JEE Journal of Ecological Engineering

Journal of Ecological Engineering, 2026, 27(11), 77–91
https://doi.org/10.12911/22998993/226217
ISSN 2299–8993, License CC-BY 4.0

Received: 2026.07.02
Accepted: 2026.09.15
Published: 2026.10.01

Environmental performance assessment of intelligent cooling control benchmarks in Moroccan residential buildings

Tarik Elyemli[1*], Jihane Gharib[2], Ahmed Ait Errouhi[1]

[1] Laboratory of Advanced Systems Engineering, National School of Applied Sciences, Ibn Tofail University, Kenitra 14000, Morocco
[2] Laboratory of Engineering Science, Ibn Tofail University, Kenitra 14000, Morocco
* Corresponding author's e-mail: tarik.elyemli1@uit.ac.ma

ABSTRACT

Benchmark reward rankings for intelligent cooling control do not necessarily represent thermal control or energy savings. This study introduces a closed-loop audit that separates available thermal authority from realised thermal effect in an occupancy-aware cooling benchmark for an 80 m² Moroccan apartment. A simplified model calibrated against EnergyPlus was evaluated using a Deep Q-Network (DQN) and deterministic setpoint rules. Across 209 hourly transitions with available authority, the DQN-derived mean setpoint rule and the deterministic rules produced no cooling activation, whereas seed-specific DQN action sequences showed limited activity for two of three seeds. Full-year EnergyPlus simulations compared four thermostat schedules in Kenitra and Marrakech. For EnergyPlus, the DQN was represented by an occupancy-conditioned mean setpoint schedule of 27/28 °C constructed from the three trained seeds; this rounded approximation is not an exact hourly replay of any trained policy. Relative to the fixed 26 °C schedule, the optimised 27/30 °C schedule reduced specific thermal cooling demand by 20.5% in Kenitra and 17.0% in Marrakech. The reward proxy ranked the 26/30 °C rule above the DQN-derived mean setpoint rule, whereas the corresponding DQN-derived schedule had lower simulated cooling demand in both climates. Benchmark reward, realised control and full-building simulated cooling demand should therefore be assessed separately.

Keywords: environmental performance, intelligent cooling control, building energy simulation, cooling demand, EnergyPlus, residential buildings, Morocco.

INTRODUCTION

Residential cooling is becoming a major energy and environmental concern in warm and hot climates, where cooling can dominate residential energy use and offers substantial saving potential (Alalouch et al., 2019). In Morocco, the residential sector faces rising cooling needs, climate-sensitive building performance, and growing attention to energy-efficient construction and operation. National building-energy regulations and recent Moroccan studies stress the role of envelope performance, thermal comfort, and energy efficiency in reducing building-related energy demand (Boumlik et al., 2024; El Asri et al., 2023). Operational cooling-setpoint control can complement passive and envelope measures, which recent Moroccan studies have examined across the country's climate zones (Lachir and Noufid, 2024; M'lahfi et al., 2020): it reduces unnecessary cooling during unoccupied periods while keeping indoor comfort acceptable.

Occupancy-aware and intelligent cooling-control strategies are therefore relevant energy-saving technologies for residential buildings. Fixed thermostat schedules remain common in practice, but they ignore occupancy patterns, outdoor temperature variation, and the thermal inertia of the building (Anand et al., 2022; Dong et al., 2018). Letting the indoor temperature rise when the dwelling is empty, then restoring comfort before or during occupancy, is a simple operational lever for cutting cooling that no occupant benefits from (Turley et al., 2020; Uddin et al., 2021). Occupancy-driven and temperature-prediction approaches to HVAC efficiency have also been investigated for building energy management (Drif et al., 2025; Sbabou et al., 2025). Rule-based control, model predictive control, and reinforcement learning have all been used to automate this decision (Afram and Janabi-Sharifi, 2014), and recent reviews report the potential of advanced control for building energy management (Taheri et al., 2022; Yao and Shekhar, 2021).

The environmental value of intelligent cooling control depends on whether a reported improvement corresponds to physically credible energy savings, not merely on whether an algorithm scores well inside a benchmark. This matters most when reduced simulation environments or reward functions are used to compare control policies. A benchmark reward may combine comfort penalties, switching penalties, or designed energy proxies that help train and rank policies, yet these quantities need not represent measured or simulated cooling demand. For energy-saving assessment in environmental engineering, benchmark rankings must therefore be checked against simulated building-energy indicators.

The reinforcement-learning literature for HVAC control has grown rapidly (Azuatalam et al., 2020; Michailidis et al., 2023). Reported applications span setpoint reset and demand-charge reduction (Fang et al., 2022; Jiang et al., 2021), as well as temperature and humidity control and residential energy management (Boutahri and Tilioua, 2025; Chen et al., 2023). Evaluation problems persist. Learned controllers are often trained and scored inside reduced environments whose fidelity to a full building model is rarely assessed end to end, and they are sometimes compared against weak or conveniently chosen baselines (Al Sayed et al., 2024; Zhang et al., 2019). Simulation-and-control frameworks have been proposed to standardise these evaluations (Blum et al., 2021; Jiménez-Raboso et al., 2021), and comparative studies have examined how different deep reinforcement-learning algorithms perform on building-control tasks (Esrafilian-Najafabadi and Haghighat, 2021; Manjavacas et al., 2024).

Most critically, a high benchmark reward is often read as evidence of energy savings or better control, even when the reward contains a designed proxy that does not represent measured or simulated HVAC energy (Dulac-Arnold et al., 2021). Evaluations rarely separate the thermal influence that actions could in principle exert from the influence the chosen policy actually exerts.

This gap is the focus of the present work. Rather than treating reinforcement learning as an end in itself, we use it as a case study for whether intelligent cooling-control benchmarks can support credible energy-saving claims. Ranking cooling-control policies by benchmark reward implicitly assumes that a reward improvement reflects improved thermal control and reduced cooling demand. That assumption can be structurally false. To test it, we introduce a closed-loop audit that separates the thermal authority available to a policy, defined as the next-state temperature change that alternative valid actions could produce, from the thermal effect it actually exerts, defined as the cooling correction realised by the selected action. This internal audit is paired with full EnergyPlus simulations under two contrasting Moroccan climates.

Applied to occupancy-aware cooling-setpoint control in an 80 m² Moroccan apartment, the audit reveals a regime in which the DQN-derived mean setpoint rule and the deterministic rules reach identical comfort and are ranked only by the designed energy proxy, without producing cooling activation. The seed-specific DQN action sequences, however, show rare cooling corrections for seeds 0 and 1. The ranking therefore reflects reward-proxy alignment, not demonstrated thermal control.

We further show that the very limited use of available thermal authority cannot be attributed to the learning algorithm alone. It arises primarily from how the simplified benchmark is calibrated and parameterised, although rare cooling corrections occurred in two of the three seed-specific DQN action sequences. A within-benchmark ranking cannot on its own support cooling-energy claims. We therefore conducted full annual EnergyPlus simulations under two contrasting Moroccan climates, Kenitra and Marrakech, and reported the simulated annual zone cooling thermal demand of four thermostat schedules. The EnergyPlus results confirm the best and worst schedules identified by the proxy but reveal an inversion of the two intermediate schedules: the 26/30 °C occupancy-based schedule ranks above the DQN-derived mean setpoint schedule under the reward proxy, whereas the DQN-derived mean setpoint schedule produces a lower simulated cooling thermal demand. Internal benchmark behaviour and full-building simulation outputs are therefore distinct quantities, and both must be examined.

