Contents list available at CBIORE journal website

International Journal of Renewable Energy Development

Journal homepage: https://ijred.cbiore.id

Research Article

Multi-objective HVAC control using genetic programming for grid-responsive commercial buildings

Sibtain Waheed* and Shuhong Li

School of Energy & Environment, Southeast University, Nanjing, China 211189, China

Abstract. Commercial buildings are significant energy consumers, with their heating, ventilation, and air conditioning (HVAC) systems being major contributors. Optimizing these systems is crucial for energy conservation, yet advanced artificial intelligence methods like Deep Reinforcement Learning (DRL) often produce opaque black-box solutions. While post-hoc explanation methods can offer some insight, they are often inexact and fail to render the core decision logic fully transparent, hindering trust and practical implementation. This paper presents a novel approach using Genetic Programming (GP) to automatically design HVAC control strategies that are both highly effective and inherently understandable. The novelty of our framework lies in its direct evolution of interpretable, multi-objective control policies that holistically co-optimize energy efficiency, occupant thermal comfort, and integrated Demand Response (DR) for a complex multi-zone system a combination not extensively explored in prior GP-HVAC research. We applied this framework to manage the central air handling unit of a simulated multi-zone office building, enabling it to dynamically adjust key settings like air temperature and fan pressure. Rigorous testing in a validated EnergyPlus simulation environment showed that the GP-designed control policies reduced annual HVAC energy use by 40.9% compared to standard ASHRAE A2006 guidelines, 28.4% against the advanced ASHRAE G36 standard, and a notable 9.3% more than a state-of-the-art DRL controller. These substantial energy savings were achieved while maintaining excellent occupant thermal comfort for 98.8% of occupied hours. Furthermore, the GP controller demonstrated robust performance during Demand Response scenarios, achieving a 72.1% reduction in peak power draw. A key outcome is that these high-performing strategies are expressed in a transparent format allowing direct inspection and understanding. This research establishes Genetic Programming as a compelling method for creating intelligent HVAC controls that are not only efficient and grid-responsive but also transparent, fostering greater confidence in advanced building automation.

Keywords: Genetic Programming; HVAC Control; Energy Efficiency; Transparent Control; Demand Reponses; Building Automation

Received: 20th May 2025; Revised: 7th Sept 2025; Accepted: 4th Oct 2025; Available online: 19th Oct 2025

1 INTRODUCTION

Commercial buildings use a lot of energy, and their heating, ventilation, and air conditioning (HVAC) systems are often the main reason, accounting for about 40-50% of total building energy consumption in many cases (Ghaderian & Veysi, 2021; Kaushik *et al.,* 2022). With growing concerns about climate change, rising energy costs, and the need for smarter power grids, making these systems more efficient is a big deal. Traditional HVAC controls, like those based on fixed rules from standards such as ASHRAE 90.1, work fine but often miss opportunities to save energy because they don't adapt well to changing conditions like weather, occupancy, or peak demand times (Pérez-Lombard *et al.,* 2008; Amer *et al.,* 2024; Yoon *et al.,* 2024). This can lead to wasted energy, uncomfortable indoor spaces, and higher bills.

Over the years, researchers have turned to advanced methods to make HVAC smarter. Model predictive control (MPC) use math models to predict and adjust settings ahead of time, which can cut energy use by 20-30% in some studies (Afroz *et al.,* 2018; Bitar *et al.,* 2024). Then there's artificial intelligence, especially reinforcement learning (RL), where systems learn from trial and error to balance energy savings with comfort (Xie, Ajagekar, & You, 2023; Al Sayed *et al.,* 2024). Deep reinforcement learning (DRL) takes this further by handling complex data, showing promising results in simulations and even real buildings (Lu *et al.,* 2022; Sanzana *et al.,* 2022). But the problem with many AI methods is they're "black-box" models. You get great performance, but it's hard to understand why the system makes certain decisions. This lack of transparency can make building managers hesitant to trust and implement them, especially in critical setups where safety and reliability matter (Pinto *et al.*, 2022; Pinthurat, Surinkaew, & Hredzak, 2024).

To address these limitations, advanced artificial intelligence (AI) techniques have gained prominence. Deep Reinforcement Learning (DRL) has emerged as a leading model-free approach, demonstrating 15-30% energy savings over conventional controls in simulated multi-zone buildings (Yu *et al.,* 2021; Hou *et al.,* 2024). Algorithms like Deep Q-Networks (DQN), Soft Actor-Critic (SAC), and multi-agent variants enable adaptive policies that learn from interactions with the environment, optimizing actions such as SAT and DSP setpoints (Kumar *et al.*, 2025; Niazi *et al.,* 2025; Sun *et al.,* 2025). Recent studies have

* Corresponding author
Email: sibtainwaheed@seu.edu.cn (S. Waheed)

https://doi.org/10.61435/ijred.2025.61410
ISSN: 2252-4940/© 2025.The Author(s). Published by CBIORE

extended DRL to incorporate DR, achieving 40-45% peak load reductions by preemptively adjusting HVAC operations during high-price signals (Kumar *et al.,* 2025; Niazi *et al.,* 2025; Çinar & Abut, 2025). However, DRL's reliance on opaque neural networks poses a major barrier: the "black-box" nature hinders interpretability, trust, and deployment in safety-critical systems (Kargar & Bahamin, 2025). Efforts to enhance transparency through post-hoc Explainable AI (XAI) methods, such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), have provided partial insights into feature importance (Mariano-Hernández *et al.,* 2021; Yao *et al.,* 2024). Yet, these approximations often fail to capture holistic decision logic, leading to incomplete or unreliable explanations (Yan *et al.,* 2016).

Parallel advancements in model-based optimization, particularly Model Predictive Control (MPC), offer structured alternatives. MPC uses physics-based models to forecast and optimize HVAC operations, reporting up to 25% energy savings while integrating DR (Bouabdallaoui *et al.,* 2021; Tomás, Lämmle, & Pfafferott, 2025). Hybrid approaches combining MPC with machine learning further improve robustness to uncertainties (Alimohammadisagvand, Jokisalo, & Sirén, 2018). However, developing accurate models is resource-intensive, and scalability remains a challenge for large buildings (Chaturvedi, Rajasekar, & Natarajan, 2020; Bouabdallaoui *et al.,* 2021).

In the DR domain, rule-based strategies provide transparency but often compromise comfort during load shedding (Cheraghi & Jahangir, 2023; Cho, Lee, & Heo, 2023; Choi *et al.,* 2023). DRL-enhanced DR, while effective, inherits interpretability issues (Ding, Cerpa, & Du, 2025). Emerging VPP concepts aggregate buildings with renewables but require interpretable controls for market participation (Pang *et al*., 2025).

Genetic Programming (GP), an evolutionary computation technique, addresses these gaps by evolving interpretable expression trees or decision rules directly from data (Cpalka, Łapa, & Przybył, 2018; Sipper & Moore, 2020). Prior GP applications in buildings include single-objective optimizations, such as chiller sequencing (yielding 10-20% savings) or thermal comfort in passive designs (Gao *et al*., 2020; Es-sakali *et al.,* 2024). Multi-objective GP using NSGA-II has optimized energy and comfort in residential settings (Pang *et al.,* 2025). However, GP has not been extensively applied to integrated multi-zone HVAC control with DR, nor benchmarked against DRL in renewable-integrated grids.

This study bridges these gaps by proposing a GP framework to evolve transparent, multi-objective policies for AHU control in grid-responsive buildings. Key contributions include:

- Development of a comprehensive GP framework for evolving interpretable multi-objective HVAC control policies with integrated Demand Response capabilities.
- Rigorous evaluation of the evolved GP controllers against both conventional rule-based approaches (ASHRAE 2006 and Guideline 36) and state-of-the-art DRL controllers in a validated EnergyPlus simulation environment.
- Analysis of the evolved control strategies, revealing how GP discovers sophisticated yet transparent operational patterns that effectively balance energy efficiency, comfort, and grid responsiveness.
- Validation of GP as a compelling approach for creating intelligent building controls that are not only efficient and grid-responsive but also transparent and trustworthy.

The remainder of this paper is organized as follows: Section 2 details the methodology, including the benchmark building scenario, simulation environment, GP Figure framework, and baseline controllers. Section 3 presents the results and discussion, analyzing the evolutionary process, comparative performance, and characteristics of the evolved policies., followed by concluding remarks in Section 4.

2 Methodology

This section demonstrates the comprehensive methodology developed and employed for the direct evolution, simulation, and rigorous evaluation of interpretable, multi-objective HVAC control policies using Genetic Programming (GP), with a specific focus on integrated Demand Response (DR) capabilities. The overall research process is visually summarized in Fig.1.

We commence by detailing the benchmark building scenario and the high-fidelity simulation environment. Subsequently, the GP-based control strategy formulation is presented, encompassing its representation, multi-objective fitness evaluation, and evolutionary algorithm configuration. we describe the implementation of state-of-the-art Deep Reinforcement Learning (DRL) and established ASHRAE standard controllers, which serve as baselines for comparative analysis, along with the key performance indicators used for evaluation.

2.1 Benchmark Scenario and Simulation Environment

To ensure a realistic and challenging testbed for the control algorithms, a high-fidelity simulation environment was meticulously constructed. This environment leverages EnergyPlus (Version 9.6.0) for dynamic building thermal and HVAC system simulation. The control algorithms, including the novel GP framework and baseline controllers, were implemented in Python (Version 3.9), interfacing with EnergyPlus via the Functional Mock-up Interface (FMI) standard.

