023. Data-driven key performance indicators and datasets for building energy flexibility: A review and perspectives. Applied Energy 343, 121217. URL: https://www.sciencedirect.com/science/article/pii/S0306261923005810, doi:10.1016/j.apenergy.2023.121217.

[53] Li, H., Wang, Z., Hong, T., Piette, M.A., 2021. Energy flexibility of residential buildings: A systematic review of characterization and quantification methods and applications. Advances in Applied Energy 3, 100054. URL: https://www.sciencedirect.com/science/article/pii/S2666792421000469, doi:https://doi.org/10.1016/j.adapen.2021.100054.

[54] Li, R., Satchwell, A.J., Finn, D., Christensen, T.H., Kummert, M., Le Dréau, J., Lopes, R.A., Madsen, H., Salom, J., Henze, G., Wittchen, K., 2022. Ten questions concerning energy flexibility in buildings. Building and Environment 223, 109461. URL: https://www.sciencedirect.com/science/article/pii/S0360132322006928, doi:10.1016/j.buildenv.2022.109461.

[55] Liang, E., Liaw, R., Moritz, P., Nishihara, R., Fox, R., Goldberg, K., Gonzalez, J.E., Jordan, M.I., Stoica, I., 2017. RLlib: Abstractions for Distributed Reinforcement Learning. URL: https://arxiv.org/abs/1712.09381v4.

[56] Liu, W., Gunay, H.B., Ouf, M.M., 2021. Modeling window and thermostat use behavior to inform sequences of operation in mixed-mode ventilation buildings. Science and Technology for the Built Environment 27, 1204–1220. URL: https://doi.org/10.1080/23744731.2021.1936629, doi:10.1080/23744731.2021.1936629. publisher: Taylor & Francis _eprint: https://doi.org/10.1080/23744731.2021.1936629.

[57] Lu, X., Fu, Y., O'Neill, Z., 2023. Benchmarking high performance HVAC Rule-Based controls with advanced intelligent Controllers: A case study in a Multi-Zone system in Modelica. Energy and Buildings 284, 112854. URL: https://www.sciencedirect.com/science/article/pii/S0378778823000841, doi:10.1016/j.enbuild.2023.112854.

[58] Luo, N., Hong, T., 2022. Ecobee Donate Your Data 1,000 homes in 2017. Technical Report 69035. Pacific Northwest National Lab. (PNNL), Richland, WA (United States); Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States). URL: https://www.osti.gov/biblio/1854924, doi:10.25584/ecobee/1854924.

[59] Maguire, J., Blonsky, M., Ericson, S., Farthing, A., Manogaran, I., Ramaraj, S., 2024. Nova Analysis: Holistically Valuing the Contributions of Residential Efficiency, Solar and Storage. Technical Report NREL/TP-5500-84658. National Renewable Energy Laboratory (NREL), Golden, CO (United States). URL: https://www.osti.gov/biblio/2311138, doi:10.2172/2311138.

[60] Mason, K., Grijalva, S., 2019. A review of reinforcement learning for autonomous building energy management. Computers & Electrical Engineering 78, 300–312. URL: https://www.sciencedirect.com/science/article/pii/S0045790618333421, doi:10.1016/j.compeleceng.2019.07.019.

[61] Miller, C., Arjunan, P., Kathirgamanathan, A., Fu, C., Roth, J., Park, J.Y., Balbach, C., Gowri, K., Nagy, Z., Fontanini, A.D., Haberl, J., 2020. The ASHRAE Great Energy Predictor III competition: Overview and results. Science and Technology for the Built Environment 26, 1427–1447. URL: https://doi.org/10.1080/23744731.2020.1795514, doi:10.1080/23744731.2020.1795514. publisher: Taylor & Francis _eprint: https://doi.org/10.1080/23744731.2020.1795514.

[62] Munankarmi, P., Maguire, J., Jin, X., 2023. Control of Behind-the-Meter Resources for Enhancing the Resilience of Residential Buildings, in: 2023 IEEE Power & Energy Society General Meeting (PESGM), pp. 1–5. URL: https://ieeexplore.ieee.org/abstract/document/10253443, doi:10.1109/PESGM52003.2023.10253443. iSSN: 1944-9933.

[63] Nagy, Z., Henze, G., Dey, S., Arroyo, J., Helsen, L., Zhang, X., Chen, B., Amasyali, K., Kurte, K., Zamzam, A., Zandi, H., Drgoňa, J., Quintana, M., McCullogh, S., Park, J.Y., Li, H., Hong, T., Brandi, S., Pinto, G., Capozzoli, A., Vrabie, D., Bergés, M., Nweye, K., Marzullo, T., Bernstein, A., 2023. Ten questions concerning reinforcement learning for building energy management. Building and Environment 241, 110435. URL: https://www.sciencedirect.com/science/article/pii/S0360132323004626, doi:10.1016/j.buildenv.2023.110435.

[64] Nagy, Z., Vázquez-Canteli, J.R., Dey, S., Henze, G., 2021. The citylearn challenge 2021, in: Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, ACM, New York, NY, USA. pp. 218–219. URL: https://dl.acm.org/doi/10.1145/3486611.3492226, doi:10.1145/3486611.3492226.

[65] Neukomm, M., Nubbe, V., Fares, R., 2019a. Grid-Interactive Efficient Buildings URL: https://www.osti.gov/biblio/1508212, doi:10.2172/1508212.

[66] Neukomm, M., Nubbe, V., Fares, R., 2019b. Grid-Interactive Efficient Buildings Technical Report Series: Overview of Research Challenges and Gaps. Technical Report NREL/TP-5500-75470; DOE/GO-102019-5227. National Renewable Energy Lab. (NREL), Golden, CO (United States). URL: https://www.osti.gov/biblio/1577966, doi:10.2172/1577966.

[67] Nweye, K., Kaspar, K., Buscemi, G., Pinto, G., Li, H., Hong, T., Ouf, M., Capozzoli, A., Nagy, Z., 2023a. A framework for the design of representative neighborhoods for energy flexibility assessment in CityLearn, in: Proceedings of building simulation 2023: 18th conference of IBPSA, IBPSA, Shanghai, China. pp. 3351–3358. URL: https://publications.ibpsa.org/conference/paper/?id=bs2023_1404, doi:https://doi.org/10.26868/25222708.2023.1404. iSSN: 2522-2708 tex.organisation: IBPSA.

[68] Nweye, K., Liu, B., Stone, P., Nagy, Z., 2022a. Real-world challenges for multi-agent reinforcement learning in grid-interactive buildings. Energy and AI 10, 100202. URL: https://www.sciencedirect.com/science/article/pii/S2666546822000489, doi:10.1016/j.egyai.2022.100202.

[69] Nweye, K., Nagy, Z., 2022. MARTINI: Smart meter driven estimation of HVAC schedules and energy savings based on Wi-Fi sensing and clustering. Applied Energy 316, 118980. URL: https://www.

sciencedirect.com/science/article/pii/S0306261922003890, doi:10.1016/j.apenergy.2022.118980.

[70] Nweye, K., Nagy, Z., 2024. Applications in CityLearn Gym Environment for Multi-Objective Control Benchmarking in Grid-Interactive Buildings and Districts, IBPSA-USA. pp. 417–428. URL: https://publications.ibpsa.org/conference/paper/?id=simbuild2024_2283.

[71] Nweye, K., Nagy, Z., Mohanty, S., Chakraborty, D., Sankaranarayanan, S., Hong, T., Dey, S., Henze, G., Drgona, J., Lin, F., Jiang, W., Zhang, H., Yi, Z., Zhang, J., Yang, C., Motoki, M., Khongnawang, S., Ibrahim, M., Zhumabekov, A., May, D., Yang, Z., Song, X., Zhang, H., Dong, X., Zheng, S., Bian, J., 2022b. The CityLearn Challenge 2022: Overview, Results, and Lessons Learned, in: Proceedings of the NeurIPS 2022 Competitions Track, PMLR. pp. 85–103. URL: https://proceedings.mlr.press/v220/nweye23a.html. iSSN: 2640-3498.

[72] Nweye, K., Sankaranarayanan, S., Nagy, Z., 2023b. MERLIN: Multi-agent offline and transfer learning for occupant-centric operation of grid-interactive communities. Applied Energy 346, 121323. URL: https://www.sciencedirect.com/science/article/pii/S0306261923006876, doi:10.1016/j.apenergy.2023.121323.

[73] Nweye, K.E., Wu, A., Park, H., Almilaify, Y., Nagy, Z., 2023c. CityLearn: A tutorial on reinforcement learning control for grid-interactive efficient buildings and communities, in: ICLR 2023 workshop on tackling climate change with machine learning. URL: https://www.climatechange.ai/papers/iclr2023/2.

[74] NYSERDA, 2022. NYSERDA RTEM Hackathon Demo Day. URL: https://be-exchange.org/nyserda-rtem-hackathon-demo-day/.

[75] Ouzeau, G., Soubeyroux, J.M., Schneider, M., Vautard, R., Planton, S., 2016. Heat waves analysis over France in present and future climate: Application of a new method on the EURO-CORDEX ensemble. Climate Services 4, 1–12. URL: https://www.sciencedirect.com/science/article/pii/S2405880716300309, doi:10.1016/j.cliser.2016.09.002.

[76] Panchabikesan, K., Ouf, M., Eicker, U., Newsham, G., Knudsen, H., 2021. Investigating thermostat setpoint preferences in Canadian households, in: Proceedings of the 17th IBPSA Conference, Bruges, Belgium. URL: https://publications.ibpsa.org/conference/paper/?id=bs2021_30433, doi:10.26868/25222708.2021.30433.