The contribution of this work is methodological. We propose a transparent audit workflow that pulls apart four things RL-for-HVAC studies often merge: how accurately the environment is calibrated, whether the reward is optimised, how much thermal influence a policy could exert, and how much it actually exerts. The internal audit is paired with EnergyPlus-based simulation-based assessment. This work does not set out to evaluate RL control performance. It evaluates the validity of the evaluation itself, asking whether a reward-based ranking can be trusted as evidence of thermal control and energy saving.

MATERIALS AND METHODS

Figure 1 summarises the methodology. The workflow comprises five phases: an initial EnergyPlus baseline and data extraction; the calibration of a lightweight thermal environment with an independent holdout; the formulation of the RL task and the rule-based policies to be compared; a closed-loop audit separating available thermal authority from realised thermal effect; and a simulation-based assessment of the candidate thermostat schedules with full EnergyPlus simulations under two contrasting climates.

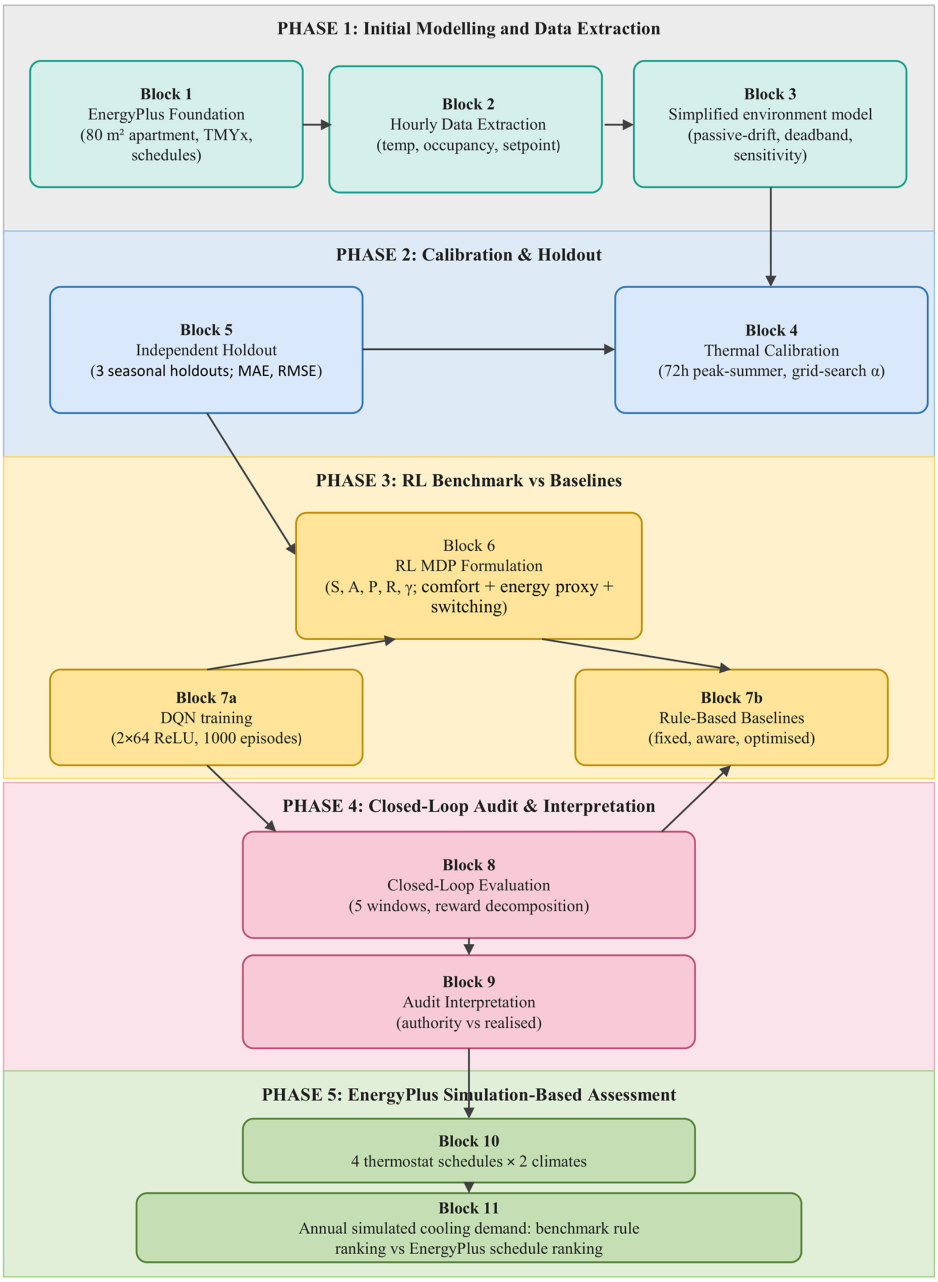

Figure 1. Overview of the proposed audit-based benchmarking and simulation-based assessment workflow

Case study and EnergyPlus baseline

The case study is an 80 m² residential apartment in Kenitra, Morocco, modelled in OpenStudio and EnergyPlus (Crawley et al., 2001) with its geometry, envelope, two thermal zones, occupancy schedule, internal gains, and thermostat schedules. Exterior walls are double hollow brick (U = 1.21 W/m²·K), the flat roof has 50 mm of expanded polystyrene (U = 0.62 W/m²·K), and windows are double glazing (U = 2.80 W/m²·K, SHGC = 0.60), with a window-to-wall ratio of about 3.4%. Occupancy follows a deterministic binary schedule (occupied 18:00–08:00, unoccupied 08:00–18:00). Reference cooling setpoints are 26 °C (occupied) and 30 °C (unoccupied); heating setpoints are 20 °C and 15 °C. The OpenStudio geometry of the modelled apartment is shown in Figure 2.

Annual EnergyPlus simulations were performed using the Ideal Loads Air System. This idealised system supplies the heating or cooling load required to maintain the prescribed zone thermostat setpoints without representing the efficiency, capacity limitations, or electricity consumption of actual HVAC equipment. The "initial constant-setpoint Ideal Loads configuration" denotes the original model configuration before introducing occupancy-dependent thermostat schedules, with constant cooling and heating setpoints of 26 °C and 20 °C, respectively. The occupancy-scheduled configuration applies cooling setpoints of 26 °C during occupied hours and 30 °C during unoccupied hours, together with heating setpoints of 20 °C and 15 °C, respectively. Table 1 reports total delivered thermal demand, calculated as the sum of annual heating and cooling loads and normalised by the conditioned floor area of 80 m².

The value of 46.35 kWh_{th}/m²·yr in Table 1 is not directly comparable with the fixed-26 value of approximately 45.3 kWh_{th}/m²·yr reported in the subsequent policy comparison. The former represents total annual delivered thermal demand, including both heating and cooling, for the initial constant 26/20 °C thermostat configuration. The latter represents annual cooling demand alone for the fixed-26 °C cooling strategy; heating is excluded from that indicator. The difference therefore results primarily from the different energy-accounting boundaries. The two values should not be interpreted as two estimates of the same cooling demand.