2.1.1 Building Model

The architectural testbed is a representative five-zone commercial office building, geometrically and

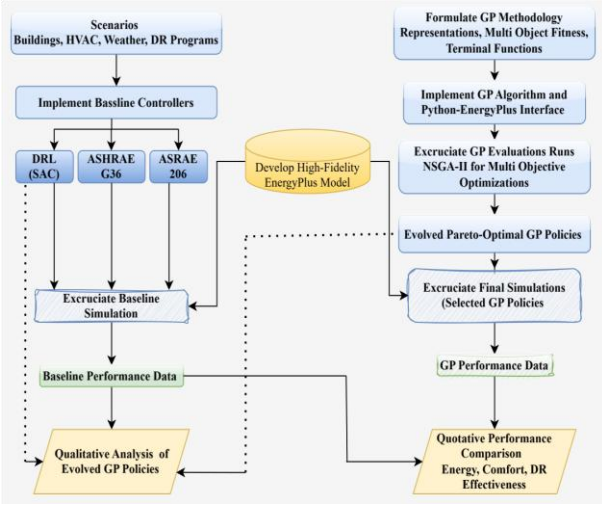

Fig. 1 Methodological Framework

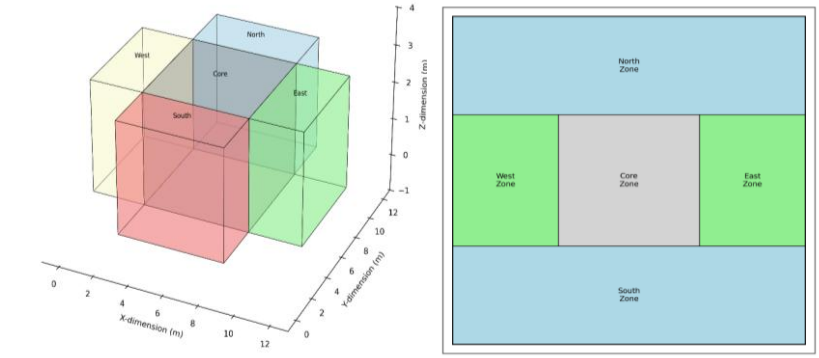

Fig. 2 Building Schematic diagram

thermodynamically adapted from the U.S. Department of Energy (DOE) Commercial Reference Building specification for a Medium Office (New Construction, Post-1980, Climate Zone 5A equivalent) (Yu *et al.*, 2021). The model features a total conditioned floor area of approximately 550 m², distributed across one thermally distinct interior core zone and four perimeter zones (North, East, South, and West), each subject to varying solar exposures and envelope heat transfer characteristics. A schematic plan view illustrating the building layout and zonal configuration is presented in Fig.2. Detailed construction assemblies for walls, roof, floor, and fenestration, along with their respective thermal properties (U-values, R-values, thermal mass characteristics), adhere to the reference building specifications designed to meet ASHRAE Standard 90.1-2019. Air infiltration rates are modeled based on ASHRAE standards for air changes per hour (ACH).

2.1.2 HVAC System Model

The building model is equipped with a centralized Variable Air Volume (VAV) Air Handling Unit (AHU) that conditions and distributes air to the five thermal zones. A high-level overview illustrating the primary functional blocks of the HVAC system and the main interaction points with the Genetic Programming (GP) controller is provided in Fig.3.

Delving into the specifics of the air-side system, the AHU, whose internal schematic and detailed GP control intervention points are depicted in Fig.4, comprises an outdoor air economizer section, a chilled water cooling coil, and a variable-speed supply fan. The economizer operation is based on a dry-bulb temperature comparison between outdoor and return air, with minimum outdoor air ventilation rates continuously maintained according to ASHRAE Standard 62.1-2019 during occupied periods. The cooling coil is supplied with chilled water from an electric water-cooled chiller plant. This plant includes the primary chiller unit, an open-loop cooling tower for heat rejection, and associated variable-speed primary and secondary chilled water pumps, as well as condenser water pumps. The operational logic and setpoints for this chiller plant, particularly the chiller supply water temperature, are influenced by the evolved GP policies as detailed in Section 2.2. Air distribution to the conditioned spaces is managed by five pressure-independent VAV terminal units, one serving each thermal zone. For baseline operations and scenarios within this study, the VAV terminal units modulate their dampers to meet zone-specific cooling demands based on standard ASHRAE 2006 control sequences; The GP controller focuses on optimizing the

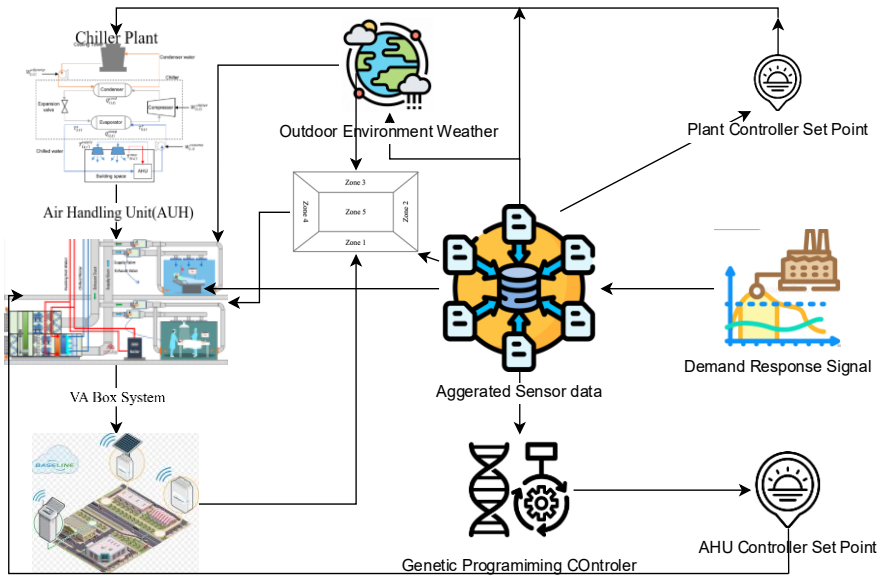

Fig. 3 High-Level HVAC System Overview with GP Controller Interaction

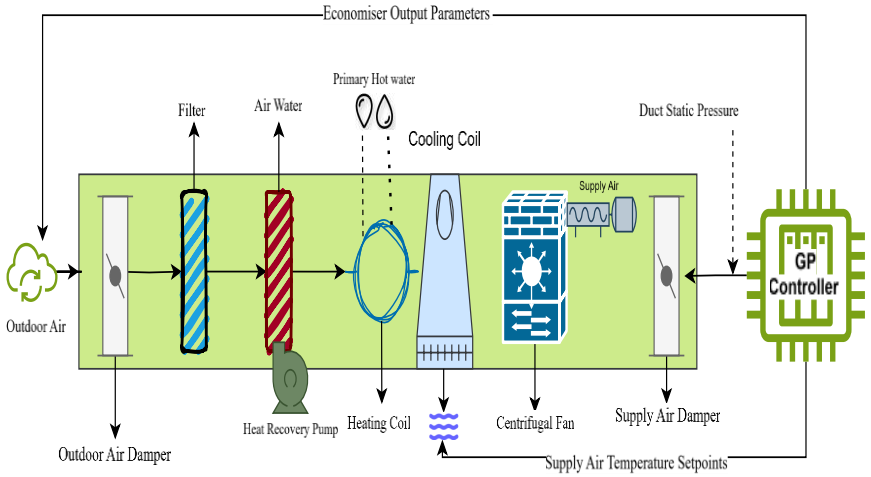

Fig. 5 Detailed AHU Schematic with GP Control Points

central AHU and plant operation, while the VAV boxes react to the supplied air conditions and zone loads according to this established baseline logic.The performance characteristics fan curves, coil capacity curves, chiller COP curves) and nominal design parameters for all primary HVAC components are modeled using standard EnergyPlus objects and empirically-derived performance curves representative of typical commercial equipment. Specific details regarding these component models are comprehensively tabulated in Appendix Table A1.

2.1.3 Operational Conditions and Location

All simulations were conducted for a continuous one-month period representative of a significant cooling season, specifically July. The meteorological conditions are based on a Typical Meteorological Year 3 (TMY3) weather file for Turin, Italy (EPW file: ITA_Torino-Caselle.160590_TMYx.epw), providing hourly data for temperature, humidity, solar radiation, and wind conditions. Standardized commercial office operational schedules were implemented for weekday (Monday-Friday) occupancy (08:00-19:00 peak, with scheduled ramp-up/down), lighting (based on ASHRAE 90.1-2019 Lighting Power Density allowances), and miscellaneous equipment loads (plug loads). Weekend operation assumes an unoccupied building state. The occupied period thermostat cooling setpoint for all zones is maintained at 24°C, with a heating setpoint of 21°C (though active heating demand is minimal during the simulated July period). The central HVAC system operates from 06:00 to 19:00 on weekdays, allowing for a pre-cooling period before nominal occupancy begins. To evaluate DR capabilities, a simulated DR program was integrated. This program introduces a high electricity price signal, three times the baseload electricity price, during peak demand hours (14:00-17:00) on selected high-load weekdays within the simulation month. This DR signal serves as an explicit input to the intelligent controllers (GP and DRL), prompting them to modulate HVAC energy consumption.

2.2 Genetic Programming (GP) Framework for HVAC Control

This research proposes a GP framework to directly evolve interpretable, multi-objective control policies for the AHU. The GP aims to identify policies that holistically optimize energy efficiency, occupant thermal comfort, and effective participation in DR events. The iterative interaction of the GP algorithm with the EnergyPlus simulation environment during the evolutionary process is conceptually illustrated in Fig.5.