[77] Pickering, B., Choudhary, R., 2021. Quantifying resilience in energy systems with out-of-sample testing. Applied Energy 285, 116465. URL: https://www.sciencedirect.com/science/article/pii/S0306261921000313, doi:10.1016/j.apenergy.2021.116465.

[78] Pigman, M., Frick, N.M., Wilson, E., Parker, A., Present, E., 2022. End-use load profiles for the us building stock: Practical guidance on accessing and using the data. Technical Report. Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA (United States .... URL: https://emp.lbl.gov/publications/end-use-load-profiles-us-building-1.

[79] Pigott, A., Crozier, C., Baker, K., Nagy, Z., 2022. GridLearn: Multiagent reinforcement learning for grid-aware building energy management. Electric Power Systems Research 213, 108521. URL: https://www.sciencedirect.com/science/article/pii/S0378779622006320, doi:https://doi.org/10.1016/j.epsr.2022.108521.

[80] Pinto, G., Deltetto, D., Capozzoli, A., 2021. Data-driven district energy management with surrogate models and deep reinforcement learning. Applied Energy 304, 117642. URL: https://www.sciencedirect.com/science/article/pii/S0306261921010096, doi:10.1016/j.apenergy.2021.117642.

[81] Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., Dormann, N., 2021. Stable-Baselines3: Reliable Reinforcement Learning Implementations. Journal of Machine Learning Research 22, 1–8. URL: http://jmlr.org/papers/v22/20-1364.html.

[82] Reynders, G., Diriken, J., Saelens, D., 2017. Generic characterization method for energy flexibility: Applied to structural thermal storage in residential buildings. Applied Energy 198, 192–202. URL: https://www.sciencedirect.com/science/article/pii/S0306261917304555, doi:10.1016/j.apenergy.2017.04.061.

[83] Richard S. Sutton and Andrew G. Barto, 2018. Reinforcement Learning, Second Edition An Introduction.

[84] Rolnick, D., Donti, P.L., Kaack, L.H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A.S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A.S., Maharaj, T., Sherwin, E.D., Mukkavilli, S.K., Kording, K.P., Gomes, C.P., Ng, A.Y., Hassabis, D., Platt, J.C., Creutzig, F., Chayes, J., Bengio, Y., 2022. Tackling Climate Change with Machine Learning. ACM Computing Surveys 55, 42:1–42:96. URL: https://dl.acm.org/doi/10.1145/3485128, doi:10.1145/3485128.

[85] Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65. URL: https://www.sciencedirect.com/science/article/pii/0377042787901257, doi:https://doi.org/10.1016/0377-0427(87)90125-7.

[86] Scharnhorst, P., Schubnel, B., Fernández Bandera, C., Salom, J., Taddeo, P., Boegli, M., Gorecki, T., Stauffer, Y., Peppas, A., Politi, C., 2021. Energym: A building model library for controller benchmarking. Applied Sciences 11, 3518. Publisher: MDPI.

[87] Siano, P., 2014. Demand response and smart grids—A survey. Renewable and Sustainable Energy Reviews 30, 461–478. URL: https://www.sciencedirect.com/science/article/pii/S1364032113007211, doi:10.1016/j.rser.2013.10.022.

[88] Suberu, M.Y., Mustafa, M.W., Bashir, N., 2014. Energy storage systems for renewable energy power sector integration and mitigation of intermittency. Renewable and Sustainable Energy Reviews 35, 499–514. URL: https://www.sciencedirect.com/science/article/pii/S1364032114002366, doi:https://doi.org/10.1016/j.rser.2014.04.009.

[89] Towers, M., Terry, J.K., Kwiatkowski, A., Balis, J.U., Cola, G.d., Deleu, T., Goulão, M., Kallinteris, A., KG, A., Krimmel, M., Perez-Vicente, R., Pierré, A., Schulhoff, S., Tai, J.J., Shen, A.T.J., Younis, O.G., 2023. Gymnasium. URL: https://zenodo.org/record/8127025, doi:10.5281/zenodo.8127026.

[90] U.S. Department of Energy, 2021. Next-Generation Grid Technologies. Technical Report. U.S. Department of Energy. URL: https://www.energy.gov/sites/default/files/2022-05/Next%20Generation%20Grid%20Technologies%20Report%20051222.pdf.

[91] U.S. Energy Information Administration, . Table 11.3 Reliability Metrics Using Any Method of U.S. Distribution System by State, 2022 and 2021. URL: https://www.eia.gov/electricity/annual/html/epa_11_03.html.

[92] Varesi, K., Radan, A., Ghayebloo, A., Nikzad, M.R., 2016. An Efficient Methodology Proposed For Deciding About the Number of Battery Modules In Hybrid Electric Vehicles. Automatika 57, 99–108. URL: https://doi.org/10.7305/automatika.2016.07.693, doi:10.7305/automatika.2016.07.693. publisher: Taylor & Francis _eprint: https://doi.org/10.7305/automatika.2016.07.693.

[93] Vazquez-Canteli, J.R., Henze, G., Nagy, Z., 2020. MARLISA: Multi-Agent Reinforcement Learning with Iterative Sequential Action Selection for Load Shaping of Grid-Interactive Connected Buildings. BuildSys 2020 - Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation , 170–179doi:10.1145/3408308.3427604. iSBN: 9781450380614.

[94] Vigna, I., Lollini, R., Pernetti, R., 2021. Assessing the energy flexibility of building clusters under different forcing factors. Journal of Building Engineering 44, 102888. URL: https://www.sciencedirect.com/science/article/pii/S2352710221007464, doi:10.1016/j.jobe.2021.102888.

[95] Vázquez-Canteli, J.J.R., Kämpf, J., Henze, G., Nagy, Z., 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 , 356–357doi:10.1145/3360322.3360998. iSBN: 9781450370059.

[96] Vázquez-Canteli, J.R., Dey, S., Henze, G., Nagy, Z., 2020. The CityLearn Challenge 2020, in: Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, ACM, New York, NY, USA. pp. 320–321. URL: https://dl.acm.org/doi/10.1145/3408308.3431122, doi:10.1145/3408308.3431122.

[97] Vázquez-Canteli, J.R., Nagy, Z., 2019. Reinforcement learning for demand response: A review of algorithms and modeling techniques. Applied Energy 235, 1072–1089. URL: https://www.sciencedirect.com/science/article/pii/S0306261918317082, doi:10.1016/j.apenergy.2018.11.002.

[98] Wilson, E.J., 2017. ResStock - Targeting Energy and Cost Savings for U.S. Homes URL: https://www.osti.gov/biblio/1398250.

[99] Wilson, E.J.H., Parker, A., Fontanini, A., Present, E., Reyna, J.L., Adhikari, R., Bianchi, C., CaraDonna, C., Dahlhausen, M., Kim, J., LeBar, A., Liu, L., Praprost, M., Zhang, L., DeWitt, P., Merket, N., Speake, A., Hong, T., Li, H., Mims Frick, N., Wang, Z., Blair, A., Horsey, H., Roberts, D., Trenbath, K., Adekanye, O., Bonnema, E., El Kontar, R., Gonzalez, J., Horowitz, S., Jones, D., Muehleisen, R.T., Platthotam, S., Reynolds, M., Robertson, J., Sayers, K., Li, Q., 2022. End-Use Load Profiles for the U.S. Building Stock: Methodology and Results of Model Calibration, Validation, and Uncertainty Quantification URL: https://www.osti.gov/biblio/1854582, doi:10.2172/1854582.

[100] Wu, A., Nweye, K., Nagy, Z., 2023. Heterogeneous Multi-Agent Reinforcement Learning for Grid-Interactive Communities, in: Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, Association for Computing Machinery, New York, NY, USA. pp. 314–315. URL: https://doi.org/10.1145/3600100.3626276, doi:10.1145/3600100.3626276.

[101] Wölfle, D., Vishwanath, A., Schmeck, H., 2020. A Guide for the Design of Benchmark Environments for Building Energy Optimization, in: Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, Association for Computing Machinery, New York, NY, USA. pp. 220–229. URL: https://dl.acm.org/doi/10.1145/3408308.3427614, doi:10.1145/3408308.3427614.

[102] Yu, Z., Lu, F., Zou, Y., Yang, X., 2022. Quantifying energy flexibility of commuter plug-in electric vehicles within a residence–office coupling virtual microgrid. Part I: System configuration, quantification framework, and optimization model. Energy and Buildings 254, 111551. URL: https://www.sciencedirect.com/science/article/pii/S0378778821008355, doi:10.1016/j.enbuild.2021.111551.

[103] Zanocco, C., Flora, J., Rajagopal, R., Boudet, H., 2021. When the lights go out: Californians' experience with wildfire-related public safety power shutoffs increases intention to adopt solar and storage. Energy Research & Social Science 79, 102183. URL: https://www.sciencedirect.com/science/article/pii/S2214629621002760, doi:10.1016/j.erss.2021.102183.

[104] Zhan, S., Lei, Y., Chong, A., 2023. Comparing model predictive control and reinforcement learning for the optimal operation of building-PV-battery systems. E3S Web of Conferences 396, 04018. URL: https://www.e3s-conferences.org/articles/e3sconf/abs/2023/33/e3sconf_iaqvec2023_04018/e3sconf_iaqvec2023_04018.html, doi:10.1051/e3sconf/202339604018. publisher: EDP Sciences.

[105] Zhang, H., Wu, D., Boulet, B., 2022. MetaEMS: A Meta Reinforcement Learning-based Control Framework for Building Energy Management System URL: https://arxiv.org/abs/2210.12590, doi:10.48550/ARXIV.2210.12590.