Calibrated lightweight thermal environment

To allow repeated RL training and policy evaluation at low computational cost, a first-order thermal environment was implemented in Python. It is intended as a transparent control-oriented benchmark, not as a reduced-order EnergyPlus surrogate. At each hourly step, the indoor

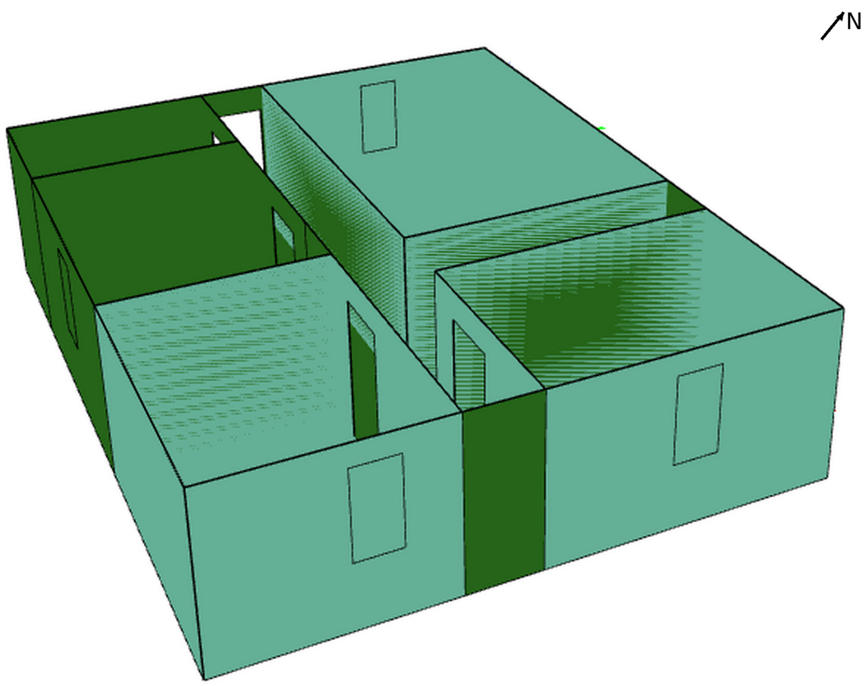

Figure 2. OpenStudio 3D geometry of the modelled apartment

Table 1. Annual total delivered thermal demand under the EnergyPlus Ideal Loads reference configurations

Configuration	Total specific thermal demand, heating + cooling ($kWh_{th}/m^2 \cdot yr$)
Initial constant-setpoint Ideal Loads configuration (26 °C cooling; 20 °C heating)	46.35
Occupancy-scheduled Ideal Loads configuration (26/30 °C cooling; 20/15 °C heating)	45.24

Note: The values reported in Table 1 are the sum of the annual heating and cooling thermal loads delivered by the EnergyPlus Ideal Loads Air System, divided by the conditioned floor area of 80 m². They represent delivered zone thermal demand, not HVAC electricity consumption. By contrast, the subsequent policy comparison reports cooling thermal demand alone.

temperature first drifts passively toward the outdoor temperature according to Equation 1:

$$T_t^{passive} = T_t + \alpha\left(T_{out,t} - T_t\right) \quad (1)$$

where: T_t is the current mean indoor temperature, $T_{out,t}$ is the outdoor dry-bulb temperature, and α is an effective thermal-response coefficient; a lower α represents stronger effective thermal inertia.

Let ΔT_{db} denote the thermostat deadband. A cooling correction is applied only when the passively drifting temperature exceeds the selected setpoint $T_{sp,t}$ by more than ΔT_{db} = 0.5 °C, retained as a representative residential hysteresis value (Ramaraj and Sparn, 2025). The maximum setpoint-dependent cooling strength per hourly step is defined by Equation 2:

$$s(T_{sp,t}) = clip\ \left[2.0 - 0.4\left(T_{sp,t} - 24\right),\ 0,\ 2.0\right] \quad (2)$$

where: *clip* (x, l, u) – min (max (x, l), u).

Thus, lower cooling setpoints give stronger corrective action, bounded between 0 and 2.0 °C per step. The cooling correction is therefore given by Equation 3:

$$C_t = \begin{cases} \min\ \left(T_t^{passive} - T_{sp,t},\ s\left(T_{sp,t}\right)\right), & \text{if } T_t^{passive} > T_{sp,t} + \Delta T_{db} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

The next indoor temperature follows from Equation 4:

$$T_{t+1} = T_t^{passive} - C_t \quad (4)$$

C_t is a simplified control-oriented temperature correction. It does not represent HVAC equipment operation, compressor power, coefficient of performance, electricity consumption, or system-level energy use.

The coefficient α was calibrated against the EnergyPlus indoor-temperature trajectory over a dedicated 72-hour peak-summer window by minimising RMSE, giving α^* = 0.005. The calibrated environment was then tested on three independent 72-hour holdout windows (warm shoulder, early summer, late summer); the resulting accuracy is reported under "Calibration of the simplified environment" in the Results and Discussion section. The model reproduces the mean indoor-temperature level more reliably than the response amplitude, hence its use as a policy-ranking benchmark only.

The primary calibration grid covered $\alpha \in [0.001, 0.020]$ in increments of 0.001. A diagnostic extension to α = 0.300 was also examined to verify that no lower-error solution existed at larger response coefficients. The trajectory-based optimum was α^* = 0.005, for which the peak-summer calibration window gave MAE = 0.144 °C, RMSE = 0.196 °C, maximum absolute error = 0.529 °C, and MBE = −0.057 °C. The extended scan showed no secondary minimum. Supporting one-step diagnostics also placed the coefficient below 0.01: the global-regression estimate was 0.00839, the passive-filtered regression estimate was 0.00780, and the one-step RMSE optimum was 0.0085. These diagnostics support α = 0.005 as a benchmark-specific effective coefficient; it is not a measurable material or envelope property.

Sensitivity checks considered deadbands of 0.3, 0.5, and 1.0 °C and cooling-strength schedules scaled by −50% and +50%. For the deterministic rules and the DQN-derived mean setpoint rule examined in the aggregate audit, these changes did not alter the evaluated temperature trajectories because the selected actions did not activate cooling. This result is trajectory-specific and does not imply that the deadband or cooling strength is unimportant in other operating regimes.

RL formulation and evaluated policies

The cooling-setpoint task is a discrete-action Markov Decision Process: M = (S, A, P, R, γ = 0.99). Here S is eight-dimensional (indoor and outdoor temperature, binary occupancy, cyclical encodings of hour and month, and previous setpoint). Actions are occupancy-constrained: {24, 25, 26, 27} °C when occupied and {26, 27, 28, 29, 30} °C when unoccupied. The reward combines a comfort term penalising departures from the 25–27 °C occupied band adopted here in line with standard thermal-comfort criteria (ISO, 2005; ANSI/ASHRAE, 2020), a designed setpoint-dependent energy proxy, and a setpoint-switching penalty.

The instantaneous reward is defined by Equation 5 as the weighted sum of three components:

$$R_t = -4 \times v_t^2 - 0.35 \times E_{proxy,t} - \\ - 0.02 \times |T_{sp,\,t} - T_{sp,\,t-1}| \tag{5}$$

where: v_t is the comfort violation, $E_{proxy,\,t}$ is the designed setpoint-dependent energy proxy, and the final term penalises abrupt setpoint changes.

The comfort violation is defined using the predicted next indoor temperature T_{t+1}. During occupied periods, the accepted comfort band is 25–27 °C (ISO, 2005; ANSI/ASHRAE, 2020), and v_t is given by Equation 6:

$$v_t = \begin{cases} 25 - T_{t+1}, & T_{t+1} < 25 \\ 0, & 25 \le T_{t+1} \le 27 \\ T_{t+1} - 27, & T_{t+1} > 27 \end{cases} \tag{6}$$

During unoccupied periods, only overheating above 30 °C is penalised, as given by Equation 7:

$$v_t = max\ (0, T_{t+1} - 30) \tag{7}$$

The setpoint-dependent scaling factor and the resulting energy proxy are defined by Equations 8 and 9, respectively:

$$f_{sp,t} = \text{clip}\ \left(\frac{31 - T_{sp,t}}{7},\ 0,\ 1\right) \tag{8}$$

$$E_{proxy,t} = f_{sp,t}\ \max\ \left(0, T_{out,t} - T_{sp,t}\right) + \\ + 0.5\max\ \left(0, T_t - T_{sp,t}\right) \tag{9}$$

The proxy therefore depends on the selected setpoint and on the contemporaneous outdoor and indoor temperatures. It carries no physical unit, was not calibrated against EnergyPlus cooling loads, measured HVAC electricity, compressor power, or equipment-level energy use, and is not intended to estimate kWh. The resulting reward ranking is an internal benchmark quantity and cannot by itself support a simulation-based energy-saving claim.