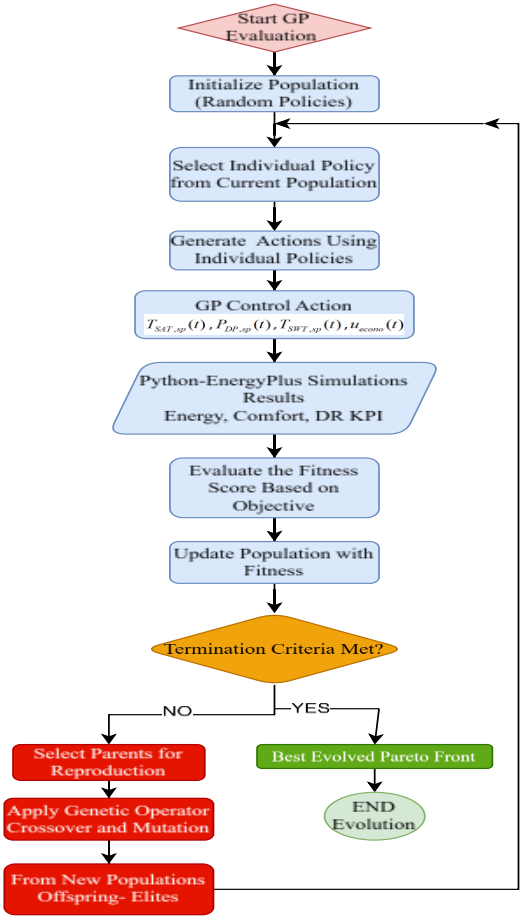

Fig. 4 GP Interaction with Simulation Environment

2.2.1 Problem Formulation for GP

The GP-evolved policies are responsible for determining the following primary AHU control outputs at each discrete control timestep, Δt (set to 15 minutes):

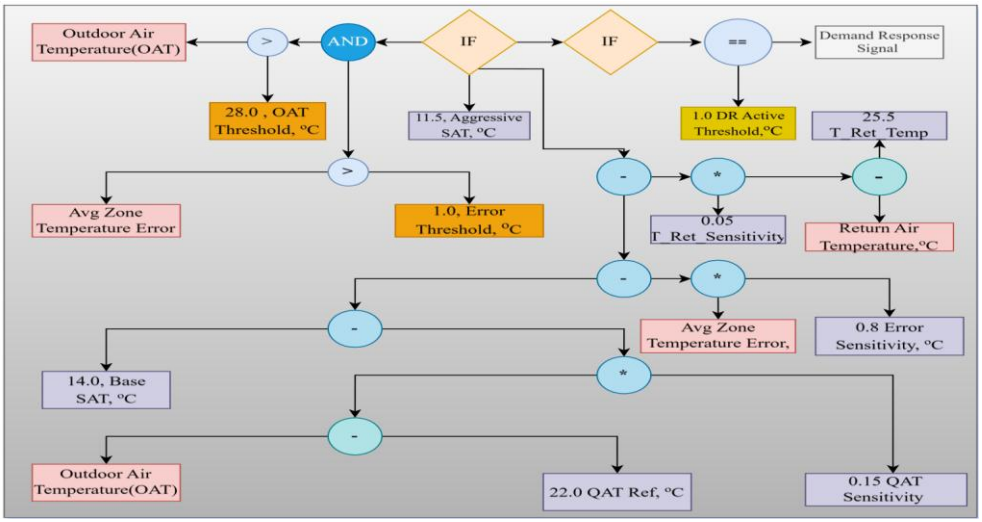

Fig. 6 Evolved GP Policy Tree

- AHU Supply Air Temperature Setpoint $T_{SAT,sp}(t)$,: The target temperature for air leaving the AHU cooling coil.
- AHU Duct Static Pressure Setpoint $P_{DP,sp}(t)$: The target static pressure to be maintained in the main supply duct by the variable-speed fan.
- Chiller Supply Water Temperature Setpoint $T_{SWT,sp}(t)$: The target temperature for water leaving the chiller.
- Economizer Control Parameter $u_{econo}(t)$: A parameter guiding economizer operation, such as a maximum allowable mixed air temperature or a direct outdoor air fraction command, depending on the chosen GP representation.

The set of input variables available to the GP for constructing these control policies encompasses real-time and predicted environmental conditions, building thermal states, and operational signals. A comprehensive list of all input terminals and control outputs, along with their descriptions and units, is provided in Appendix Table B.1. and include current outdoor air temperature T_{OAT}, predicted outdoor air temperatures for the next [1, 2, and 4] hours ($T_{OAT,pred+1h}$, etc.), average and maximum zone air temperatures ($T_{zone,avg}$, $T_{zone,max}$), maximum deviation from zone cooling setpoint ($\Delta T_{zone,max\_dev}$), current supply air temperature ($T_{SAT,act}$), return air temperature (T_{Ret}), an indicator of maximum VAV box damper position across zones ($VAV_{damp,crit}$), current time of day, day of the week, and the active Demand Response signal ($S_{DR}(t)$).

2.2.2 GP Representation

The GP representation uses a tree-based structure where each control output is determined by a separate expression tree as shown in Fig.6. These trees combine input variables (terminals) through mathematical operations (addition, subtraction, multiplication, division), comparison operators (greater than, less than, equal to), and conditional logic (if-then-else). This representation allows for the evolution of complex, non-linear control strategies while maintaining human interpretability.

2.2.3 Fitness Evaluation and Multi-Objective Optimization

The fitness of each GP individual evolved policy is evaluated based on its performance over the simulation period one representative week in July including DR events, or the full month. The operational interaction between a candidate GP policy and the simulated EnergyPlus environment over a typical 24-hour cycle is depicted in Fig.7, the Environment light green bar represents the continuous building simulation. During Fixed Action Periods (hours 00:00-06:00 for night setback and 19:00-23:00 for system off), pre-defined control actions are applied. During the GP Control Active period (light blue bar, 06:00-19:00, indicated by green vertical lines on the timeline), the evolved GP policy is engaged. At each control timestep within this active period, the GP policy receives Observations (Env. to GP) (brown downward arrows) from the EnergyPlus environment and, based on its evolved logic, issues Actions (GP to Env.) (blue upward arrows) back to the simulation to control the AHU. The performance (energy, comfort, DR KPIs) resulting from these actions over the evaluation period T_{eval} is then used to calculate the objective function values: J_{Energy}, $J_{Comfort}$ and J_{DR}.

- *Integrated HVAC Energy Consumption* **(J_{Energy}):** This objective from equation (1) reflects the total electrical energy consumed by the chiller plant (chiller, cooling tower fans, pumps), the AHU supply fan, and any auxiliary HVAC components over the evaluation period T_{eval}.

$$J_{Energy} = \int_0^{T_{eval}} \left(P_{chiller\_plant}(t, u_{GP}(t), x(t)) + P_{fan}(t, u_{GP}(t), x(t)) \right) dt \quad (1)$$

Here, $P_{chiller\_plant}(t)$ and $P_{fan}(t)$ are the instantaneous power demands (kW) of the chiller plant and AHU fan, respectively. These are functions of the GP-dictated control actions $U_{GP}(t)$ and the overall system state $x(t)$. The integral is numerically approximated from the discrete-time simulation outputs.

- *Aggregated Thermal Discomfort* **($J_{Comfort}$):** This objective from equation (2) quantifies occupant dissatisfaction due to deviations from the defined thermal comfort band. It is formulated as the sum of Time-Integrated Absolute Error (TIAE) or Integrated Squared Error (ISE) of zone temperatures from the comfort band limits ($T_{LL,i} = T_{setpoint,i} - \delta T_{comfort}$ and

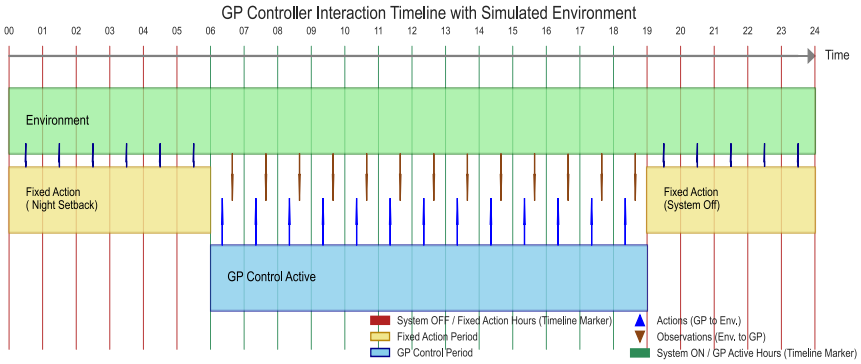

Fig. 7 GP Interaction with Simulation Environment

$T_{UL,i} = T_{setpoint,i} + \delta T_{comfort}$ for zone i, with $T_{setpoint,i} = 24°C$ and $\delta T_{comfort} = 1°C$ during occupied hours $Occ_i(t)$.