A. Appendix

A.1. Distributed energy resources

This section describes the energy models for the DERs in CityLearn to include the heat pump (Appendix A.1.1), electric heater (Appendix A.1.2), TES (Appendix A.1.3), BESS (Appendix A.1.4), EV (Appendix A.1.5), and PV system (Appendix A.1.6). All DERs but the PV system are controllable.

A.1.1. Heat pump

CityLearn uses a heat pump model that is based on an ideal refrigeration cycle. Thus, the COP (Equations (A.1) and (A.2)) at any given time step, t, is a function of only outdoor air temperature, T^{out}, and target supply temperatures for cooling and heating modes ($T^{\text{cooling, supply}}$, $T^{\text{heating, supply}}$), where the supply temperature is constant at all time steps. The heat pump can be used in either control or ideal mode. In the control mode, the electricity consumption of the heat pump, $E_t^{\text{HP, control}}$, is a function of the control action a_{t-1}^{HP} where $a \in [0, 1]$ denotes the proportion of heat pump's nominal power, $P^{\text{HP, nominal}}$, made available (Equation (A.3)). The supplied cooling or heating energy for the controlled power is then defined by Equation (A.4) as the product of the COP, electricity consumption and technical efficiency, $\eta^{\text{HP, technical}}$.

$$\text{COP}_t^{\text{cooling}} = \frac{T^{\text{cooling, supply}}}{T_t^{\text{out}} - T^{\text{cooling, supply}}} \tag{A.1}$$

$$\text{COP}_t^{\text{heating}} = \frac{T^{\text{heating, supply}}}{T^{\text{heating, supply}} - T_t^{\text{out}}} \tag{A.2}$$

$$E_t^{\text{HP, control}} = a_{t-1}^{\text{HP}} \times P^{\text{HP, nominal}} \tag{A.3}$$

$$Q_t^{\text{HP, control}} = \eta^{\text{HP, technical}} \times \text{COP}_t^{\text{cooling | heating}} \times E_t^{\text{HP, control}} \tag{A.4}$$

In the ideal mode, the supplied cooling or heating energy is known a priori from BEPS software e.g. EnergyPlus [17], or real world measurements, and only the electricity consumption for the supplied energy needs to be calculated as defined in Equation (A.5).

$$E_t^{\text{HP, ideal}} = \frac{Q_t^{\text{HP, ideal}}}{\eta^{\text{technical}} \times \text{COP}_t^{\text{cooling | heating}}} \tag{A.5}$$

A.1.2. Electric heater

The electric heater is defined by Equations (A.6) to (A.8) where like the heat pump, can be used in control or ideal mode. Similarly, in the control mode, the electricity consumption of the electric heater is a function of the control action a_{t-1}^{EH} where $a \in [0, 1]$ denotes the proportion of the heater's nominal power, $P^{\text{EH, nominal}}$, that is made available (Equation (A.6)). The supplied heating energy is the product of the heater's technical efficiency, $\eta^{\text{EH, technical}}$, and controlled electricity consumption, $E_t^{\text{EH, control}}$ (Equation (A.7)). In the ideal case, $Q_t^{\text{EH, ideal}}$ is known and $E_t^{\text{EH, ideal}}$ is evaluated as Equation (A.8).

$$E_t^{\text{EH, control}} = a_{t-1}^{\text{EH}} \times P^{\text{EH, nominal}} \tag{A.6}$$

$$Q_t^{\text{EH, control}} = \eta^{\text{EH, technical}} \times E_t^{\text{EH, control}} \tag{A.7}$$

$$E_t^{\text{EH, ideal}} = \frac{Q_t^{\text{EH, ideal}}}{\eta^{\text{EH, technical}}} \tag{A.8}$$

A.1.3. Thermal energy storage

Thermal loads are shifted by storing (charging) and releasing (discharging) chilled or hot water energy in a TES. The stored energy at any time step, Q_t^{tes}, is a piecewise function (Equation (A.9)) driven by the control action a_{t-1}^{TES}, where $a \in [-1, 1]$ prescribes the proportion of the TES capacity, C^{TES}, to be charged ($a > 0$) or discharged ($a < 0$). The energy stored after charging is defined by Equation (A.10) as the minimum of the capacity and an expected energy. The expected energy is the sum of the initial energy after losses, $Q_{t-1}^{\text{TES}} \times (1 - \theta^{\text{TES}})$, where θ^{TES} is the thermal loss coefficient, and the product of the action, capacity and round-trip efficiency, $\eta^{\text{TES, round-trip}}$. The round-trip efficiency defined in Equation (A.11) is the square root of the technical efficiency, $\eta^{\text{TES, technical}}$.

$$Q_t^{\text{TES}} = \begin{cases} Equation\ (A.10) & \text{if } a_{t-1}^{\text{TES}} > 0.0 \\ Equation\ (A.12) & \text{if } a_{t-1}^{\text{TES}} < 0.0 \\ 0.0 & \text{otherwise} \end{cases} \tag{A.9}$$

$$Q_t^{\text{TES, +}} = \min\left(C^{\text{TES}}, Q_{t-1}^{\text{TES}} \times \left(1 - \theta^{\text{TES}}\right) + a_{t-1}^{\text{TES}} \times C^{\text{TES}} \times \eta^{\text{TES, round-trip}} \right) \tag{A.10}$$

$$\eta^{\text{TES, round-trip}} = \sqrt{\eta^{\text{TES, technical}}} \tag{A.11}$$

The stored energy after discharging the TES is defined in Equation (A.12) as the sum of the initial energy after losses and quotient of the energy equivalent of the control action ($a_{t-1}^{\text{TES}} \times C^{\text{TES}}$) and the round-trip efficiency. A lower limit of 0 is imposed to satisfy the energy balance in the case of a completely discharged TES.

$$Q_t^{\text{TES, -}} = \max\left(0.0, Q_{t-1}^{\text{TES}} \times \left(1 - \theta^{\text{TES}}\right) + a_{t-1}^{\text{TES}} \times C^{\text{TES}} \div \eta^{\text{TES, round-trip}}\right) \quad \text{(A.12)}$$

Consequently, the SoC (Equation (A.13)) $\in [0, 1]$ is the ratio of the stored energy calculated in Equation (A.9) and the capacity.

$$\text{SoC}_t^{\text{TES}} = \frac{Q_t^{\text{TES}}}{C^{\text{TES}}} \quad \text{(A.13)}$$

Finally, the energy balance i.e., energy supplied to the TES by an electric device e.g., heat pump or electric heater before adjustment for efficiency or energy supplied from the TES to a building after adjustment for efficiency is defined by Equation (A.14).

$$Q_t^{\text{TES, balance}} = \left(Q_t^{\text{TES}} - Q_{t-1}^{\text{TES}} \times \left(1 - \theta^{\text{TES}}\right)\right) \times \begin{cases} 1 \div \eta^{\text{TES, round-trip}} & \text{if } a_{t-1} > 0.0 \\ \eta^{\text{TES, round-trip}} & \text{if } a_{t-1} < 0.0 \\ 0.0 & \text{otherwise} \end{cases} \quad \text{(A.14)}$$

A.1.4. Battery energy storage system

The BESS model builds up on the TES model but has a time-dependent capacity, C^{BESS}, round-trip efficiency, $\eta^{\text{BESS, round-trip}}$, and maximum input as well as output power, $P^{\text{BESS, max}}$. The stored energy at any time step, Q_t^{bess}, is a piecewise function (Equation (A.15)) driven by the control action a_{t-1}^{BESS}, where $a \in [-1, 1]$ prescribes the proportion of the maximum capacity (before degradation), C_0^{BESS}, to be charged ($a > 0$) or discharged ($a < 0$). The stored energy after charging is defined by Equation (A.16) as the minimum of the degraded capacity, C_t^{BESS} (Equation (A.21)), and an expected energy. The expected energy is the sum of the initial energy after losses, $Q_{t-1}^{\text{BESS}} \times (1 - \theta^{\text{BESS}})$, and the energy to be added, where θ^{BESS} is the thermal loss coefficient. This added energy is the product of the power-dependent round-trip efficiency, $\eta^{\text{BESS, round-trip}}$, and the minimum of two quantities: the energy equivalent of the control action ($a_{t-1}^{\text{BESS}} \times P^{\text{BESS, nominal}}$) and the SoC-dependent maximum input and output power, $P_t^{\text{BESS, max}}$.

$$Q_t^{\text{BESS}} = \begin{cases} \textit{Equation (A.16)} & \text{if } a_{t-1}^{\text{BESS}} > 0.0 \\ \textit{Equation (A.17)} & \text{if } a_{t-1}^{\text{BESS}} < 0.0 \\ 0.0 & \text{otherwise} \end{cases} \quad \text{(A.15)}$$

$$Q_t^{\text{BESS, +}} = \min\left(C_t^{\text{BESS}}, Q_{t-1}^{\text{BESS}} \times \left(1 - \theta^{\text{BESS}}\right) + \min\left(a_{t-1}^{\text{BESS}} \times C_0^{\text{BESS}}, P_t^{\text{BESS}}\right) \times \eta^{\text{BESS, round-trip}}\right) \quad \text{(A.16)}$$

The stored energy after discharging the BESS is defined by Equation (A.17) as the sum of the initial energy after losses and quotient of the energy equivalent of the control action (after adjustment for maximum output power, P_t^{BESS}) and the round-trip efficiency. The stored energy is limited to a depth-of-discharge (DoD) such that the BESS is never completely drained for DoD > 0. Equation (A.18) is the BESS SoC as a function of the stored energy and C_0^{BESS}.

$$Q_t^{\text{BESS, -}} = \min\left(C_0^{\text{BESS}} \times \text{DoD}^{\text{BESS}}, Q_{t-1}^{\text{BESS}} \times \left(1 - \theta^{\text{BESS}}\right) + \min\left(a_{t-1}^{\text{BESS}} \times C_0^{\text{BESS}}, -P_t^{\text{BESS}}\right) \div \eta^{\text{BESS, round-trip}}\right) \quad \text{(A.17)}$$

$$\text{SoC}_t^{\text{BESS}} = \frac{Q_t^{\text{BESS}}}{C_0^{\text{BESS}}} \quad \text{(A.18)}$$

The energy balance, Equation (A.19), is same as that of the TES except that the round-trip efficiency is time-dependent. The energy balance is also the electricity consumed when charging or avoided consumption when discharging (Equation (A.20)).