The Deep Q-Network (DQN; Mnih et al., 2015) received the eight-dimensional state vector and produced one Q-value for each candidate cooling setpoint. The network comprised two fully connected hidden layers of 64 ReLU units and a linear output layer. Training used experience replay, a target network, epsilon-greedy exploration, the Adam optimiser, and the Huber loss. The network comprised seven action outputs. Occupancy-invalid actions were masked during action selection, while the bootstrap target was computed from the maximum of the seven target-network outputs. Each episode contained 72 hourly steps, and training used 1,000 episodes for each of three random seeds. The discount factor was 0.99, the learning rate was 1×10^{-3}, the batch size was 64, and the replay-buffer capacity was 20,000 transitions. Epsilon decreased from 1.0 to 0.05 with a decay factor of 0.992, and the target network was updated every 20 episodes. The five evaluation windows were excluded from training, and evaluation used deterministic greedy action selection.

Throughout this study, "policy" denotes a state-to-action mapping evaluated within the simplified benchmark. The three neural networks obtained from random seeds 0–2 are referred to as the trained DQN policies. A "seed-specific DQN action sequence" denotes the actual hourly setpoints selected by one of these policies over the evaluation windows. "Rule" denotes deterministic setpoint-selection logic evaluated within the simplified benchmark, whereas "schedule" denotes a prescribed sequence of thermostat setpoints used in the EnergyPlus simulations. The DQN-derived mean setpoint approximation was implemented as a 27/28 °C rule in the simplified benchmark and as a 27/28 °C thermostat schedule in EnergyPlus. It is not an exact hourly replay of any trained DQN policy.

Closed-loop policy-influence audit

The audit determines whether a policy merely selects different setpoints or actually changes the thermal state. At each evaluated state, the next indoor temperature is computed counterfactually for every valid action a, and the available thermal authority is the span between the maximum and minimum counterfactual next-state temperatures; authority is considered present when this span exceeds 0.10 °C. The realised thermal effect of the selected action is the magnitude of its cooling correction relative to the passive trajectory; an actual-effect hour is counted when it exceeds 0.10 °C. The authority-use ratio is the number of actual-effect hours divided by the number of available-authority hours. A cooling activation is recorded whenever the cooling correction is non-zero.

The audit was applied to five 72-state windows: the peak-summer calibration window, three seasonal holdout windows (warm shoulder, early summer, and late summer), and an independent high-demand stress-test window. The stress-test window extended from 30 July 2001 at 00:00 to 1 August 2001 at 23:00 in the TMY calendar; these timestamps identify positions in the representative weather sequence rather than an observed historical event. Its outdoor dry-bulb temperature had a mean of 24.59 °C, a minimum of 22.24 °C, and a maximum of 28.15 °C. The window was selected independently of the month-constrained calibration and holdout windows because it had the highest overnight minimum outdoor temperature among the five evaluation windows, 22.24 °C during the 00:00–07:59 period. It was not the window with the highest peak temperature, as the peak-summer calibration window reached 31.63 °C. The high overnight minimum limits passive nighttime cooling and therefore increases the likelihood that an active cooling correction could become necessary. At the initial time step, the indoor temperature was 26 °C in both thermal zones, the cooling setpoint was 26 °C, and the apartment was occupied by four occupants. Because the final state has no successor within the selected window, each 72-state trajectory provides 71 evaluable hourly transitions. The four calibration and seasonal holdout windows each contributed 42 available-authority hours, whereas the stress-test window contributed 41 h, corresponding to 19.6% of the total 209 h. No realised thermal effect or cooling activation occurred during the stress-test window under any evaluated policy.

Simulation-based assessment with EnergyPlus

Because a within-benchmark ranking cannot by itself support building-level energy claims, the candidate thermostat schedules were evaluated using full annual EnergyPlus simulations. Four cooling thermostat schedules were simulated under two contrasting Moroccan climates: the optimised schedule (27/30 °C), the occupancy-based schedule (26/30 °C), the fixed setpoint schedule (26 °C), and the DQN-derived mean setpoint schedule (27/28 °C).

All annual cases used the same building model, occupancy and internal-gain assumptions, Ideal Loads system, and, within each climate, the same weather file. Their occupied cooling setpoints remained within the prescribed 25–27 °C comfort band. The comparison is therefore comfort-constrained at the prescribed setpoint level. However, it does not establish exact equivalence of operative temperature, PMV/PPD, or occupant-experienced comfort among the schedules. The cooling-demand ranking is interpreted subject to this limitation.

The two climates were Kenitra (Atlantic, mild) and Marrakech (semi-arid, hot), represented using TMYx EPW files downloaded from Climate.OneBuilding.org. According to the provider metadata, these files were constructed from NCEI ISD and ERA5 data over the 2011–2025 period of record. The Kenitra simulations used MAR_RK_Kenitra.AP-Port.Lyautey.601200_TMYx.2011-2025.epw, corresponding to Kenitra Airport–Port Lyautey (WMO 601200; 34.300° N, 6.600° W; elevation 6.0 m; GMT+1). The Marrakech simulations used MAR_MS_Marrakesh-Menara.AP.602300_TMYx.2011-2025.epw, corresponding to Marrakesh–Menara Airport (WMO 602300; 31.607° N, 8.036° W; elevation 467.9 m; GMT+1). The downloaded EPW files were used directly in EnergyPlus without modification; no bias correction, interpolation, resampling, or gap filling was applied. Calendar years attached to the processed hourly timestamps identify positions in the representative TMYx sequence rather than observed historical events.

Two modelling choices must be stated explicitly. First, the term "DQN-derived mean setpoint approximation" denotes an occupancy-conditioned representation constructed from the cooling setpoints selected by the three independently trained DQN policies over the five evaluation windows. Across 1,065 evaluated actions, the mean selected setpoint was 26.632 °C during occupied periods (n = 627) and 28.048 °C during unoccupied periods (n = 438). These values were rounded to the nearest degree, producing the two-level 27/28 °C approximation. Within the simplified benchmark, this approximation was implemented as the “DQN-derived mean setpoint rule”. In the EnergyPlus simulations, it was implemented as the “DQN-derived mean setpoint schedule”. The approximation therefore does not denote a mean discrete-action index, a mean audit result across random seeds, or a single representative seed. It is not an exact hourly replay of any trained DQN policy. Accordingly, its EnergyPlus results are interpreted as a simulation-based consistency assessment of the mean setpoint behaviour.

Second, the cooling values reported by OpenStudio/EnergyPlus represent the annual zone cooling thermal load delivered by the Ideal Loads Air System. They were converted to kWh_{th} and do not represent equipment-level electricity consumption. An indicative electrical estimate, expressed in kWh_{el}, was calculated by post-processing the thermal load using a uniform seasonal COP of 3.0. Because the same COP was applied to every schedule, this conversion does not affect their relative ranking. Such EnergyPlus-based workflows can be implemented flexibly using Python-enabled tools (Bonnema et al., 2025).