-

$$J_{Comfort} = \sum_{i=1}^{N_z} \int_0^{T_{eval}} Occ_i(t) \cdot \left(\begin{matrix} \max(0, T_{zone,i}(t, u_{GP}(t), x(t)) - T_{UL,i}) \\ + \max(0, T_{LL,i} - T_{zone,i}(t, u_{GP}(t), x(t))) \end{matrix} \right) dt \quad (°C \cdot hr) \tag{2}$$

The evolutionary search for optimal control policies is guided by a multi-objective fitness function, designed to concurrently satisfy performance criteria related to energy consumption, thermal comfort, and DR participation. Each candidate GP policy is evaluated by deploying it in the Energy Plus simulation for a representative period T_{eval} (carefully selected week in July encompassing diverse load conditions and DR events). The performance is quantified by the following three objective functions, all of which are to be minimized:

- *Demand Response Performance Deficit* **(J_{DR}):** This objective penalizes the failure to meet DR targets or maximizes the benefit from DR participation in equation (3) and (4). One formulation is to minimize the average HVAC power consumption during DR event periods, T_{DR}, relative to a baseline power consumption, $P_{baseline,DR}(t)$, or to penalize exceeding a DR power cap $P_{cap,DR}$:

$$J_{DR} = \frac{1}{|T_{DR}|} \int_{t \in T_{DR}} P_{HVAC,total}(t, u_{GP}(t), x(t)) dt \quad (\text{kW}) \tag{3}$$

$$J_{DR} = \int_{t \in T_{DR}} \max(0, P_{HVAC,total}(t, u_{GP}(t), x(t)) - P_{cap,DR}) dt \tag{4}$$

To address these multiple, often conflicting, objectives, the Non-dominated Sorting Genetic Algorithm II (NSGA-II) (Ghaderian & Veysi, 2021) is employed as the evolutionary search mechanism. NSGA-II identifies a set of Pareto-optimal solutions, representing the best achievable trade-offs between J_{Energy}, $J_{Comfort}$, and J_{DR}. From this final Pareto front, a single representative GP policy was selected for the detailed comparative analysis against the baseline controllers. This selection was based on a balanced performance criterion, specifically targeting the solution that minimized the Euclidean distance to the utopia point (0, 0, 0) in the normalized three-objective space. This approach was chosen to identify the policy offering the most compelling compromise across all three objectives, rather than a policy that might excel in one objective at the significant expense of others.

Furthermore, to address the potential for overfitting, it is important to note that for this study, both the evolution and final evaluation of the GP policies were conducted on the representative month of July. This approach was deliberately chosen to create a controlled and reproducible 'level playing field' for directly comparing the learning capabilities of GP against the DRL agent, ensuring both were optimized under identical conditions. The critical question of generalization to unseen weather data is a primary focus for future work, as discussed in Section 4.

2.2.4 Evolutionary Algorithm Configuration

The GP system was implemented using the DEAP (Distributed Evolutionary Algorithms in Python) library, Version 1.3.1. Key evolutionary parameters were set as follows: population size of 200, number of generations set to 100, tournament selection with tournament size of 3. The entire evolution process, conducted on an Intel Core i9-12900K CPU and 32GB RAM, took approximately 30 hours to complete. Genetic operators included one-point subtree crossover with probability 0.7 and subtree mutation with probability 0.2. Additionally, point mutation for ephemeral random constants was applied with a probability of 0.1. The initial population was generated using the ramped half-and-half method with tree depths ranging from 2 to 6, and a maximum tree depth of 12 was enforced to prevent excessive bloating. Elitism, preserving the top 2% non-dominated solutions, was incorporated to ensure convergence towards high-quality solutions. These parameters were chosen based on a combination of established practices in GP literature for complex control problems (Cpalka, Łapa & Przybył, 2018) and a series of preliminary tuning experiments. The population size and number of generations were selected to provide a sufficient search diversity and convergence time, while remaining within feasible computational limits for the high-fidelity simulation environment. The crossover and mutation rates were set to standard values that encourage a balance between exploration of new solutions and exploitation of high-performing genetic material.

2.3 Baseline Controller Implementations

For a comprehensive performance benchmark, the evolved GP policies were compared against three distinct baseline controllers, simulated under identical environmental and operational conditions.

Table 1
Comprehensive Performance Comparison of Control Strategies

Performance Indicator	Unit	Evolved GP Policy	DRL (SAC) Baseline	ASHRAE G36 Baseline	ASHRAE A2006 Baseline
Energy Efficiency					
Total HVAC Energy Consumption	kWh	6,800	7,500	9,500	11,500
Chiller Plant Energy Consumption	kWh	4,080	4,500	5,800	7,000
AHU Fan Energy Consumption	kWh	1,840	2,025	2,470	3,000
% Savings vs. A2006 (Total)	%	40.9%	34.8%	17.4%	—
% Savings vs. G36 (Total)	%	28.4%	21.1%	—	-21.1% (Worse)
Thermal Comfort					
ZAT Violation Degree-Hours	°C·hr	75	80	150	280
Occupied Hours in Comfort Band	%	98.8%	98.5%	96.5%	93.0%
Demand Response Effectiveness					
Average Peak Load Reduction	kW	13.7	13.2	1.0*	0.2*
% Peak Load Reduction	%	72.1%	69.5%	~5%*	~1%*
Total Energy Saved/Shifted	kWh	40.5	38.8	2.5*	0.5*

2.3.1 Deep Reinforcement Learning (DRL) Baseline

A DRL agent based on the Soft Actor-Critic (SAC) algorithm was developed as a state-of-the-art learning-based benchmark.

- *State and Action Spaces*: The DRL agent utilized the same state observation space as the GP terminals (Appendix Table B.1). Its continuous action space corresponded to the four AHU control outputs ($T_{SAT,sp}, P_{DP,sp}, T_{SWT,sp}, u_{econo}$), normalized to [-1, 1] and subsequently scaled to physical operational limits.
- *Reward Function*: The instantaneous reward R_t as shown in equation (5) was formulated to align with the GP's multi-objective nature, typically as a negatively weighted sum of penalties reflecting energy consumption ($P_{HVAC,t}$), thermal discomfort ($D_{comfort,t}$), and DR non-compliance ($D_{DR,t}$):

$$R_t = -\begin{bmatrix} w_E \cdot P_{HVAC,t} \\ +w_C \cdot \sum_{i=1}^{N_z} Occ_i(t) \cdot D_{comfort,i,t+1} \\ +w_{DR} \cdot S_{DR}(t) \cdot D_{DR,t} \end{bmatrix} \quad (5)$$

  The terms *comfort* I ,t+1 and $\boldsymbol{D_{DR,t}}$ are analogous to the integrands in $\boldsymbol{J_{Comfort}}$ and $\boldsymbol{J_{DR}}$ respectively, evaluated at time t or t+1. The weights ($\boldsymbol{w_E, w_C, w_{DR}}$) were determined through an iterative tuning process. A grid search over a range of plausible weight ratios was conducted in short, preliminary training runs (100k timesteps each). The set of weights that demonstrated the most stable learning curve and resulted in agents achieving a balanced performance across energy, comfort, and DR objectives during these initial runs was selected for the full, long-duration training.
- *Network Architecture and Training:* Both actor and critic networks in SAC were implemented as fully connected multi-layer perceptions (MLPs) with two hidden layers of 256 neurons each, employing ReLU activation functions, and appropriate output activations (Tanh for bounded actions)]. The agent was trained for [2×10^6 simulation timesteps] using an experience replay buffer of size [10^5]. This training process took approximately 48 hours on the same computational hardware. Further details on DRL hyperparameters (learning rates, discount factor γ, target smoothing τ, entropy coefficient α) are provided in Appendix C.1 *(DRL Agent Hyperparameters)*.

2.3.2 ASHRAE Guideline 36 Baseline

The advanced rule-based control sequences specified in ASHRAE Guideline 36-2021 (Yoon *et al.*, 2024) were implemented as a high-performance conventional baseline. This encompassed:

- Dynamic supply air temperature (SAT) reset using the Trim & Respond algorithm, responsive to aggregate zone cooling demand and outdoor air temperature.
- Dynamic duct static pressure (DP) reset using Trim & Respond logic based on VAV terminal damper positions, thereby minimizing fan energy while ensuring terminal authority.
- Economizer control based on differential dry-bulb or enthalpy comparison, with enforcement of minimum ventilation requirements.
- Chiller plant sequencing and chilled water temperature reset consistent with standard Guideline 36 specifications.

As illustrated in Fig.8a, the SAT and DP reset logic ensures supervisory-level efficiency by adjusting setpoints dynamically in response to load conditions and zone demands. A more conventional rule-based controller consistent with ASHRAE 2006 sequences was also considered as a fundamental baseline. This configuration features a simpler SAT linear reset tied directly to outdoor air temperature, potentially fixed or less adaptive DP setpoints, and standard economizer control. The VAV terminal unit control logic (shown in Fig.8b) was applied consistently across all baseline scenarios, including those served by the GP-controlled AHU. Each pressure-independent VAV box modulated damper position to maintain the local zone temperature setpoint, subject to minimum ventilation requirements.

2.4 Performance Evaluation Metrics

The comparative performance of the evolved Genetic Programming (GP) policies and the baseline controllers was

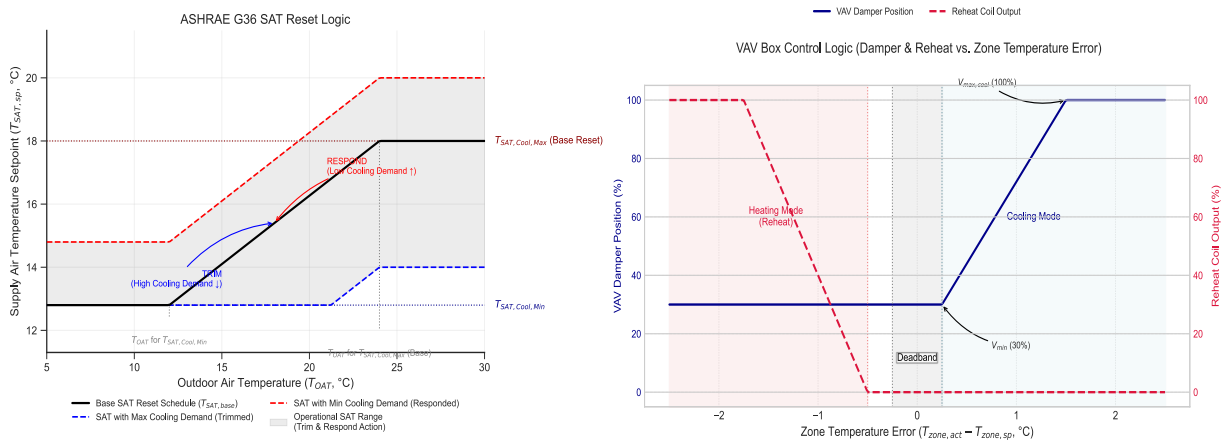

Fig. 8 ASHRAE Guideline 36 SAT and VAV Box Control Logic

quantitatively assessed using a suite of Key Performance Indicators (KPIs), aggregated over the entire simulation period (July). These indicators were selected to provide a holistic evaluation across energy efficiency, occupant thermal comfort, Demand Response (DR) effectiveness, and controller characteristics.