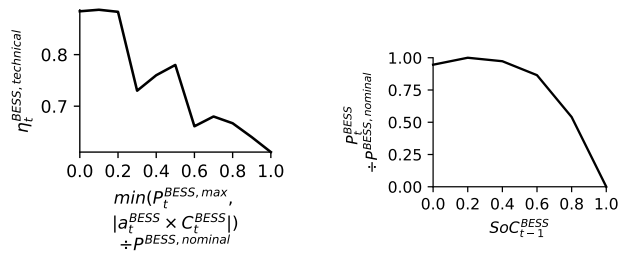
$$Q_t^{\text{BESS, balance}} = \left(Q_t^{\text{BESS}} - Q_{t-1}^{\text{BESS}} \times \left(1 - \theta^{\text{BESS}}\right)\right) \times \begin{cases} 1 \div \eta_t^{\text{BESS, round-trip}} & \text{if } a_{t-1}^{\text{BESS}} > 0.0 \\ \eta_t^{\text{BESS, round-trip}} & \text{if } a_{t-1}^{\text{BESS}} < 0.0 \\ 0.0 & \text{otherwise} \end{cases} \quad \text{(A.19)}$$

$$E_t^{\text{BESS}} = Q_t^{\text{BESS, balance}} \quad \text{(A.20)}$$

The time-dependency of the capacity is as a result of degradation from charging and discharging cycles. It is defined in Equation (A.21) as the difference between the capacity at the previous time step, and the loss quantity. This quantity is a function of the capacity loss coefficient, ϕ, maximum capacity, energy balance at the current time step, and capacity at the previous time step.

$$C_t^{\text{BESS}} = C_{t-1}^{\text{BESS}} - \frac{\phi \times C_0^{\text{BESS}} \times \left|Q_t^{\text{BESS, balance}}\right|}{2 \times C_{t-1}^{\text{BESS}}} \quad \text{(A.21)}$$

The BESS round-trip efficiency, $\eta_t^{\text{BESS, round-trip}}$, (Equation (A.22)) is similar to that of the TES but with the



(a) Technical efficiency as a function of controlled proportion of nominal power (Equation (A.23)).

(b) Proportion of nominal power used to define the maximum input and output power as a function of SoC (Equation (A.24)).

Figure A.1: Example curves defining BESS technical efficiency, $\eta_t^{\text{BESS, technical}}$, and maximum input and output power, $P_t^{\text{BESS, max}}$.

technical efficiency, $\eta_t^{\text{BESS, technical}}$, as a function of a power-efficiency curve (Equation (A.23)), where for a given controlled proportion of the nominal power, $P^{\text{BESS, nominal}}$, there is a limitation on the battery's efficiency. An example of this function from [50] is shown in Figure A.1a for a BESS subjected to continuous charge and discharge cycles.

$$\eta_t^{\text{BESS, round-trip}} = \sqrt{\eta_t^{\text{BESS, technical}}} \tag{A.22}$$

$$\eta_t^{\text{BESS, technical}} = f\left(a_t^{\text{BESS}}, C_0^{\text{BESS}}, P_t^{\text{BESS, max}}, P^{\text{BESS, nominal}}\right) \tag{A.23}$$

Equation (A.24) defines the maximum input and output power $P_t^{\text{BESS, max}}$ as the product of the nominal power, $P^{\text{BESS, nominal}}$, and SoC-power curve. The curve's dependent variable is the proportion of available nominal power, $P^{\text{BESS, nominal}}$, at any SoC, $\text{SoC}_{t-1}^{\text{BESS}}$ value.

$$P_t^{\text{BESS, max}} = P^{\text{BESS, nominal}} \times f\left(\text{SoC}_{t-1}^{\text{BESS}}\right) \tag{A.24}$$

Figure A.1b shows an example of an SoC-dependent power curve from [92] where the BESS charges faster and close to the nominal power for lower SoC ($\text{SoC}_{t-1} \leq 0.6$), while the charge rate decreases to zero at higher SoC. This is consistent with the charging behaviour of a BESS where it takes longer to charge per unit SoC after exceeding an SoC threshold of ≈ 0.8 [47].

A.1.5. Electric vehicle

The energy management of EVs is designed to allow the simulation of three distinct modes: (1) V2G; (2) G2V; (3) no control (i.e., where EVs act as a load without any possible control over their charging). For that purpose, the simulation has three fundamental parts: the Electric Vehicle Chargers (EVCs), which serve as a connection between a building and an EV, theEV model itself, which acts as a DER, and an input schedule, which dictates the plug in/out energy flexibility routine for each EV and introduces formulation for energy flexibility and simulation modes. For a detailed explanation of the EV modes and simulation please refer to [27].

Electric vehicle charger: The EVC serves as an interface between the buildings in the simulation environment and the EVs, and is the component where control is applied.

Just like a house, an office or any building can have multiple installed chargers in the real world, a single building in the simulation can have more than one charger simulated. Each charger is assigned a unique identifier $\text{EVC}_{b_n_p}$, where b is the building index where the charger is installed, n is the EVC index in building b, and p is the plug index in EVC n). This facilitates the appropriate linkage between the EVC and the EVs during simulation.

The model supports the specification of distinct power output levels for charging and discharging modes. When an EV is plugged in, the electricity consumption of the EVC, $E_t^{\text{EVC}_{b_n_p}}$, is a function of the control action $a_{t-1}^{\text{EVC}_{b_n_p}}$ where $a \in [-1, 1]$ for V2G $a \in [0, 1]$ for G2V and no-control modes. It denotes the proportion of EVC nominal power for charging, $P^{\text{EVC, nominal, charging}}$, when if $a_{t-1}^{\text{EVC}_{b_n_p}} \geq 0$ and the proportion of the EVC nominal power for discharging $P^{\text{EVC, nominal, discharging}}$ when if $a_{t-1}^{\text{EVC}_{b_n_p}} < 0$ (Equation (A.25)). The supplied charging energy from the EVC to the connected EV BESS, for a controlled input power is then defined by Equation (A.26) as the product of the electricity consumption $E_t^{\text{EVC}_{b_n_p}}$ and technical efficiency $\eta^{\text{EVC, technical}}$.

$$E_t^{\text{EVC}_{b_n_p}} = a_{t-1}^{\text{EVC}_{b_n_p}} \times P^{\text{EVC}_{b_n_p}\text{,nominal, charging | discharging}} \tag{A.25}$$

$$Q_t^{\text{EVC}_{b_n_p}} = \eta^{\text{EVC}_{b_n_p}\text{,technical}} \times E_t^{\text{EVC}_{b_n_p}} \tag{A.26}$$

Electric vehicle: The EV uses the same BESS model as the stationary BESS (Appendix A.1.4). The EVs in the simulation can be connected to an EVC (Appendix A.1.5) and consume energy (in all modes) and discharge back to the grid (in V2G mode). An EVs will connect and disconnect from the chargers as per the modes pre-simulated file described in Appendix A.1.5.

Modes: In the simulation modes for EV energy management, we establish that the control over EV charging is exercised exclusively when the EVs are plugged into a building's EVC. Control agents regulate the power of the EVCs, which in turn charges or discharges the connected EVs. It is important to note that these agents will not have an influence on the EV arrival and departure schedule, which are pre-simulated and made available to CityLearn through a time series file.

A.1.6. Photovoltaic system

The PV model makes use of pre-simulated inverter output power per unit of installed PV capacity, that is reported as a time series, P_t^{I}. For PV generation estimates, the System Advisory Model (SAM) [10] is used to generate location specific data. The PV generation is thus, the negative product of a user-defined nominal power, $P^{\mathrm{PV,\,nominal}}$ and P_t^{I} (Equation (A.27)).

$$E_t^{\mathrm{PV}} = -P^{\mathrm{PV,\,nominal}} \times P_t^{\mathrm{I}} \tag{A.27}$$

A.2. Thermal dynamics model

A.2.1. Fitting data generation

To generate the LSTM fitting data, we first build an EnergyPlus model of the building or source a representative model from open-source datasets like [99]. The energy model is paired with a weather file for the desired building location and simulated multiple times to obtain ideal cooling and heating loads as well as the effect that supplying thermal energy below or above the ideal loads has on the indoor dry-bulb temperature. The first simulation (*energyplus.simulation.1*) uses the as-provided mechanical HVAC system in the energy model as a reference point for subsequent simulations.