RESULTS AND DISCUSSION

Calibration of the simplified environment

Calibration on the peak-summer window selected $\alpha^* = 0.005$, with MAE = 0.144 °C and RMSE = 0.196 °C on that trajectory. On the three independent holdouts the calibrated environment achieved a pooled MAE of 0.744 °C and RMSE of 0.816 °C, all per-window RMSE values staying below 1 °C. The per-window correlations (0.857–0.950) are reported as a secondary indicator of temporal tendency only; agreement is assessed primarily through the residual-based error metrics and the amplitude differences visible in Figure 3, since correlation alone does not establish amplitude agreement. The holdout trajectories show a visibly damped response and a systematic negative bias: pooled MAE and RMSE were 0.744 and 0.816 °C, respectively, pooled MBE was −0.743 °C, and the largest holdout error reached 1.442 °C. The amplitude discrepancy is therefore explicitly visible in Figure 3 and limits the model to control oriented benchmarking; it should not be interpreted as a reduced-order EnergyPlus surrogate.

Internal audit: Available authority with limited realised effect

Across the five evaluation windows, alternative valid actions could have changed the next thermal state during 209 hourly transitions, confirming the presence of measurable thermal authority. The DQN-derived mean setpoint rule and the three deterministic rules produced no realised thermal effect above the 0.10 °C threshold and no cooling activation. The seed-specific DQN action sequences, however, revealed limited but non-zero activity: seed 0 produced one realised-effect hour and one cooling activation, corresponding to an authority-use ratio of 0.005; seed 1 produced four realised-effect hours and four activations, corresponding to a ratio of 0.019; and seed 2 produced no realised effect or cooling activation. The DQN-derived mean setpoint rule (27/28 °C) therefore masks rare control activity present in two of the three trained DQN policies. Accordingly, the trained policies, their seed-specific action sequences, and the mean setpoint approximation are reported separately in Table 2. The DQN-derived mean setpoint rule and the three deterministic rules produced the same indoor-temperature trajectory and an identical cumulative comfort component (−9.071): with $\alpha = 0.005$, their selected actions did not activate cooling. The seed-specific DQN action sequences, however, showed rare trajectory departures for seeds 0 and 1, consistent with the one and four realised-effect hours reported in Table 2, respectively, while seed 2 showed none. Consequently, the zero-activation and identical-trajectory result applies to the DQN-derived mean rule and the deterministic rules, not to every trained DQN policy. For the aggregate mean-rule and deterministic-rule cases, the ranking differences arose primarily from the designed energy proxy and, to a lesser extent, the switching penalty. The ranking obtained inside the benchmark is thus a ranking of setpoint-selection behaviour under the reward proxy, *not* of demonstrated thermal control. While reward decomposition is common in DRL–HVAC studies, we are not aware of prior work that separately quantifies available thermal authority and realised thermal effect within the same benchmark.

Reward ranking versus realised control evidence

Under the benchmark reward, the optimised 27/30 °C rule obtained the highest mean reward

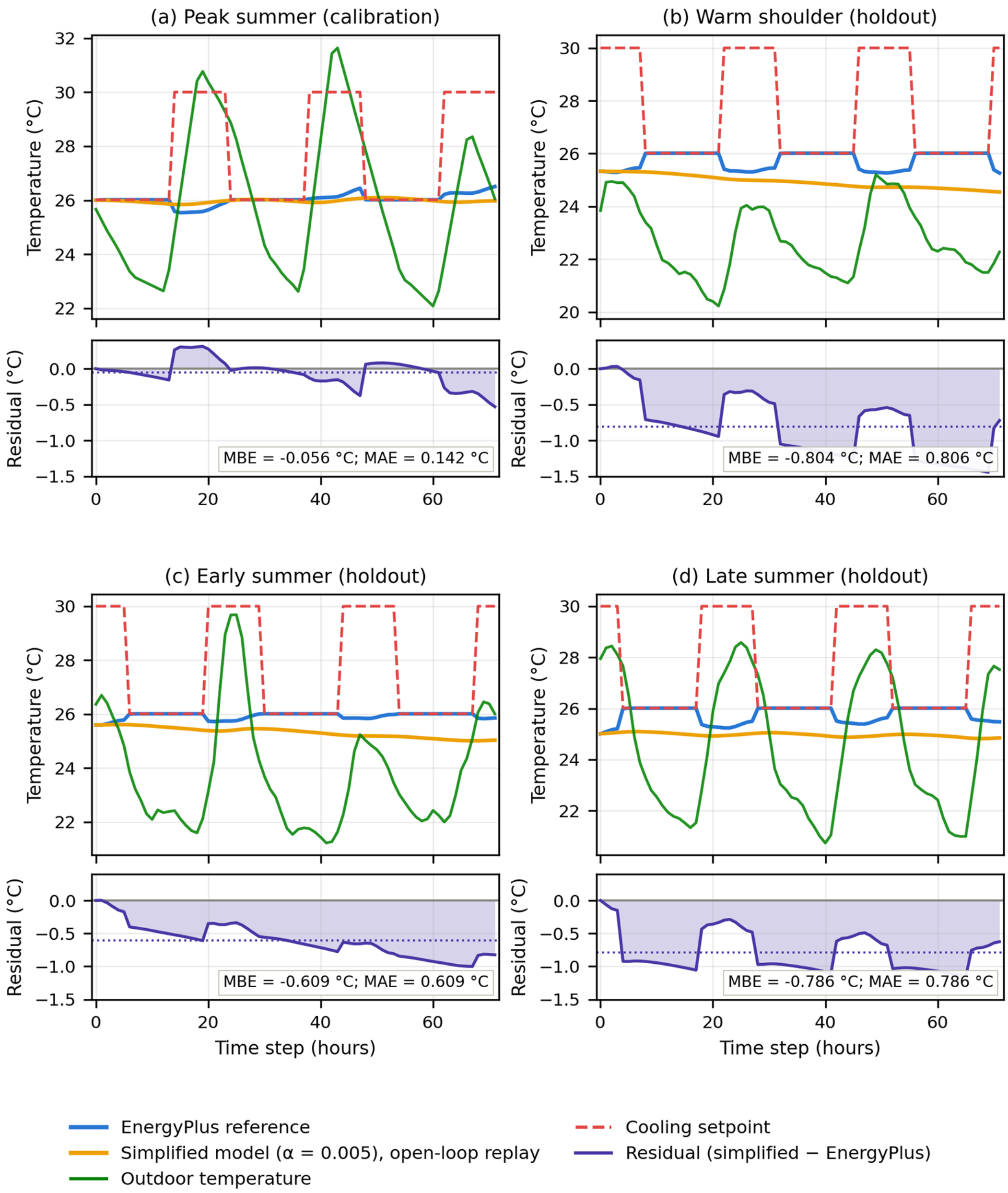

Figure 3. Comparison of the hourly indoor-temperature trajectories obtained from EnergyPlus and the simplified open-loop model for the peak-summer calibration window and three independent holdout windows. Residuals are calculated as the simplified-model temperature minus the EnergyPlus reference temperature. The dotted horizontal lines in the residual panels indicate the mean bias error (MBE), while MBE and mean absolute error (MAE) are reported for each window. The residuals reveal that the simplified model reproduces the general temporal tendency but systematically underestimates the response amplitude during the holdout periods. The separately selected stress-test window is documented in the closed-loop audit analysis.