- Total HVAC Energy Consumption ($E_{HVAC,total}$): This primary energy KPI from equation (6) represents the sum of all electrical energy (in kWh) consumed by the HVAC system components, including the chiller plant (chiller, cooling tower fans, pumps) and the Air Handling Unit (AHU) fan, over the simulation period T_{sim}:

$$E_{HVAC,total} = \int_0^{T_{sim}} \left(P_{chiller\_plant}(t) + P_{fan}(t) + P_{aux\_pumps}(t) \right) dt \quad (6)$$

where $P_{chiller\_plant}(t)$, $P_{fan}(t)$ and $P_{aux\_pumps}(t)$ are the instantaneous power demands of the respective components.

- Thermal Comfort - ZAT Violation Degree-Hours (VDH_{ZAT}): This metric quantifies the integrated magnitude and duration of thermal discomfort in equation (7). It is calculated as the sum of absolute temperature deviations outside the defined comfort band (23°C - 25°C) during occupied hours ($Occ_i(t) = 1$) for all N_z zones:

$$VDH_{ZAT} = \sum_{i=1}^{N_z} \int_0^{T_{sim}} Occ_i(t) \cdot \begin{pmatrix} \max(0, T_{zone,i}(t) - T_{UL,i}) \\ + \max(0, T_{LL,i} - T_{zone,i}(t)) \end{pmatrix} dt \quad (7)$$

- Thermal Comfort - Percentage of Occupied Hours within Comfort Band ($\%T_{comfort}$): This provides an intuitive measure of comfort, representing the percentage of total occupied hours during which all monitored zone temperatures were maintained within the [23°C - 25°C] comfort band.

- Demand Response - Average Peak Load Reduction ($\Delta P_{DR,peak}$): This KPI (in kW) measures the average reduction in total HVAC power consumption during active DR event periods compared to a defined baseline power consumption level that would have occurred without DR intervention (average power during similar non-DR peak hours or a simulated "no-DR" scenario).

- Demand Response - Total Energy Saved/Shifted ($\Delta P_{DR,peak}$)): This metric (in kWh) quantifies the total net reduction or shifting of energy consumption achieved specifically during all DR event periods throughout the simulation month, calculated by comparing the actual energy consumed during DR events with the energy that would have been consumed under a non-DR operational baseline.

- GP Policy Complexity: ΔE_{DR} To assess the interpretability and conciseness of the evolved solutions, the complexity of the final selected GP policy is quantified by the total number of nodes (functions and terminals) in its constituent expression tree(s).

3 Results and discussion

This section presents the empirical findings from the application of the proposed Genetic Programming (GP) framework and its comprehensive comparative evaluation against established baseline controllers. The analysis begins with an examination of the GP evolutionary process and the characteristics of the resultant policies, followed by a detailed benchmarking of performance across key metrics including energy efficiency, thermal comfort, and Demand Response (DR) effectiveness, and concludes with a statistical validation of the observed performance differentials.

3.1 GP Evolutionary Process and Characteristics of Evolved Policies

The foundation of the GP framework's success rests on the efficacy of its learning process. Before analyzing the final controller's performance, it is crucial to validate that the evolutionary algorithm effectively navigated the vast search space to discover and refine superior control policies. This ensures the final result is the product of a robust optimization process, not a random outcome.

Fig.9 provides the visual evidence of this successful learning journey over 100 generations. A detailed analysis of Fig.9a tracks the convergence of the three primary objective components for the best-performing individual in each generation, and All objectives, which are formulated for minimization, exhibit a clear and significant improvement. The initial, randomly generated policies perform poorly, as shown by the high penalty scores at generation 0. However, substantial reductions in both the (green line) and (orange line) penalties are observed within the initial 50 generations. This signifies that the best individuals progressively learned to achieve better thermal comfort and more effective Demand Response performance. Notably, the (crimson line) penalty stabilizes at a minimized value relatively early in the process. This suggests the GP algorithm quickly identified a baseline for energy-

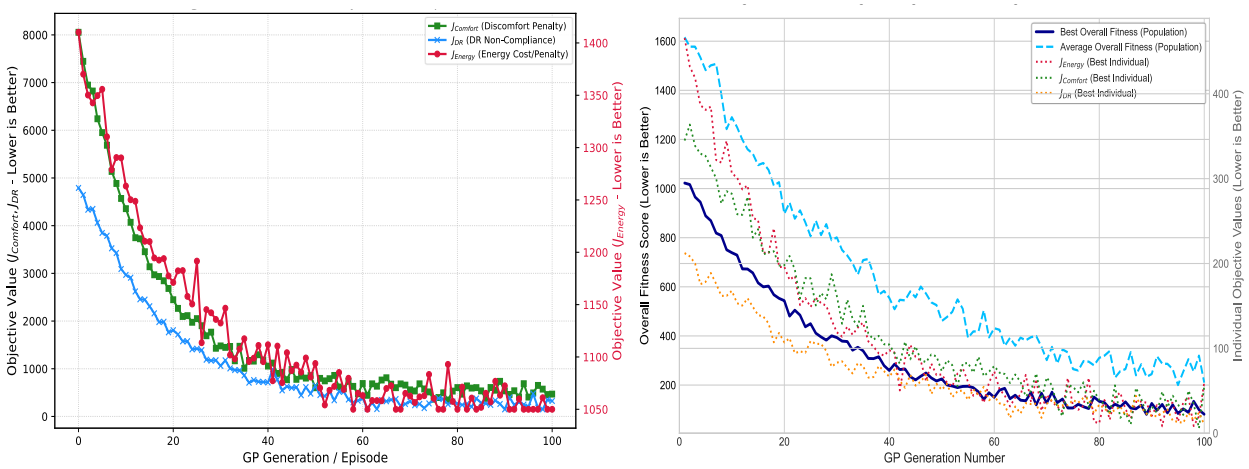

Fig. 9 Programming Evolutionary Process and Convergence: (a) Convergence of the individual fitness objective components for the best individual; (b) Convergence of the overall fitness score for both the best individual and the population average.

efficient operation and then dedicated the majority of its subsequent evolutionary effort to fine-tuning the more complex and dynamic trade-offs between occupant comfort and grid responsiveness.

Further illustrating the overall optimization progress, Fig.9b shows the performance of both the elite individuals and the population as a whole. A consistent downward trajectory is observed for both the best and average overall fitness scores, particularly within the initial 40-60 generations. This confirms successful population-wide learning and optimization towards better aggregate solutions. The narrowing gap between the two lines is particularly important, as it indicates that beneficial genetic material was being successfully distributed throughout the population, raising the quality of the entire gene pool. This robust, dual-level convergence at both the individual objective level and the aggregate population level validates the GP framework as a potent and reliable methodology for discovering holistically optimized solutions.

3.2 Comparative Performance of Energy, Comfort, and Grid-Responsiveness

Having established the robust convergence of the evolutionary process in the preceding section, the analysis now shifts to the performance of the final, representative GP policy selected from the Pareto front. This controller was rigorously benchmarked against the state-of-the-art DRL agent and the ASHRAE standards under identical operational conditions for the entire simulated month of July. Table 1 provides a comprehensive summary of the key performance indicators (KPIs) across all three primary objectives energy efficiency, thermal comfort, and Demand Response effectiveness.

A primary benchmark for evaluation is energy performance, where a distinct hierarchy between the controllers is immediately apparent from the data in Table 1. The GP policy's total consumption of 6,800 kWh represents a substantial 40.9% reduction over the A2006 baseline and a 28.4% saving over the G36 standard. Most critically, the 9.3% energy saving relative to the state-of-the-art DRL (SAC) baseline highlights its superior energy optimization capabilities. Fig.10a provides a clear, month-long visualization of this hierarchy, showing the GP controller's cumulative energy use consistently tracking below all others. The hourly operational dynamics depicted in the Fig.10b heatmaps offer further insights into how these savings were achieved. Both the GP and DRL controllers effectively curtailed energy use during shoulder periods (06:00-09:00 and 16:00-18:00), as indicated by the predominantly darker, lower-energy regions. This suggests a more adept part-load operation

Table 2
Statistical Significance of Key Performance Indicator (KPI) Differences Between Controllers (p-values)

Controller Pair 1	Controller Pair 2	Statistical Test Used†	p-value	Significance Level‡
Total HVAC Energy Consumption (kWh)				
Evolved GP Policy	ASHRAE A2006	Paired t-test	< 0.001	***
Evolved GP Policy	ASHRAE G36	Paired t-test	0.002	**
Evolved GP Policy	DRL (SAC)	Paired t-test	0.045	*
DRL (SAC)	ASHRAE G36	Paired t-test	0.005	**
ZAT Violation Degree-Hours (°C·hr)				
Evolved GP Policy	ASHRAE A2006	Paired t-test	< 0.001	***
Evolved GP Policy	ASHRAE G36	Paired t-test	0.008	**
Evolved GP Policy	DRL (SAC)	Paired t-test	0.350	NS
DRL (SAC)	ASHRAE G36	Paired t-test	0.012	*
DR Peak Load Reduction (kW)				
Evolved GP Policy	ASHRAE G36*	Independent t-test	< 0.001	***
Evolved GP Policy	DRL (SAC)	Paired t-test	0.650	NS
DRL (SAC)	ASHRAE G36*	Independent t-test	< 0.001	***