In the second simulation (*energyplus.simulation.2*), the mechanical HVAC system is replaced with an *HVACTemplate:Zone:IdealLoadsAirSystem* EnergyPlus object, which is an artificial HVAC system that completely supplies cooling or heating air to a building in sufficient quantity and 100% efficiency to meet the building load in full (ideal load). We run the ideal load simulation to determine the ideal cooling and heating loads that satisfy the indoor dry-bulb temperature setpoint, and the results from this simulation is validated against that of *energyplus.simulation.1*. The validation ensures that the as-provided energy model's integrity is not tampered with when replacing mechanical systems with an ideal load system. Our check for integrity is that for the ideal load case, cooling and heating energy delivered to the building is equal or slightly less than that for the mechanical case. Indoor dry-bulb temperature in the building using the ideal load system should be similar to that from using the mechanical system.

The remaining simulations make up the LSTM training dataset where we replace the ideal load system with an *OtherEquipment* EnergyPlus object which is a generic internal heat gain and loss equipment to simulate undercooling, under-heating, over-cooling and overheating with respect to the building's ideal loads. The first of these simulations (*energyplus.simulation.3*) supplies the ideal load from *energyplus.simulation.2* to validate the operation of the generic equipment compared to the ideal load system, as the effect on temperature should closely match. The next simulation (*energyplus.simulation.4*) free-floats the temperature by supplying neither cooling nor heating energy to the building while the remaining simulations (*energyplus.simulation.5-n*) have the ideal load varied using the inequality, $A \cdot I_t \leq I_t \leq B \cdot I_t$ with a probability, $p^{L_t \neq I_t}$, that the ideal load is not satisfied. A and B are coefficients that define the variation range where $p^{L_t \neq I_t} = 0.6$, $A = 0.3$ and $B = 1.7$ have been used in the literature [18].

Finally, we query the results of the generic internal heat gain/loss equipment simulations to extract features that make up the LSTM fitting data, including the following variables: direct and diffuse solar irradiance, outdoor dry-bulb temperature, occupant count, cooling and heating loads, month, day-of-week, hour, and indoor dry-bulb temperature (target variable). Note that while we use EnergyPlus for our simulations, these data can be generated with any other BEPS software.

A.2.2. Training and testing

We split the generated data into training, validation and testing sets. The method of splitting ensures that each of the simulations with reference *energyplus.simulation.3-n* are equally represented in each set. The sliding window method is used in training where we partition the training data into sequences of l time steps. $l = 12$ has been used in practice [80]. Each sequence represents a window of independent features. Each sequence is then paired with its corresponding target, forming input-output pairs. By sliding the window along the time axis, multiple input-output pairs are generated, covering the entirety of the training data. Through this iterative process, the LSTM model learns to predict the indoor dry-bulb temperature at a given time step, t, leveraging the lagged input sequence leading up to the previous $t-l$ time steps. This enables the model to effectively capture temporal dynamics based on past observations.

We test the model in a closed loop, whereby the model's predictions are used as input for subsequent predictions. The model's performance is evaluated on its minimization of its root mean square error (RMSE) (Equation (A.28)) and mean absolute percentage error (MAPE) (Equation (A.29)) where n is the number of time steps in the testing period, T_t^{in} is the true temperature at time step t and $\hat{T}_t^{\mathrm{in}}$ is the predicted temperature.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=0}^{n-1}\left(T_t^{\mathrm{in}} - \hat{T}_t^{\mathrm{in}}\right)^2}{n}} \tag{A.28}$$

$$\mathrm{MAPE} = \frac{100}{n} \times \sum_{t=0}^{n-1}\left|\frac{T_t^{\mathrm{in}} - \hat{T}_t^{\mathrm{in}}}{T_t^{\mathrm{in}}}\right| \tag{A.29}$$

Figure A.2 shows the RMSE and MAPE distribution for LSTM models included in the Alameda Co., Travis Co., and Chittenden Co. datasets (see Table A.6 for description of datasets). These errors are an aggregate of the ideal load scenario using the generic internal heat gain and loss equipment (energyplus.simulation.3) and partial load scenarios (energyplus.simulation.5-n). Lower median RMSEs in all three datasets are observed during the summer while the higher median RMSEs are observed during the shoulder

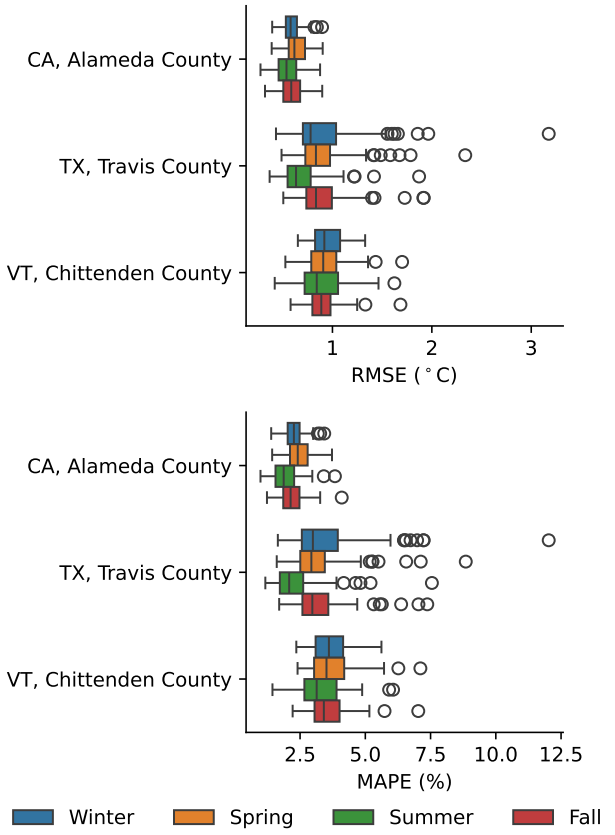


Figure A.2: Seasonal RMSE and MAPE for LSTM models included in the Alameda Co., Travis Co., and Chittenden Co. datasets (see Table A.6 for description of datasets). The errors are an aggregate of the ideal load scenario using the generic internal heat gain and loss equipment (energyplus.simulation.3) and partial load scenarios (energyplus.simulation.5-n). Winter season includes December, January-February, spring season includes March-May, summer season includes June-August, and fall season includes September-November.

spring and fall seasons. Nevertheless, the RMSE for the 75th percentile are generally below 1.5°C for the three counties while Chittenden County's 75th percentile RMSE is 1.0°C on average. These observations suggest that the LSTM models in the three datasets will typically predict within 1.0°C of the actual indoor dry-bulb temperature for all seasons. There are 7-27/100, and 11-18/47 models in Travis and Chittenden Counties respectively where their RMSEs are $\geq$ 1.0°C during one or more seasons. For such models, the user may opt to exclude their respective buildings from the environment setup.

The MAPE distribution is similar to the RMSE distribution but shows that the predicted temperature values are within 10% of the actual values on average and are as low as 1.0%-2.4% for any given season, which we deem sufficient for our application.

We show in Figure A.3, the comparison between actual and predicted indoor dry-bulb temperature for the building with the least RMSE in Figure A.2 for each county, over a one week period, starting from the beginning of each season. The time series are shown only for the ideal load scenario using the generic internal heat gain and loss equipment (energyplus.simulation.3). The profiles are resultant of real-world dry-bulb temperature setpoint profiles from the ecobee dataset [58] that were assigned to each building using the methodology described in [67]. The profiles differ from day-to-day and across seasons, yet the LSTM models are able to predict the unique profiles across all four seasons, while staying within the 0.5°C error band in most cases and deviating into the 1.0°C error band for a few time steps. In Chittenden County, the actual temperature remains approximately constant for extended periods however, the model is unable to predict the constant-value trend but instead, predicts temperature change in the order of less than 0.1°C. Although the predicted change magnitude is small, it occurs long enough for the predicted temperature to deviate by up to 1.0°C from the actual temperature. Only a few time steps deviate past 1.0°C for the shown snapshot. The models predict moderately well the change points in temperature, but may slightly underestimate higher temperatures, and overestimate lower temperatures that last for a few time steps.

A.3. Observation space

There are over 50 observable continuous states in CityLearn as summarized in Table A.1 and categorized as one of the following: calendar, weather, environmental, cost, load, consumption, flexibility, efficiency, comfort, occupancy, and resilience observations. The weather and cost observations include 6, 12, and 24 hour forecasts that can inform future self-generation capability, loads and electric device efficiency. In ideal load control, the load and consumption observations inform the building loads and the energy used to satisfy the loads whereas in a controlled load scenario, where the load observations are calculated, the load and consumption observations inform the impact of the agent(s) control action(s) on energy use. Given that simultaneous cooling and heating is not possible in a building, the `hvac_mode` is used to determine which of these two thermal modes a building is operating in at any given time step. The thermal loads can also be unmet and the indoor dry-bulb temperature allowed to free-float when `hvac_mode=off`. The flexibility observations provide the control agent with the building's capacity for upward or downward energy flexibility. There are as many EV-related observations as there are EV chargers in a building. The efficiency observations are the COP of the cooling, heating and DHW devices when they are heat pump type otherwise, it is their technical efficiency e.g., electric heater. The comfort observations provide information on occupant thermal and humidity comfort as well as occupant thermostat interaction. `occupant_count` specifies if the building is occupied or not and the number of people during occupancy. The `power_outage` observation is the power outage signal discussed in Section 2.2.