(−2.501), followed by the occupancy rule (26/30 °C; −3.159), the DQN-derived mean setpoint rule (−4.178), and the fixed-26 °C rule (−9.960). This ordering remained unchanged when alternative reward weights were applied post hoc to the fixed evaluated trajectories, without DQN retraining (details are provided in the repository). Two facts qualify this ranking. First, the leading cases share the same cumulative comfort component, so their reward differences are driven primarily by the designed energy proxy. Second, the DQN-derived mean setpoint rule and the three deterministic rules produced no cooling activation or realised thermal effect, whereas the seed-specific DQN action sequences showed limited but non-zero activity for seeds 0 and 1, using at most 1.9% of the available thermal authority. The benchmark ranking therefore cannot by itself support claims

Table 2. Closed-loop policy-influence audit across the five evaluation windows

Audit case	Available authority (h)	Realised effect (h)	Cooling activations	Authority-use ratio
DQN seed 0 action sequence	209	1	1	0.005
DQN seed 1 action sequence	209	4	4	0.019
DQN seed 2 action sequence	209	0	0	0.000
DQN-derived mean setpoint rule (27/28 °C)	209	0	0	0.000
Optimised rule (27/30 °C)	209	0	0	0.000
Occupancy rule (26/30 °C)	209	0	0	0.000
Fixed setpoint rule (26 °C)	209	0	0	0.000

Note: available thermal authority was evaluated over 209 hourly transitions for each audit case. The seed-specific rows report the actual hourly action sequences selected by the three trained DQN policies. The DQN-derived mean setpoint rule is a two-level occupancy-conditioned approximation obtained by rounding the occupied and unoccupied mean selected setpoints to 27 °C and 28 °C, respectively. It is not an exact replay of any trained DQN policy.

of superior thermal control or building cooling-energy savings, which motivates the subsequent simulation-based EnergyPlus assessment.

Simulation-based assessment with EnergyPlus

Table 3 reports the annual cooling demand of the four thermostat schedules under both climates, with the indicative COP-based electrical estimate. Three observations follow.

The cooling-demand reductions relative to the fixed 26 °C schedule provide an energy-saving interpretation of the EnergyPlus results. The optimised 27/30 °C schedule reduced specific thermal cooling demand by 20.5% in Kenitra and 17.0% in Marrakech, while the DQN-derived mean setpoint schedule reduced it by 13.0% and 11.6%, respectively. These results show that the EnergyPlus stage provides a simulation-based environmental-engineering assessment of the energy-saving potential of the evaluated thermostat schedules. The apparent contrast arises because the closed-loop audit evaluates the internal dynamics of the simplified benchmark, whereas the full-building EnergyPlus simulations evaluate annual zone cooling thermal loads under scheduled thermostat setpoints. The DQN-related values therefore apply to the mean setpoint schedule and do not represent an exact hourly replay of any trained DQN policy.

First, simulated cooling differences exist in both climates, not only in the hotter one. At Kenitra the cooling demand ranges from 36.0 $kWh_{th}/m^2 \cdot yr$ (27/30) to 45.3 $kWh_{th}/m^2 \cdot yr$ (fixed 26);

Table 3. Annual thermal cooling demand and energy-saving potential of the four thermostat schedules under two contrasting Moroccan climates, based on full-year EnergyPlus simulations using the Ideal Loads Air System

Thermostat schedule	Climate	Thermal cooling demand (kWh_{th}/yr)	Specific thermal cooling demand ($kWh_{th}/m^2 \cdot yr$)	Indicative annual electricity use (kWh_{el}/yr, COP=3)	Reduction vs fixed 26 (%)
Optimised schedule (27/30 °C)	Kenitra	2881	36.0	960	20.5
DQN-derived mean setpoint schedule (27/28 °C)	Kenitra	3150	39.4	1050	13.0
Occupancy schedule (26/30 °C)	Kenitra	3614	45.2	1205	0.2
Fixed setpoint schedule (26 °C)	Kenitra	3628	45.3	1209	0.0
Optimised schedule (27/30 °C)	Marrakech	4064	50.8	1355	17.0
DQN-derived mean setpoint schedule (27/28 °C)	Marrakech	4325	54.1	1442	11.6
Occupancy schedule (26/30 °C)	Marrakech	4808	60.1	1603	1.8
Fixed setpoint schedule (26 °C)	Marrakech	4895	61.2	1632	0.0

Note: The DQN-derived row reports the 27/28 °C mean setpoint schedule constructed from the occupied and unoccupied mean setpoints selected by the three trained DQN policies. It is a rounded occupancy-conditioned approximation and not an exact hourly replay of any trained DQN policy.

at Marrakech it ranges from 50.8 to 61.2 kWh_{th}/m²·yr. Marrakech provides a higher-demand contrast case (about +41% for the 27/30 schedule relative to Kenitra) but does not introduce a qualitatively different ranking.

Second, the simulated cooling-demand ranking is stable across both climates: 27/30 < DQN < 26/30 < fixed 26. The most relaxed occupancy-aware schedule consumes least, and the always-on fixed setpoint consumes most, as expected. The gap between 26/30 and fixed 26 is small in both climates: at Kenitra the two are near-equivalent in simulated demand (45.2 versus 45.3 kWh_{th}/m²·yr, within EnergyPlus rounding), and only at Marrakech does a clear separation appear.

Figure 4 shows the corresponding cooling demand by thermostat schedule and climate. Third, the EnergyPlus schedule ranking only partly matches the within-benchmark rule ranking reported in Table 4. The proxy correctly identifies the optimised 27/30 °C case as the best and the fixed 26 °C case as the worst, but it inverts the two intermediate cases. Within the simplified benchmark, the 26/30 °C occupancy rule outranks the DQN-derived mean setpoint rule (−3.159 versus −4.178). In the EnergyPlus simulations, however, the DQN-derived mean setpoint schedule has a lower cooling demand than the 26/30 °C occupancy schedule and therefore ranks second rather than third. This inversion is consistent across both climates.

The mean setpoint schedule receives a larger proxy penalty mainly because its approximate unoccupied setpoint of 28 °C is lower than the 30 °C unoccupied setpoint of the occupancy-based schedule. The proxy penalises this difference strongly, whereas the EnergyPlus results indicate that it has a smaller effect on annual cooling thermal demand than the proxy implies. The proxy therefore preserves the lowest- and highest-demand schedules while materially misranking the two intermediate schedules. This inversion is consistent across both climates because the EnergyPlus-based schedule ranking is identical in Kenitra and Marrakech.

Why the decoupling appears, and why the audit is needed

The internal decoupling is not an algorithmic failure but a structural property of the calibrated benchmark. The value $\alpha = 0.005$ was selected because it best matches the EnergyPlus