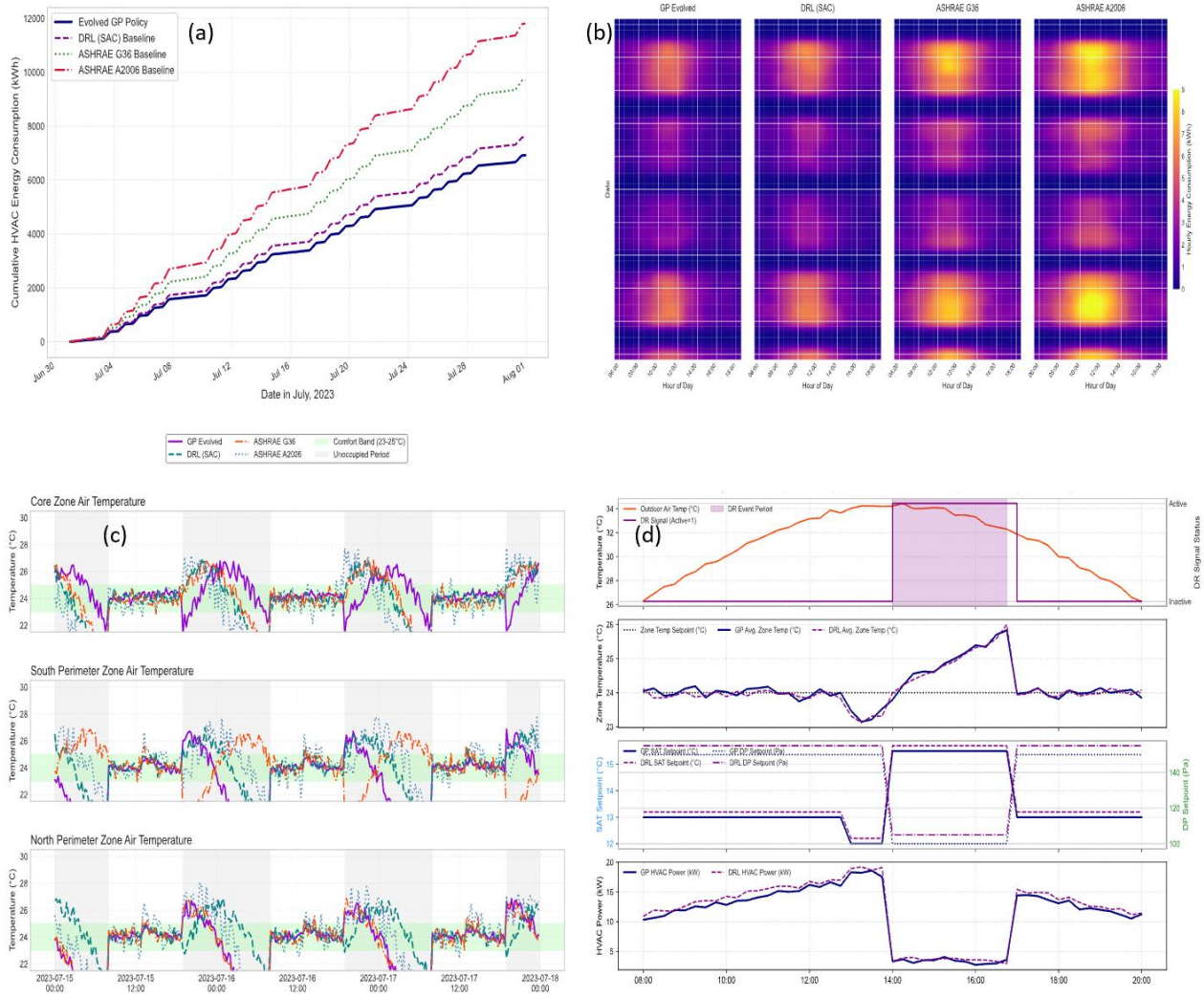

Fig. 10 Comparative Performance of Control Strategies: (a) Cumulative monthly HVAC energy consumption; (b) Hourly energy consumption patterns; (c) Zone air temperature control on representative days; (d) System dynamic response during a peak Demand Response event.

and aggressive utilization of energy-saving measures. In stark contrast, the ASHRAE A2006 panel shows consistently high energy consumption across the entire occupied daytime block, indicative of its less adaptive, more rigid control logic.

Crucially, these substantial energy savings were achieved without compromising occupant comfort. As the data in Table 1 confirms, the Evolved GP Policy achieved a notably low ZAT Violation Degree-Hour value of 75 °C·hr and maintained zone temperatures within the desired [23°C-25°C] comfort band for 98.8% of all occupied hours. This high level of thermal comfort was statistically comparable to the DRL baseline (80 °C·hr, 98.5%) and significantly better than the ASHRAE G36 (150 °C·hr) and A2006 (280 °C·hr) approaches. The dynamic thermal performance is further elucidated in Fig.10c, which presents indoor air temperature profiles for representative summer days. It is visually evident that the GP and DRL policies consistently regulate zone temperatures more tightly within the comfort band, exhibiting smoother profiles with minimal overshoot or undershoot. In contrast, the ASHRAE baselines display more pronounced temperature fluctuations and larger deviations outside the designated comfort band. This visual evidence strongly suggests that the learning-based approaches achieve superior comfort stability due to their ability to learn more nuanced and anticipatory responses to varying load conditions.

Finally, the GP policy's capacity to actively participate in Demand Response (DR) events was exceptional. As summarized in Table 1, the controller achieved an average peak load reduction of 13.7 kW (72.1%), a performance comparable to the DRL baseline while the standard ASHRAE baselines showed negligible active participation. The dynamic response to a representative DR event is visualized in Fig.10d, providing a clear, step-by-step illustration of the learned strategy. Upon activation of the DR signal, both intelligent controllers aggressively increased their Supply Air Temperature (SAT) setpoints from approximately 12.5°C to 15.5°C. As a direct result, total HVAC power consumption decreased dramatically from a peak of ~17-19 kW to a minimal ~3-4 kW for the duration of the event. During this period, the average zone temperatures experienced a controlled, slow drift, peaking around 25.5-25.8°C before being brought back down. This clearly illustrates the GP framework's ability to evolve effective, explicit strategies for demand-side management, successfully completing the trifecta of holistic, multi-objective optimization.

3.3 *Analysis of Evolved GP Operational Strategies, Policy Complexity, and Interpretability*

The superior performance documented in the previous section is not a mystery. A unique and powerful advantage of the Genetic Programming approach is the inherent transparency of its resulting policies. While the DRL agent's logic remains an opaque "black-box," the fundamental structure of the evolved GP policies, represented as expression trees, allows for direct inspection, analysis, and human understanding. This section deconstructs the evolved strategies to reveal the drivers of its success and discuss the critical importance of interpretability.

First, it is important to address the complexity of the evolved solution. The complete multi-output GP policy, comprising distinct expression trees for each of the four AHU control variables, aggregates to a total of 185 nodes. This moderate complexity represents a successful balance: the policy is sophisticated enough to execute nuanced, high-performance control, yet it remains entirely human-inspectable. This indicates that the GP evolved functionally effective policies without an unmanageable degree of structural complexity, thereby preserving interpretability.

Fig.11 offers a multi-faceted visualization of the controller's evolved "intelligence," revealing the specific, learned strategies that led to its superior performance. Fig.11a visualizes the multi-dimensional control strategy for the AHU SAT setpoint. The surface reveals a distinctly non-linear and adaptive strategy. When zones are at or below their setpoint (zero or negative error), the GP maintains a higher, energy-saving SAT (approximately 15-16.5°C). However, as zones become warmer, the GP policy enacts a sharp reduction in the SAT, driving it towards its lower operational limit of approximately 11°C when the error reaches +2.0°C. This aggressive cooling response demonstrates a learned prioritization, a sophisticated, threshold-influenced behavior that would be challenging to hand-craft. Further insights into emergent operational behavior are provided in Fig.11b and 11c clearly delineates distinct operational modes: a dense cluster of points shows operation with a low SAT (11.5°C-12.5°C) and an