Observations are (1) either shared i.e., equal value in all buildings or building specific, (2) either control action

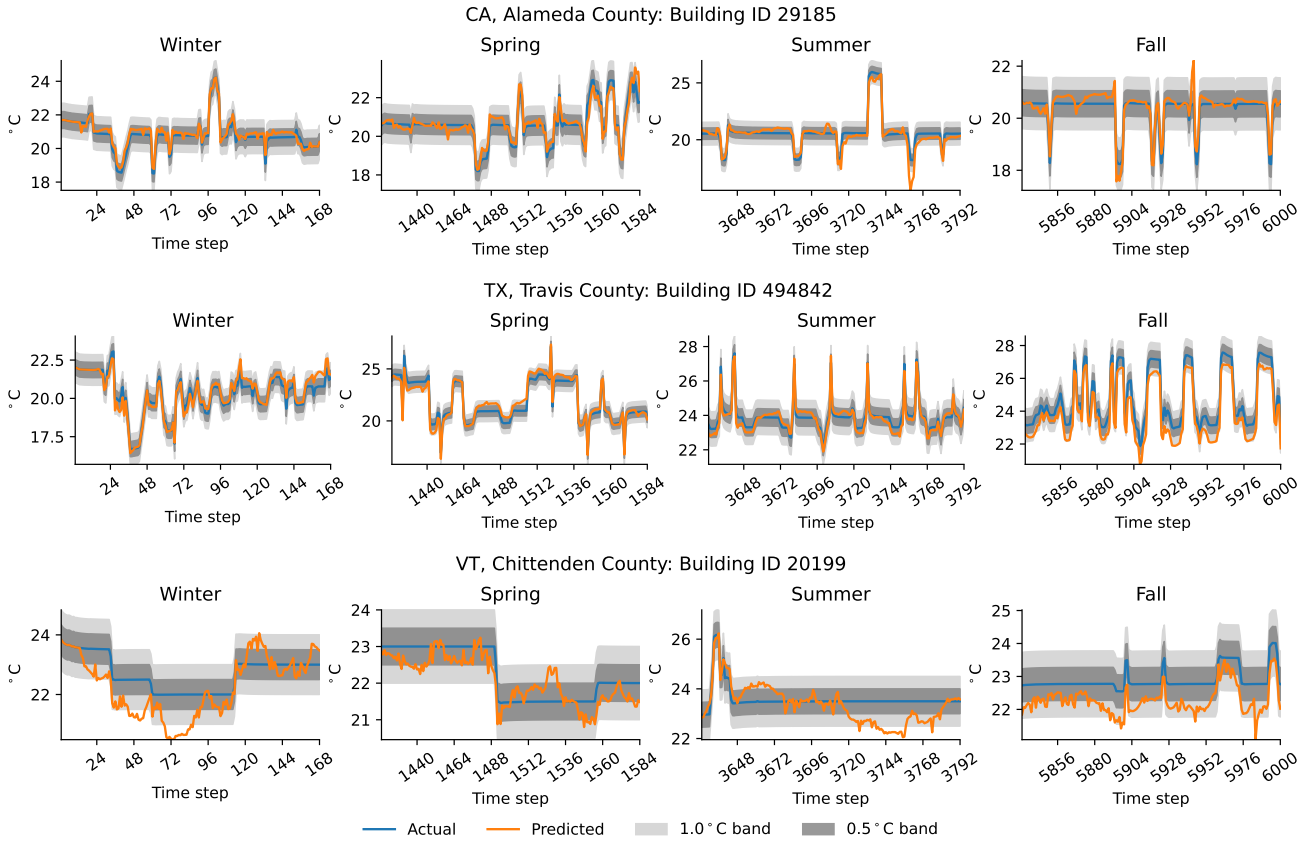


Figure A.3: Comparison between actual and predicted indoor dry-bulb temperature for the building with the least RMSE in Figure A.2 for each county, over a one week period starting from the beginning of each season. The shaded regions show either a 0.5°C or 1.0°C error band around the actual temperature. The time series are shown only for the ideal load scenario using the generic internal heat gain and loss equipment (energyplus.simulation.3) and the building IDs match the same building IDs in the *resstock–amy2018–2021–release–1* version of [99].

dependent or control agnostic, and (3) either statically defined in a time series file or calculated during runtime. The inclusion of an observation in a control problem's observation space is dependent on the control problem formulation including the controlled DERs and control objective and is customizable by the user.

A.4. Action space

Table A.2 summarizes the continuous action space where there are five ESS-related actions controlling the proportion of storage capacity to be charged or discharged, and two HVAC electric device actions controlling the proportion of nominal power to be supplied. There are as many `electric_vehicle_storage` actions as there are EV chargers in a building. Depending on the control problem and the DERs in a building, some actions may be excluded in the action space.

A.5. Reward functions

Table A.3 summarizes four reward functions defined in CityLearn for RLC although, other user-defined reward functions are possible depending on the control problem formulation. Equations (A.31) to (A.34) define these reward functions at the building-level, r_t^{building}, and for centralized control architecture, the district reward, r_t^{district}, is the sum of the building-level rewards (Equation (A.30)).

The reward function guides the learning process in RLC and is akin to the objective function in MPC.

$$r_t^{\text{district}} = \sum_{i=0}^{b-1} r_t^{\text{building } i} \tag{A.30}$$

A.6. Control algorithms

Table A.4 summarizes the control agent algorithms that are defined in CityLearn as well as supported standardized RLC algorithm libraries. The Baseline agent is used to establish a BAU scenario without DER control while the Random agent takes arbitrary actions. All RBC agents in CityLearn are descendants of the Hour-RBC agent that prescribes actions with respect to the hour-of-day. CityLearn provides three main types of model-free RLC agents: Tabular Q-Learning [83] for discretized observation and action spaces, SAC [34] and MARLISA [93] for continuous observation and action spaces. The SAC algorithm is adapted for either single-agent or independent multi-agent configurations

Table A.1
Observation space.

Name	Unit	Shared	Control-dependent	Source
Calendar				
month	-	✓		building_id.csv
day_type	-	✓		building_id.csv
hour	-	✓		building_id.csv
daylight_savings_status	true/false	✓		building_id.csv
Weather				
outdoor_dry_bulb_temperature$_{t+(0,6,12,24)}$	°C	✓		weather.csv
outdoor_relative_humidity$_{t+(0,6,12,24)}$	%	✓		weather.csv
diffuse_solar_irradiance$_{t+(0,6,12,24)}$	W/m^2	✓		weather.csv
direct_solar_irradiance$_{t+(0,6,12,24)}$	W/m^2	✓		weather.csv
Environmental				
carbon_intensity	kgCO_2e/kWh	✓		carbon_intensity.csv
Cost				
electricity_pricing$_{t+(0,6,12,24)}$	$/kWh			pricing.csv
Load				
cooling_demand	kWh		✓	building_id.csv/calculated
heating_demand	kWh		✓	building_id.csv/calculated
dhw_demand	kWh			building_id.csv
Consumption				
cooling_electricity_consumption	kWh		✓	calculated
heating_electricity_consumption	kWh		✓	calculated
dhw_electricity_consumption	kWh		✓	calculated
electrical_storage_electricity_consumption	kWh		✓	calculated
non_shiftable_load	kWh			building_id.csv
net_electricity_consumption	kWh		✓	calculated
hvac_mode	cooling/heating/off			building_id.csv
Flexibility				
cooling_storage_soc	-		✓	calculated
heating_storage_soc	-		✓	calculated
dhw_storage_soc	-		✓	calculated
electrical_storage_soc	-		✓	calculated
electrical_vehicle_soc	-		✓	calculated
electrical_vehicle_estimated_arrival_soc	-			electric_vehicle_id.csv
electrical_vehicle_required_departure_soc	-			electric_vehicle_id.csv
electrical_vehicle_estimated_arrival_time	-			electric_vehicle_id.csv
electrical_vehicle_estimated_departure_time	-			electric_vehicle_id.csv
electrical_vehicle_charger_state	-			electric_vehicle_id.csv
solar_generation	kWh			calculated
Efficiency				
cooling_device_efficiency	-			calculated
heating_device_efficiency	-			calculated
dhw_device_efficiency	-			calculated
Comfort				
indoor_dry_bulb_temperature	°C		✓	building_id.csv/calculated
indoor_dry_bulb_temperature_set_point	°C		✓	building_id.csv/calculated
indoor_dry_bulb_temperature_delta	°C		✓	calculated
indoor_dry_bulb_temperature_set_point_override_delta	°C		✓	calculated
indoor_relative_humidity	%			building_id.csv
Occupancy				
occupant_count	people			building_id.csv
Resilience				
power_outage	true/false			building_id.csv/calculated

while the MARLISA algorithm is for purely multi-agent configurations with the ability to share information amongst agents. SAC-RBC and MARLISA-RBC are sub-types of the SAC and MARLISA algorithms that make use of an Hour-RBC for offline training to initialize their learning [68]. For a wider range of algorithm options, we provide wrappers for interfacing with algorithms implemented in Stable-Baselines3 [81] and RLlib [55] libraries.

Examples of how to use these control algorithms are provided in our web-based documentation[3].

A.7. Energy flexibility key performance indicators

Equation (A.35) is a thermal comfort KPI that quantifies the proportion of total simulation time steps, n, where the difference between a building's indoor dry-bulb temperature, T_t^{in}, and set point, T_t^{spt}, falls outside a comfort band, T^{Δ}.

[3] https://github.com/intelligent-environments-lab/CityLearn/blob/master/examples/quickstart.ipynb

Table A.2
Action space.