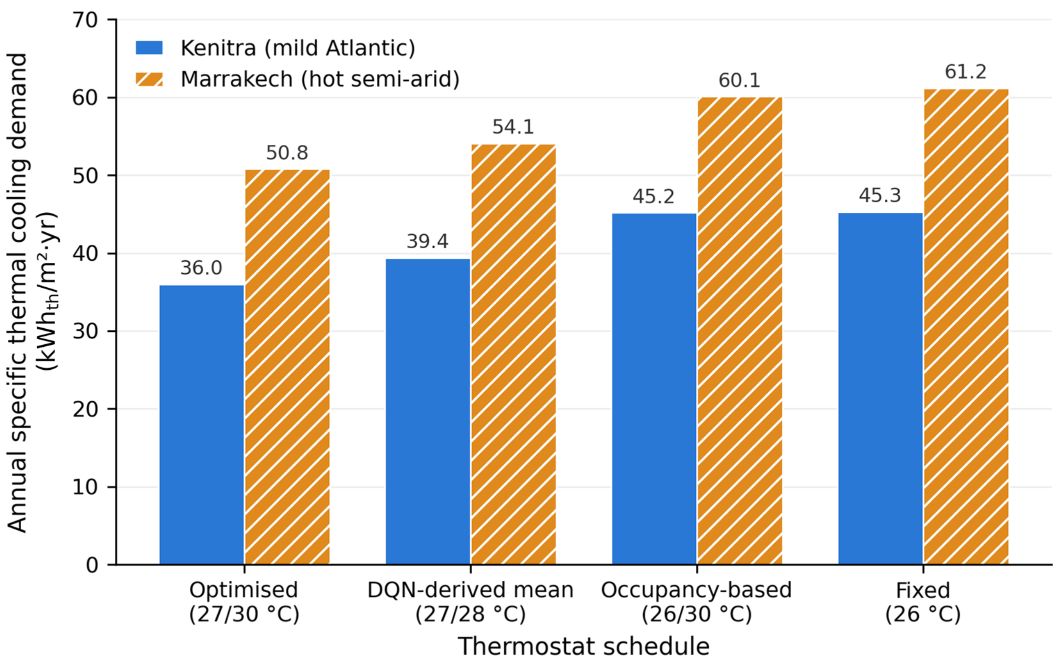

Figure 4. Annual specific thermal cooling demand of the four thermostat schedules under the two contrasting Moroccan climates, Kenitra and Marrakech. The DQN-derived mean setpoint schedule is a rounded 27/28 °C occupancy-conditioned approximation and not an exact hourly replay of any trained DQN policy. Each bar represents one deterministic annual EnergyPlus simulation using one TMYx weather file and the same occupancy profile and internal-gain assumptions. No statistical error bars are shown because no probabilistic uncertainty or repeated-simulation distribution was generated. The separate Kenitra and Marrakech results provide the two-climate ranges examined in this study

Table 4. Within-benchmark proxy ranking of the setpoint rules versus the EnergyPlus cooling-demand ranking of the corresponding thermostat schedules for Kenitra

Setpoint case	Proxy reward (Kenitra)	Proxy rank	Simulated cooling, Kenitra (kWh_{th} /m²·yr)	Simulated rank
Optimised 27/30 °C	−2.501	1	36.0	1
Occupancy 26/30 °C	−3.159	2	45.2	3
DQN-derived mean approximation (27/28 °C)	−4.178	3	39.4	2
Fixed 26 °C	−9.960	4	45.3	4

Note: within the simplified benchmark, the four cases are evaluated as setpoint-selection rules. In EnergyPlus, the corresponding setpoints are implemented as thermostat schedules. The DQN-derived mean approximation is evaluated as a 27/28 °C rule in the benchmark and as the corresponding 27/28 °C schedule in EnergyPlus; it is not an exact replay of a trained DQN policy. The EnergyPlus ranking is identical for Marrakech.

indoor-temperature trajectory, where thermostat control keeps the indoor temperature below the setpoint; in that regime, passive drift rarely crosses the cooling-activation threshold, so few transitions produce distinct thermal outcomes and the reward gradient gives a weak learning signal, consistent with the persistent inter-seed variance observed during training. A further design factor is that the occupied-period action ceiling (27 °C) coincides with the optimised rule's occupied setpoint. This narrows the range over which the DQN can differentiate itself. The decoupling therefore emerges from the joint effect of thermal calibration, action space, and reward formulation.

Together, the two analyses point in the same direction. Inside the benchmark, reward differences exist with very limited realised thermal control: the DQN-derived mean setpoint rule and the three deterministic rules produced no cooling activation or realised thermal effect, while the seed-specific DQN action sequences used at most 1.9% of the available thermal authority. Outside the benchmark, EnergyPlus shows that the schedules do differ in simulated cooling demand, and that the proxy ranking is only partly faithful: correct on the best and worst schedules, but misleading for the intermediate ranking between the DQN-derived mean setpoint schedule and the 26/30 °C occupancy schedule in the EnergyPlus simulations. Neither view alone is sufficient: a within-benchmark reward ranking can hide how little control is exercised, while a full-building simulation alone would miss the proxy-induced misranking. This is the central argument for the proposed audit: it prevents reward quality, calibration quality, available authority, and realised effect from collapsing into a single number. The framework does not claim that RL is inferior to rule-based control; it shows that, for this environment and reward, reward improvement must be audited against simulation-based cooling-energy indicators before any control or savings claim is made.

These findings are subject to three limitations that bound their scope. First, the audit results describe a single 80 m² apartment under two Moroccan climates and one reward formulation; the zero-activation regime reported here is a property of this calibrated benchmark rather than a general feature of DQN control, and it should not be generalised to other buildings, climates, or reward designs. Second, the EnergyPlus stage evaluated the DQN through a rounded 27/28 °C occupancy-conditioned mean setpoint schedule, not an exact hourly replay of any trained policy, so the reported ranking inversion applies to that approximation rather than to a specific trained network. Third, the comparison is comfort-constrained at the prescribed occupied-setpoint level: under the Ideal Loads Air System the occupied cooling setpoints are met at every occupied hour by construction, but this does not establish equivalence of operative temperature, PMV, or PPD across schedules, which would require a separate thermal-comfort analysis.

The methodological contribution, by contrast, is not restricted in the same way. Separating the thermal authority available to a policy from the thermal effect it actually exerts, and checking a benchmark reward ranking against full-building simulation, are transferable to other buildings, climates, reward formulations, and control algorithms.

CONCLUSIONS

This study developed and applied a closed-loop audit for reinforcement-learning and rule-based cooling-setpoint policies, and paired it with simulation-based assessment in a detailed building energy model. The motivating concern was that ranking such policies by benchmark reward implicitly assumes reward improvement reflects improved thermal control and reduced cooling demand. We show that this assumption is structurally fragile.

A lightweight first-order thermal environment was calibrated against EnergyPlus data for an 80 m² Moroccan apartment, reproducing mean indoor-temperature levels on independent holdout windows with sub-degree error. Within this environment, we compared the three trained DQN policies with three deterministic setpoint rules and used the audit to separate the thermal authority available to each policy from the thermal effect it actually exerted.

The audit showed that measurable thermal authority was available during 209 hourly steps. The DQN-derived mean setpoint rule and the three deterministic rules produced no cooling activation or realised thermal effect above the selected threshold. In contrast, the seed-specific DQN action sequences showed limited but non-zero activity for seeds 0 and 1, while seed 2 showed none. The mean setpoint rule therefore masks rare seed-specific control activity, and the zero-activation finding should not be generalised to all trained DQN policies.

Full annual EnergyPlus simulations under two contrasting Moroccan climates confirmed that simulated cooling differences between the schedules do exist, in both the milder and the hotter climate. Relative to the fixed 26 °C setpoint, the optimised occupancy-aware schedule (27/30 °C) reduced the specific thermal cooling demand by 20.5% in Kenitra and 17.0% in Marrakech. These reductions provide a simulation-based basis for assessing the energy-saving relevance of intelligent cooling-control strategies in climate-sensitive residential buildings.

The quantitative findings are specific to the studied 80 m² apartment, the prescribed thermostat schedules, and the two Moroccan TMYx climates, and they do not establish performance in other buildings, HVAC configurations, occupancy patterns, or climates. The broader contribution is methodological: the proposed audit distinguishes reward optimisation, available thermal authority, realised thermal control, and simulation-based cooling demand.

Acknowledgements

The authors thank the Laboratory of Advanced Systems Engineering, National School of Applied Sciences, Ibn Tofail University, for institutional support.