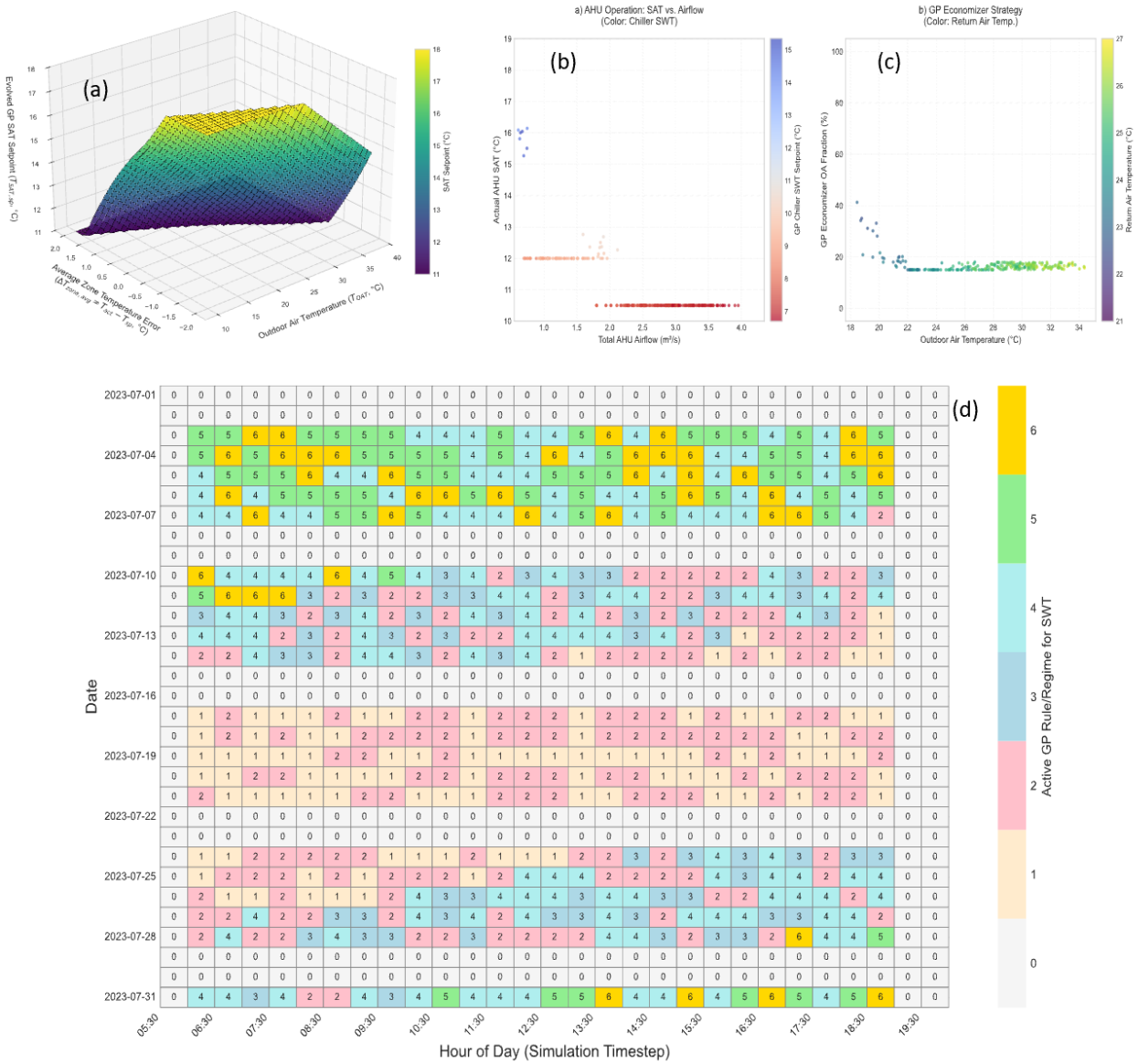

Fig. 11 Analysis of Evolved GP Operational Strategies: (a) Control surface for the AHU SAT setpoint; (b) Emergent coordination of AHU and chiller plant operation; (c) Intelligent economizer free cooling logic; (d) Daily activation patterns of Supply Water Temperature (SWT) management regimes.

aggressive, low Chiller SWT (7-9°C, reddish hues), indicating active cooling under high demand. A sparser cluster at higher SATs (~15-16°C) corresponds to a higher, more efficient Chiller SWT (10-12°C, bluish hues), representing an energy-saving mode during lower loads. This demonstrates a learned coordination between the AHU and chiller plant. A high outdoor air fraction (60-100%) is predominantly utilized when the outdoor air is cool (15-20°C). As the outdoor temperature surpasses ~22°C, the GP decisively minimizes the outdoor air fraction to avoid introducing excessive thermal loads. Finally, to understand the temporal dynamics, Fig.11d presents a heatmap illustrating which specific "GP Rule/Regime for SWT" is active at each timestep. The color coding clearly distinguishes these regimes. For example, on high-load days like 2023-07-04, a clear shift from inactive (regime '0') to active cooling (regimes '4', '5', '6', indicated by yellow/orange/green) is evident during the morning ramp-up and peak afternoon hours. Conversely, during milder periods or weekends, lower-numbered regimes ('0', '1', '2', white/pink), indicating reduced cooling, are predominantly active. This visualization highlights the GP's ability to dynamically switch between operational modes based on the time of day and prevailing conditions, providing a granular view of its adaptive behavior beyond static setpoint schedules.

This ability to deconstruct and verify the control logic is the paramount contribution of this work. It directly addresses the primary barrier to the adoption of advanced AI in critical building systems: a lack of trust and verifiability (Yu *et al.*, 2021). While a "black-box" model may perform well, its opacity creates significant implementation hurdles. The GP controller's "white-box" nature provides the transparency necessary for vetting, trust, and practical implementation by facility managers, offering a viable pathway to deploy truly intelligent and trustworthy building automation.

3.4 Statistical Significance of Performance Differences

To ascertain the statistical robustness of the observed performance advantages, a comprehensive analysis was conducted using 10 independent simulation runs for each controller, with the results summarized in Table 2. This analysis formally validates that the superior performance of the evolved GP policy is not an artifact of a single simulation but a consistent and statistically significant outcome. In the critical domain of energy efficiency, the analysis reveals a clear hierarchy. The GP controller was found to be significantly superior to all other methods, consuming less energy than the ASHRAE A2006 ($p < 0.001$) and G36 ($p = 0.002$) baselines. Most consequentially, the 9.3% energy saving achieved by the GP policy over the state-of-the-art DRL agent was also confirmed to be a statistically significant advantage ($p = 0.045$). This finding is pivotal, as it provides strong empirical evidence that, for this complex control problem, the GP's evolutionary search discovered a more globally optimal policy than the DRL's gradient-based learning. While achieving this superior efficiency, the GP policy's performance in maintaining thermal comfort and executing demand response was statistically indistinguishable from the DRL agent, with p-values of 0.350 and 0.650, respectively. This demonstrates that the GP's energy advantage was not a simple trade-off but a genuine optimization gain, achieved without any statistically significant sacrifice in other key performance areas.

These statistical findings are highly significant because they paint a clear, empirically-backed picture of a holistically superior and more intelligent controller. The GP model did not simply find a good solution; it found a better way to balance the system's competing objectives. The fact that its energy savings are statistically significant while its comfort and DR performance remain on par with the DRL agent suggests that the GP learned to successfully decouple energy efficiency from comfort provision in a way the DRL agent could not. The true importance of our model, however, is realized when coupling this statistically validated performance with its inherent interpretability, as discussed in the previous section. This combination directly addresses the primary barrier to the adoption of advanced AI in critical building systems: the black-box problem, which creates a fundamental lack of trust and verifiability (Cpalka, Łapa & Przybył, 2018). By providing a solution that is not only statistically proven to be more efficient but is also fully transparent and auditable, our GP model represents a viable and highly advantageous pathway toward intelligent HVAC control that is not just effective, but also trustworthy and practically deployable in real-world applications.
HVAC control.

3.5 Limitations and Future work

While this study robustly demonstrates the significant potential of the GP-evolved controllers, several considerations for practical application and future research warrant discussion. The findings, derived from a high-fidelity simulation environment, provide a strong performance benchmark, but on-site validation is the logical next step to confirm performance against real-world dynamics. A key methodological limitation is the lack of a separate validation dataset, which raises the potential for overfitting. Future work must therefore validate the evolved policies against different weather years and seasons to rigorously assess their generalization performance and robustness.

Translating this research into industry practice requires addressing two primary barriers: the integration with proprietary Building Automation Systems (BAS) via standardized APIs, and the upfront computational and expertise requirements for policy evolution. However, the inherent transparency of GP offers a clear advantage over opaque AI, significantly lowering these adoption hurdles. A practical pathway could involve a Control-as-a-Service model, where foundational policies are evolved for building archetypes and then presented to facility managers for verification. This white-box nature enables a phased, trust-building deployment like shadow mode with human oversight, contrasting sharply with the all-or-nothing trust demanded by black-box DRL agents.

Looking ahead, future work should explore hybrid models that combine the strengths of GP and Deep Reinforcement Learning. A promising approach involves using GP to evolve an interpretable, high-level strategic framework—defining the operational modes and primary logic while a DRL agent is tasked with fine-tuning the continuous control parameters within that GP-defined logic in real-time. Such a hybrid system could offer the best of both worlds: the robust, transparent, and verifiable strategic intelligence of GP, coupled with the adaptive, fine-grained optimization of DRL.

4 Conclusion

This research successfully pioneered and rigorously validated a Genetic Programming (GP) framework for the direct evolution of interpretable, multi-objective HVAC control policies, addressing the critical black-box limitations of contemporary AI controllers while integrating sophisticated Demand Response

capabilities. Comprehensive simulations within a validated multi-zone office building model demonstrated that the GP-evolved policies achieved superior energy efficiency, reducing total HVAC consumption by a significant 40.9% against ASHRAE A2006, 28.4% over ASHRAE G36, and a notable 9.3% compared to a state-of-the-art Deep Reinforcement Learning agent, all while maintaining excellent thermal comfort for 98.8% of occupied hours—a level comparable to DRL and markedly better than standard baselines. Furthermore, the GP policies exhibited robust DR effectiveness, delivering a 72.1% peak load reduction through learned strategies like pre-cooling and dynamic setpoint modulation, performing on par with DRL. The paramount contribution of this work is the attainment of this high operational performance through policies that are inherently transparent 185 total nodes for the complete AHU strategy, allowing for direct human inspection, verification, and trust. This contrasts sharply with opaque DRL models and obviates the need for potentially inexact post-hoc explanations. By directly evolving understandable, high-performing solutions, GP is substantiated as a potent and practical methodology for advancing intelligent building automation towards more efficient, grid-responsive, and trustworthy systems. Future investigations should prioritize real-world deployment, enhancing GP scalability for larger systems, and evolving adaptive policies with greater resilience to operational uncertainties and faults.