Name	a_t range	Description
Energy storage system		
`cooling_storage`	[-1, 1]	Proportion of `cooling_storage` capacity to be charged ($a_t > 0$) or discharged ($a_t < 0$).
`heating_storage`	[-1, 1]	Proportion of `heating_storage` capacity to be charged ($a_t > 0$) or discharged ($a_t < 0$).
`dhw_storage`	[-1, 1]	Proportion of `dhw_storage` capacity to be charged ($a_t > 0$) or discharged ($a_t < 0$).
`electrical_storage`	[-1, 1]	Proportion of `electrical_storage` capacity to be charged ($a_t > 0$) or discharged ($a_t < 0$).
`electric_vehicle_storage`	[-1, 1]	Proportion of `electric_vehicle_storage` capacity to be charged ($a_t > 0$) or discharged ($a_t < 0$).
Electric device		
`cooling_device`	[0, 1]	Proportion of `cooling_device` nominal power to be supplied.
`heating_device`	[0, 1]	Proportion of `heating_device` nominal power to be supplied.

Equations (A.36) and (A.37) are similar to Equation (A.35) but consider when the difference is either below or above T^{Δ}. Equations (A.38) to (A.43) report the descriptive statistics of $T_t^{\text{in}} - T_t^{\text{spt}}$ for either cold discomfort ($T_t^{\text{in}} < T_t^{\text{spt}}$) or hot discomfort ($T_t^{\text{in}} > T_t^{\text{spt}}$) at the building level.

The energy KPIs are defined in Equations (A.44) to (A.49) where Equation (A.44) defines a building's total electricity consumption, $\max\left(E_t^{\text{building, net}}, 0\right)$, from the grid (net import) whereas Equation (A.45) is a building's total net electricity consumption, $E_t^{\text{building, net}}$, (sum of import and export). Equation (A.46) is a district level KPI quantifying an entire district's average daily peak net electricity consumption, $E^{\text{district, net}}$, where h is the hours per day, and d is the number of days. The peak district net electricity consumption at any time step is quantified by Equation (A.47). Ramping is defined in Equation (A.48) as the positive difference in $E^{\text{district, net}}$ between two consecutive time steps. It is a district level KPI that quantifies the smoothness of the district electricity consumption during the up-ramp that typically happens in the early evening when renewable power generation depletes and fossil-fueled power plants come online. Equation (A.49) is also a district level KPI that evaluates the average ratio of daily average and peak net electricity consumption, termed, load factor. The load factor is the efficiency of electricity consumption and is bound between 0 (highly inefficient) and 1 (highly efficient) thus, the goal is to maximize the load factor or minimize 1 − load factor. These district-level energy KPIs are evaluated at the building level by substituting $E_t^{\text{district, net}}$ with $E_t^{\text{building, net}}$ in their equations.

Equation (A.50) and Equation (A.51) evaluate the cost and CO_2e emissions from net import at the building level using time-dependent electricity rate, R_t, and emissions rate, G_t.

Equations (A.52) and (A.53) evaluate the resilience provided by a control algorithm during power outages i.e., $O_t > 0$, from thermal comfort and energy perspectives. Thermal resilience (Equation (A.52)) is similar to Equation (A.35) but is evaluated only during power outage events. Total unserved energy (Equation (A.53)) quantifies the difference between the energy needed to satisfy building loads and the actual energy supplied to the building by the electric devices and ESSs.

Table A.3
Reinforcement learning reward functions in CityLearn.

Name	Description	r_t^{building}	
Electricity consumption	Encourages reduced consumption from the grid and does not reward nor penalize net export i.e., ($E_t^{\text{building, net}} < 0$). Exponent, a, used to impose greater penalty for $E_t^{\text{building, net}} >> 0$.	$-\max\left(E_t^{\text{building, net}}, 0\right)^a$	(A.31)
MARL	Multi-agent reward used to share information between the agents and rewards them for reducing the district net electricity consumption, $E_t^{\text{district, net}}$ [93]. For a building, the value is (1) negative if the building is consuming electricity from the grid ($E_t^{\text{building, net}} > 0$) while the district is also consuming electricity from the grid ($E_t^{\text{district, net}} > 0$), (2) positive when the building generates more electricity than it consumes ($E_t^{\text{building, net}} < 0$) though the district is consuming electricity from the main grid ($E_t^{\text{district, net}} > 0$), since the building is contributing to making the district self-sufficient. If the district is self-sufficient i.e., $E_t^{\text{district, net}} < 0$, the reward is 0.	$0.01 \times \left(E_t^{\text{building, net}^2}\right) \times \max\left(E_t^{\text{district, net}}, 0\right) \times \begin{cases} 1 & \text{if } E_t^{\text{building, net}} < 0 \\ -1 & \text{otherwise} \end{cases}$	(A.32)
Solar penalty	Encourages net-zero energy use by penalizing load satisfaction from the grid when there is stored energy in ESSs as well as penalizing net export when ESSs are not fully charged [67]. There is neither penalty nor reward when ESSs are fully charged during net export. Whereas, when ESSs are charged to capacity and there is net import the penalty is maximized.	$\sum_{i=0}^{n} -\left(\left(1 + \frac{E_t^{\text{building, net}}}{\left\lvert E_t^{\text{building, net}}\right\rvert} \times \text{ESS}_t^{\text{i, SoC}}\right) \times \left\lvert E_t^{\text{building, net}}\right\rvert\right)$	(A.33)
Comfort	Negative difference between indoor dry-bulb temperature, T_t^{in}, and setpoint, T_t^{spt}, raised to some exponent,$a\|b$, ($a \leq b$) if outside the comfort band, T^{Δ} [80]. If within the comfort band, the reward is the negative difference when in cooling mode and temperature is below the setpoint or when in heating mode and temperature is above the setpoint. The reward is 0 if within the comfort band and above the setpoint in cooling mode or below the setpoint and in heating mode. The exponents teach the control agent to take actions that use the least energy while keeping T_t^{in} close to T_t^{spt}.	$\begin{cases} -\left\lvert T_t^{\text{in}} - T_t^{\text{spt}}\right\rvert^b & \text{if } T_t^{\text{in}} < \left(T_t^{\text{spt}} - T^{\Delta}\right) \\ & \text{\& hvac_mode = cooling} \\ -\left\lvert T_t^{\text{in}} - T_t^{\text{spt}}\right\rvert^a & \text{if } T_t^{\text{in}} < \left(T_t^{\text{spt}} - T^{\Delta}\right) \\ & \text{\& hvac_mode = heating} \\ -\left\lvert T_t^{\text{in}} - T_t^{\text{spt}}\right\rvert & \text{if } \left(T_t^{\text{spt}} - T^{\Delta}\right) \leq T_t^{\text{in}} < T_t^{\text{spt}} \\ & \text{\& hvac_mode = cooling} \\ 0 & \text{if } \left(T_t^{\text{spt}} - T^{\Delta}\right) \leq T_t^{\text{in}} < T_t^{\text{spt}} \\ & \text{\& hvac_mode = heating} \\ 0 & \text{if } T_t^{\text{spt}} \leq T_t^{\text{in}} \leq \left(T_t^{\text{spt}} + T^{\Delta}\right) \\ & \text{\& hvac_mode = cooling} \\ -\left\lvert T_t^{\text{in}} - T_t^{\text{spt}}\right\rvert & \text{if } T_t^{\text{spt}} \leq T_t^{\text{in}} \leq \left(T_t^{\text{spt}} + T^{\Delta}\right) \\ & \text{\& hvac_mode = heating} \\ -\left\lvert T_t^{\text{in}} - T_t^{\text{spt}}\right\rvert^a & \text{if } \left(T_t^{\text{spt}} + T^{\Delta}\right) < T_t^{\text{in}} \\ & \text{\& hvac_mode = cooling} \\ -\left\lvert T_t^{\text{in}} - T_t^{\text{spt}}\right\rvert^b & \text{otherwise} \end{cases}$	(A.34)

Table A.4
Control agents in CityLearn and supported third-party standardized control algorithm libraries.

Name	Theory	Single-agent	Configuration Independent	Coordinated	Description
CityLearn					
Baseline	-	✓	✓		BAU scenario where neither ESSs nor electric devices are controlled and ideal loads are satisfied.
Random	-	✓	✓		Arbitrarily selected actions.
Hour-RBC	RBC	✓	✓		Prescribes control actions based on custom hour-of-use values.
Tabular Q-Learning	RLC	✓	✓		Model-free RLC algorithm for low-dimensional discrete observation and action spaces [83].
SAC	RLC	✓	✓		Implementation of the model-free SAC algorithm [34].
MARLISA	RLC		✓	✓	SAC extension with shared reward and internal building model to predict district electricity consumption for cooperative control [93].
SAC-RBC	RBC, RLC	✓	✓		SAC paired with an Hour-RBC for offline training or initial exploration [68].
MARLISA-RBC	RBC, RLC		✓	✓	MARLISA paired with an Hour-RBC for offline training or initial exploration.
Third-party					
Stable-Baselines3 wrapper	RLC	✓			Interface for Stable-Baselines3 single-agent algorithms [81].
RLlib wrapper	RLC	✓	✓	✓	Interface for RLlib single-agent and multi-agent algorithms [55].

Table A.5
Summary of post-simulation KPIs in CityLearn.