REFERENCES

1. Afram, A., Janabi-Sharifi, F. (2014). Theory and applications of HVAC control systems: A review of model predictive control. *Building and Environment, 72*, 343–355. https://doi.org/10.1016/j.buildenv.2013.11.016
2. Al Sayed, K., Boodi, A., Broujeny, R. S., Beddiar, K. (2024). Reinforcement learning for HVAC control in intelligent buildings. *Journal of Building Engineering, 95*, 110085. https://doi.org/10.1016/j.jobe.2024.110085
3. Alalouch, C., Al-Saadi, S., AlWaer, H., Al-Khaled, K. (2019). Energy saving potential for residential buildings in hot climates: The case of Oman. *Sustainable Cities and Society, 46*, 101442. https://doi.org/10.1016/j.scs.2019.101442
4. Anand, P., Cheong, D., Sekhar, C. (2022). A review of occupancy-based building energy and IEQ controls. *Science of the Total Environment, 804*, 150249. https://doi.org/10.1016/j.scitotenv.2021.150249
5. ANSI/ASHRAE. (2020). ANSI/ASHRAE Standard 55-2020: Thermal environmental conditions for human occupancy. American Society of Heating, Refrigerating and Air-Conditioning Engineers.
6. Azuatalam, D., Lee, W. L., De Nijs, F., Liebman, A. (2020). Reinforcement learning for whole-building HVAC control and demand response. *Energy and AI, 2*, 100020. https://doi.org/10.1016/j.egyai.2020.100020
7. 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*(5), 586–610. https://doi.org/10.1080/19401493.2021.1986574
8. Bonnema, E., Allen, A., Lee, E., Mitchell, M. S., Meyer, R. (2025). Flexible integration of diverse HVAC technologies in EnergyPlus via Python-enabled workflows (No. NREL/TP-5500-85628). National Renewable Energy Laboratory (NREL). https://doi.org/10.2172/2567047
9. Boumlik, K., Belarbi, R., Ahachad, M., Mahdaoui, M., Radoine, H., Krarti, M. (2024). Design optimization of energy-efficient residential buildings in Morocco. *Buildings, 14*(12), 3915. https://doi.

org/10.3390/buildings14123915

10. Boutahri, Y., Tilioua, A. (2025). Reinforcement learning for HVAC control and energy efficiency in residential buildings with BOPTEST simulations and real case validation. *Discover Computing, 28*(1), 45. https://doi.org/10.1007/s10791-025-09544-y
11. Chen, C., An, J., Wang, C., Duan, X., Lu, S., Che, H., Wu, Y., Yan, D. (2023). Deep reinforcement learning-based joint optimization control of indoor temperature and relative humidity in office buildings. *Buildings, 13*(2), 438. https://doi.org/10.3390/buildings13020438
12. 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*(4), 319–331. https://doi.org/10.1016/S0378-7788(00)00114-6
13. Dong, J., Winstead, C., Nutaro, J., Kuruganti, T. (2018). Occupancy-based HVAC control with short-term occupancy prediction algorithms. *Energies, 11*(9), 2427. https://doi.org/10.3390/en11092427
14. Drif, N., Ayrir, W., Ait Errouhi, A., Benhaddou, D. (2025). Data-driven classroom occupancy modeling using QR-based attendance records. In: Proceedings of the 2025 International Conference on Advanced Technologies and Interdisciplinary Innovation (ICAT2I) (pp. 1–5). IEEE. https://doi.org/10.1109/ICAT2I69744.2025.11472694
15. 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(9), 2419–2468. https://doi.org/10.1007/s10994-021-05961-4
16. El Asri, N., Abdou, N., Mharzi, M., Maghnouj, A. (2023). Moroccan public buildings and the RTCM: Insights into compliance, energy performance, and regulation improvement. *Energies, 16*(18), 6496. https://doi.org/10.3390/en16186496
17. Esrafilian-Najafabadi, M., Haghighat, F. (2021). Occupancy-based HVAC control using deep learning algorithms for estimating online preconditioning time in residential buildings. *Energy and Buildings, 252*, 111377. https://doi.org/10.1016/j.enbuild.2021.111377
18. Fang, X., Gong, G., Li, G., Chun, L., Peng, P., Li, W., Shi, X., Chen, X. (2022). Deep reinforcement learning optimal control strategy for temperature setpoint reset. *Applied Thermal Engineering, 212*, 118552. https://doi.org/10.1016/j.applthermaleng.2022.118552
19. International Organization for Standardization. (2005). ISO 7730:2005: Ergonomics of the thermal environment — Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria. International Organization for Standardization.
20. Jiang, Z., Risbeck, M. J., Ramamurti, V., Murugesan, S., Amores, J., Zhang, C., Lee, Y. M., Drees, K. H. (2021). Building HVAC control with reinforcement learning for reduction of energy cost and demand charge. *Energy and Buildings, 239*, 110833. https://doi.org/10.1016/j.enbuild.2021.110833
21. Jiménez-Raboso, J., Campoy-Nieves, A., Manjavacas-Lucas, A., Gómez-Romero, J., Molina-Solana, M. (2021). Sinergym: A building simulation and control framework for training reinforcement learning agents. In: Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys) (pp. 319–323). https://doi.org/10.1145/3486611.3486709
22. Lachir, A., Noufid, A. (2024). Assessment of building design strategies to enhance energy efficiency and thermal comfort: Case study in Morocco's climate zones. *Journal of Building Physics, 48*(2), 197–221. https://doi.org/10.1177/17442591241266608
23. Manjavacas, A., Campoy-Nieves, A., Jiménez-Raboso, J., Molina-Solana, M., Gómez-Romero, J. (2024). An experimental evaluation of deep reinforcement learning algorithms for HVAC control. *Artificial Intelligence Review, 57*, 173. https://doi.org/10.1007/s10462-024-10819-x
24. Michailidis, P., Michailidis, I., Vamvakas, D., Kosmatopoulos, E. (2023). Model-free HVAC control in buildings: A review. *Energies, 16*(20), 7124. https://doi.org/10.3390/en16207124
25. M'lahfi, B., Amegouz, D., El Qandil, M. (2020). A new approach for the mandatory application of the thermal regulation of construction in future Moroccan buildings. *SN Applied Sciences, 2*(10), 1656. https://doi.org/10.1007/s42452-020-03367-w
26. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature, 518*, 529–533. https://doi.org/10.1038/nature14236
27. Ramaraj, S., Sparn, B. (2025). Evaluating thermostats' deadbands using HVAC hardware-in-the-loop experiment for advanced control strategies. National Renewable Energy Laboratory (NREL); Northeastern University; National Energy Technology Laboratory. https://doi.org/10.2172/2588191
28. Sbabou, D., Ayrir, W., Ait Errouhi, A., Benhaddou, D. (2025). Optimizing HVAC systems through

temperature prediction: A regression model approach for energy efficiency in buildings. In: Proceedings of the 2025 4th International Conference on Embedded Systems and Artificial Intelligence (ESAI) (pp. 1–9). IEEE. https://doi.org/10.1109/ESAI67033.2025.11438619

29. Taheri, S., Hosseini, P., Razban, A. (2022). Model predictive control of HVAC systems: A state-of-the-art review. *Journal of Building Engineering, 60*, 105067. https://doi.org/10.1016/j.jobe.2022.105067
30. Turley, C., Jacoby, M., Pavlak, G., Henze, G. (2020). Development and evaluation of occupancy-aware HVAC control for residential building energy efficiency and occupant comfort. *Energies, 13*(20), 5396. https://doi.org/10.3390/en13205396
31. Uddin, M. N., Wei, H. H., Chi, H. L., Ni, M. (2021). Influence of occupant behavior for building energy conservation. *Buildings, 11*(2), 41. https://doi.org/10.3390/buildings11020041
32. Yao, Y., Shekhar, D. K. (2021). State of the art review on model predictive control in HVAC. *Building and Environment, 200*, 107952. https://doi.org/10.1016/j.buildenv.2021.107952
33. Zhang, C., Kuppannagari, S. R., Kannan, R., Prasanna, V. K. (2019). Building HVAC scheduling using reinforcement learning via neural network-based model approximation. In: Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys) (pp. 287–296). https://doi.org/10.1145/3360322.3360861