Name	Spatial resolution		Equation
	Building	District	
Thermal comfort			
Discomfort time steps	✓		$\frac{1}{n} \times \sum_{t=0}^{n-1} \begin{cases} 1 & \text{if } \left\lvert T_t^{\text{in}} - T_t^{\text{spt}} \right\rvert > T^{\Delta} \\ 0 & \text{otherwise} \end{cases}$ (A.35)
Cold discomfort time steps	✓		$\frac{1}{n} \times \sum_{t=0}^{n-1} \begin{cases} 1 & \text{if } T_t^{\text{in}} - T_t^{\text{spt}} < -T^{\Delta} \\ 0 & \text{otherwise} \end{cases}$ (A.36)
Hot discomfort time steps	✓		$\frac{1}{n} \times \sum_{t=0}^{n-1} \begin{cases} 1 & \text{if } T_t^{\text{in}} - T_t^{\text{spt}} > T^{\Delta} \\ 0 & \text{otherwise} \end{cases}$ (A.37)
Minimum cold discomfort	✓		$\min\left(\left\lvert \min\left(0, T_0^{\text{in}} - T_0^{\text{spt}}\right)\right\rvert, \ldots, \left\lvert \min\left(0, T_{n-1}^{\text{in}} - T_{n-1}^{\text{spt}}\right)\right\rvert\right)$ (A.38)
Maximum cold discomfort	✓		$\max\left(\left\lvert \min\left(0, T_0^{\text{in}} - T_0^{\text{spt}}\right)\right\rvert, \ldots, \left\lvert \min\left(0, T_{n-1}^{\text{in}} - T_{n-1}^{\text{spt}}\right)\right\rvert\right)$ (A.39)
Average cold discomfort	✓		$\frac{1}{n} \times \sum_{t=0}^{n-1} \left\lvert \min\left(0, T_t^{\text{in}} - T_t^{\text{spt}}\right)\right\rvert$ (A.40)
Minimum hot discomfort	✓		$\min\left(\max\left(0, T_0^{\text{in}} - T_0^{\text{spt}}\right), \ldots, \max\left(0, T_{n-1}^{\text{in}} - T_{n-1}^{\text{spt}}\right)\right)$ (A.41)
Maximum hot discomfort	✓		$\max\left(\max\left(0, T_0^{\text{in}} - T_0^{\text{spt}}\right), \ldots, \max\left(0, T_{n-1}^{\text{in}} - T_{n-1}^{\text{spt}}\right)\right)$ (A.42)
Average hot discomfort	✓		$\frac{1}{n} \times \sum_{t=0}^{n-1} \max\left(0, T_t^{\text{in}} - T_t^{\text{spt}}\right)$ (A.43)
Energy			
Total electricity consumption	✓		$\sum_{t=0}^{n-1} \max\left(E_t^{\text{building, net}}, 0\right)$ (A.44)
Total net electricity consumption	✓		$\sum_{t=0}^{n-1} E_t^{\text{building, net}}$ (A.45)
Average daily peak		✓	$\frac{h}{n} \times \sum_{d=0}^{n \div h-1} \max\left(E_{d \cdot h}^{\text{district, net}}, \ldots, E_{d \cdot h+h-1}^{\text{district, net}}\right)$ (A.46)
All-time peak		✓	$\max\left(E_0^{\text{district, net}}, \ldots, E_n^{\text{district, net}}\right)$ (A.47)
Total ramping		✓	$\sum_{t=1}^{n-1} \max\left(0, E_t^{\text{district, net}} - E_{t-1}^{\text{district, net}}\right)$ (A.48)
Average (1 - load factor)		✓	$\frac{h}{n} \times \sum_{d=0}^{n \div h} 1 - \frac{\left(\sum_{t=d \cdot h}^{d \cdot h+h-1} E_t^{\text{district, net}}\right) \div h}{\max\left(E_{d \cdot h}^{\text{district, net}}, \ldots, E_{d \cdot h+h-1}^{\text{district, net}}\right)}$ (A.49)
Cost			
Total energy cost	✓		$\sum_{t=0}^{n-1} \max\left(E_t^{\text{building, net}}, 0\right) \times R_t$ (A.50)
Environmental			
Total CO_2e emissions	✓		$\sum_{t=0}^{n-1} \max\left(E_t^{\text{building, net}}, 0\right) \times G_t$ (A.51)
Resiliency and thermal comfort			
1 - thermal resilience	✓		$\frac{1}{n^{\text{power outage}}} \times \sum_{t=0}^{n-1} \begin{cases} 1 & \text{if } \left\lvert T_t^{\text{in}} - T_t^{\text{spt}} \right\rvert > T^{\Delta} \ \&\ O_t > 0 \\ 0 & \text{otherwise} \end{cases}$ (A.52)
Resiliency and energy			
Total unserved energy	✓		$\begin{cases} \left(Q_t^{\text{cooling_demand}} + Q_t^{\text{heating_demand}} + Q_t^{\text{dhw_demand}} + Q_t^{\text{non_shiftable_load}} \right. \\ \left. + Q_t^{\text{electric_vehicle_charger}}\right)^{\text{expected}} - \left(Q_t^{\text{cooling_device}} + Q_t^{\text{cooling_storage}}\right. \\ \left. + Q_t^{\text{heating_device}}\right. \\ \left. + Q_t^{\text{heating_storage}} + Q_t^{\text{dhw_device}}\right. \\ \left. + Q_t^{\text{dhw_storage}} + Q_t^{\text{non_shiftable_load}} + Q_t^{\text{electric_vehicle_charger}}\right)^{\text{actual}} & \text{if } O_t > 0 \\ 0 & \text{otherwise} \end{cases}$ (A.53)

Table A.6: CityLearn datasets.

Name[1,2]	Loc.	CZ[3]	Date[4]	Bldgs.	Type[5]	Model			Signal		Load						ESS			PV
						Occupant	Thermal	Outage	Pricing	Emissions	Cooling	Heating	DHW	Non-shift	Cooling	Heating	DHW	Electrical	EV	
*_2020_climate_zone_1	New Orleans, LA, US	2A	Jan-Dec, 05	9	Com					✓			✓	✓	✓		✓	✓		✓
*_2020_climate_zone_2	Atlanta, GA, US	3A	Jan-Dec, 05	9	Com					✓			✓	✓	✓		✓	✓		✓
*_2020_climate_zone_3	Nashville, TN, US	4A	Jan-Dec, 05	9	Com					✓			✓	✓	✓		✓	✓		✓
*_2020_climate_zone_4	Chicago, IL, US	5A	Jan-Dec, 05	9	Com					✓			✓	✓	✓		✓	✓		✓
*_2021	New Orleans, LA, US	2A	Jan-Dec 05, 14-17	9	Com					✓			✓	✓	✓		✓	✓		✓
*_2022_phase_1	Fontana, CA, US	3B	Aug, 16-Jul, 17	5	Res				✓	✓				✓				✓		✓
*_2022_phase_2	Fontana, CA, US	3B	Aug, 16-Jul, 17	5	Res				✓	✓				✓				✓		✓
*_2022_phase_3	Fontana, CA, US	3B	Aug, 16-Jul, 17	7	Res				✓	✓				✓				✓		✓
*_2022_phase_all	Fontana, CA, US	3B	Aug, 16-Jul, 17	17	Res				✓	✓				✓				✓		✓
baeda_3dem	Albuquerque, NM, US	4B	Jun-Aug, TMY3	4	Com		✓		✓		✓		✓	✓	✓		✓			✓
*_2023_phase_1	Austin, TX, US	2A	Jun, 18	3	Res		✓	✓	✓	✓	✓		✓	✓			✓	✓		✓
*_2023_phase_2_local_evaluation	Austin, TX, US	2A	Jun, 18	3	Res		✓	✓	✓	✓	✓		✓	✓			✓	✓		✓
*_2023_phase_2_online_evaluation_1	Austin, TX, US	2A	Jun, 21-Aug, 21	3	Res		✓	✓	✓	✓	✓		✓	✓			✓	✓		✓
*_2023_phase_2_online_evaluation_2	Austin, TX, US	2A	Jun, 21-Aug, 21	3	Res		✓	✓	✓	✓	✓		✓	✓			✓	✓		✓
*_2023_phase_2_online_evaluation_3	Austin, TX, US	2A	Jun, 21-Aug, 21	3	Res		✓	✓	✓	✓	✓		✓	✓			✓	✓		✓
*_2023_phase_3_1	Austin, TX, US	2A	Jun, 21-Aug, 21	6	Res		✓	✓	✓	✓	✓		✓	✓			✓	✓		✓
*_2023_phase_3_2	Austin, TX, US	2A	Jun, 21-Aug, 21	6	Res		✓	✓	✓	✓	✓		✓	✓			✓	✓		✓
*_2023_phase_3_3	Austin, TX, US	2A	Jun, 21-Aug, 21	6	Res		✓	✓	✓	✓	✓		✓	✓			✓	✓		✓
tx_travis_county_neighborhood	Travis Co., TX, US	2A	Jun-Aug 18	100	Res		✓				✓		✓	✓			✓	✓		✓
ca_alameda_county_neighborhood	Alameda Co., CA, US	3C	Jun-Aug 18	100	Res		✓				✓		✓	✓			✓	✓		✓
vt_chittenden_county_neighborhood	Chittenden Co., VT, US	6A	Jan-Mar 18	47	Res		✓					✓	✓	✓			✓	✓		✓
quebec_neighborhood	Quebec, Canada	6A	Jan-Mar 20-22	10	Res	✓	✓		✓			✓	✓	✓						
ev_dataset	Porto, Portugal	3B	Jan-Dec 18-22	9	Com, Res									✓					✓	

[1] The datasets are available at https://github.com/intelligent-environments-lab/CityLearn/tree/master/citylearn/data/datasets.
[2] * in the name is a placeholder for citylearn_challenge.
[3] Climate zone temperature-humidity types: 2A (hot-humid), 3A (warm-humid), 3B (warm-dry), 3C (warm-marine), 4A (mixed-humid), 4B (mixed-dry), 5A (cool-humid), and 6A (cold-humid).
[4] The year is an abbreviation of 20XX.
[5] Building archetypes: Commercial (Com) and Residential (Res).