References

Afroz, Z., Shafiullah, G., Urmee, T., & Higgins, G. (2018). Modeling techniques used in building HVAC control systems: A review. *Renewable and Sustainable Energy Reviews*, 83, 64-84; https://doi.org/10.1016/j.rser.2017.10.044

Alimohammadisagvand, B., Jokisalo, J., & Sirén, K. (2018). Comparison of four rule-based demand response control algorithms in an electrically and heat pump-heated residential building. *Applied Energy*, 209, 167-179; https://doi.org/10.1016/j.apenergy.2017.10.088

Al Sayed, K., Boodi, A., Sadeghian Broujeny, R., & Beddiar, K. (2024). Reinforcement learning for HVAC control in intelligent buildings: A technical and conceptual review. *Journal of Building Engineering*, 95, 110085; https://doi.org/10.1016/j.jobe.2024.110085

Amer, A., Bayhan, S., Abu-Rub, H., Ehsani, M., & Massoud, A. (2024). Enhancing Grid Stability through Grid-Interactive Efficient Buildings with Deep Reinforcement Learning: Innovations and Challenges. In *IECON 2024 - 50th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, pp. 1-6; https://doi.org/10.1109/IECON55916.2024.10905725

Bitar, R., Youssef, N., Chamoin, J., Hage Chehade, F., & Defer, D. (2024). Simultaneous Energy Optimization of Heating Systems by Multi-Zone Predictive Control—Application to a Residential Building. *Buildings*, 14(10), 3241; https://doi.org/10.3390/buildings14103241

Bouabdallaoui, Y., Lafhaj, Z., Yim, P., Ducoulombier, L., & Bennadji, B. (2021). Predictive Maintenance in Building Facilities: A Machine Learning-Based Approach. *Sensors*, 21(4), 1044; https://doi.org/10.3390/s21041044

Chaturvedi, S., Rajasekar, E., & Natarajan, S. (2020). Multi-objective Building Design Optimization under Operational Uncertainties Using the NSGA II Algorithm. *Buildings*, 10(5), 88; https://doi.org/10.3390/buildings10050088

Cheraghi, R. & Jahangir, M. H. (2023). Multi-objective optimization of a hybrid renewable energy system supplying a residential building using NSGA-II and MOPSO algorithms. *Energy Conversion and Management*, 294, 117515; https://doi.org/10.1016/j.enconman.2023.117515

Cho, J., Lee, H., & Heo, Y. (2023). Dynamic rule-based change-over ventilation strategy with weather-responsive air-conditioning setpoints. *Building and Environment*, 246, 110966; https://doi.org/10.1016/j.buildenv.2023.110966

Choi, Y., Lu, X., O'Neill, Z., Feng, F., & Yang, T. (2023). Optimization-informed rule extraction for HVAC system: A case study of dedicated outdoor air system control in a mixed-humid climate zone. *Energy and Buildings*, 295, 113295; https://doi.org/10.1016/j.enbuild.2023.113295

Çinar, E. & Abut, T. (2025). Fuzzy LQR-based control to ensure comfort in HVAC system with two different zones. *Case Studies in Thermal Engineering*, 73, 106544; https://doi.org/10.1016/j.csite.2025.106544

Cpalka, K., Łapa, K., & Przybył, A. (2018). Genetic Programming Algorithm for Designing of Control Systems. *Information Technology And Control*, 47(4); https://doi.org/10.5755/j01.itc.47.4.20795

Deru, M., Field, K., Studer, D., Benne, K., Griffith, B., Torcellini, P., ... & Crawley, D. (2011). *U.S. Department of Energy commercial reference building models of the national building stock* (No. NREL/TP-5500-46861). National Renewable Energy Lab. (NREL), Golden, CO (United States)

Ding, X., Cerpa, A., & Du, W. (2025). Multi-Zone HVAC Control With Model-Based Deep Reinforcement Learning. *IEEE Transactions on Automation Science and Engineering*, 22, 4408-4426; https://doi.org/10.1109/TASE.2024.3410951

Es-sakali, N., Zoubir, Z., Idrissi Kaitouni, S., Mghazli, M. O., Cherkaoui, M., & Pfafferott, J. (2024). Advanced predictive maintenance and fault diagnosis strategy for enhanced HVAC efficiency in buildings. *Applied Thermal Engineering*, 254, 123910; https://doi.org/10.1016/j.applthermaleng.2024.123910

Gao, Y., Li, S., Fu, X., Dong, W., Lu, B., & Li, Z. (2020). Energy management and demand response with intelligent learning for multi-thermal-zone buildings. *Energy*, 210, 118411; https://doi.org/10.1016/j.energy.2020.118411

Ghaderian, M. & Veysi, F. (2021). Multi-objective optimization of energy efficiency and thermal comfort in an existing office building using NSGA-II with fitness approximation: A case study. *Journal of Building Engineering*, 41, 102440; https://doi.org/10.1016/j.jobe.2021.102440

Hou, F., Cheng, J. C. P., Kwok, H. H. L., & Ma, J. (2024). Multi-source transfer learning method for enhancing the deployment of deep reinforcement learning in multi-zone building HVAC control. *Energy and Buildings*, 322, 114696; https://doi.org/10.1016/j.enbuild.2024.114696

Kargar, S. M. & Bahamin, M. (2025). Advanced model predictive control strategy for thermal management in multi-zone buildings with energy storage and dynamic pricing. *Energy Exploration & Exploitation*, 43(3), 1308-1326; https://doi.org/10.1177/01445987241298534

Kaushik, E., Prakash, V., Mahela, O. P., Khan, B., El-Shahat, A., & Abdelaziz, A. Y. (2022). Comprehensive Overview of Power System Flexibility during the Scenario of High Penetration of Renewable Energy in Utility Grid. *Energies*, 15(2), 516; https://doi.org/10.3390/en15020516

Kim, Y.-J. (2020). A Supervised-Learning-Based Strategy for Optimal Demand Response of an HVAC System in a Multi-Zone Office Building. *IEEE Transactions on Smart Grid*, 11(5), 4212-4226; https://doi.org/10.1109/TSG.2020.2986539

Kumar, V., Niazi, M. A., Sadiq, M. U., Rizwan, M., Sajid, Q., & Ahmad, S. (2025). Enhancing electric vehicle battery performance through Grey Wolf Optimization and Deep Reinforcement Learning integration. *IET Conference Proceedings*, 2025(3), 72-79; https://doi.org/10.1049/icp.2025.1097

Kumar, V., Niazi, M. A., Sajid, Q., Rizwan, M., Shah, S. A. A., & Khatoon, S. (2025). Optimized Energy Management for EV Charging Stations Using Soft Actor-Critic Reinforcement Learning. In *2025 International Conference on Emerging Technologies in Electronics, Computing, and Communication (ICETECC)*. IEEE, pp. 1-6; https://doi.org/10.1109/ICETECC65365.2025.11070275

Lu, R., Bai, R., Luo, Z., Jiang, J., Sun, M., & Zhang, H.-T. (2022). Deep Reinforcement Learning-Based Demand Response for Smart Facilities Energy Management. *IEEE Transactions on Industrial Electronics*, 69(8), 8554-8565; https://doi.org/10.1109/TIE.2021.3104596

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; https://doi.org/10.1016/j.enbuild.2023.112854

Mariano-Hernández, D., Hernández-Callejo, L., Zorita-Lamadrid, A., Duque-Pérez, O., & Santos García, F. (2021). A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis. *Journal of Building Engineering*, 33, 101692; https://doi.org/10.1016/j.jobe.2020.101692

Niazi, M. A., Kumar, V., Sadiq, M. U., Sajid, Q., Rizwan, M., & Ahmad, S. (2025). Strategic management of forecast uncertainties to mitigate power imbalances in virtual power plants. *IET Conference Proceedings*, 2025(3), 64-71; https://doi.org/10.1049/icp.2025.1096

Niazi, M. A., Bukhari, S. A. A. S., Kumar, V., Zuhaib, K. M., & Aslam, U. (2025). Optimal Scheduling of Electric Vehicle Aggregators in Residential Areas: A Cost Minimization Approach. *Sukkur IBA Journal of Emerging Technologies*, 8(1), 62-69; https://doi.org/10.30537/sjet.v8i1.1602

Pang, L. M., Ishibuchi, H., Deb, K., & Shang, K. (2025). MaNSGA-II: Many-Objective NSGA-II. *IEEE Transactions on Emerging Topics in Computational Intelligence*, pp. 1-16; https://doi.org/10.1109/TETCI.2025.3576105

Pérez-Lombard, L., Ortiz, J., & Pout, C. (2008). A review on buildings energy consumption information. *Energy and Buildings*, 40(3), 394-398; https://doi.org/10.1016/j.enbuild.2007.03.007

Pinthurat, W., Surinkaew, T., & Hredzak, B. (2024). An overview of reinforcement learning-based approaches for smart home energy management systems with energy storages. *Renewable and Sustainable Energy Reviews*, 202, 114648; https://doi.org/10.1016/j.rser.2024.114648

Pinto, G., Wang, Z., Roy, A., Hong, T., & Capozzoli, A. (2022). Transfer learning for smart buildings: A critical review of algorithms, applications, and future perspectives. *Advances in Applied Energy*, 5, 100084; https://doi.org/10.1016/j.adapen.2022.100084

Sanzana, M. R., Maul, T., Wong, J. Y., Abdulrazic, M. O. M., & Yip, C.-C. (2022). Application of deep learning in facility management and maintenance for heating, ventilation, and air conditioning. *Automation in Construction*, 141, 104445; https://doi.org/10.1016/j.autcon.2022.104445

Sipper, M. & Moore, J. H. (2020). Genetic programming theory and practice: a fifteen-year trajectory. *Genetic Programming and Evolvable Machines*, 21(1-2), 169-179; https://doi.org/10.1007/s10710-019-09353-5

Sun, H., Hu, Y., Luo, J., Guo, Q., & Zhao, J. (2025). Enhancing HVAC Control Systems Using a Steady Soft Actor-Critic Deep Reinforcement Learning Approach. *Buildings*, 15(4), 644; https://doi.org/10.3390/buildings15040644

Tomás, L., Lämmle, M., & Pfafferott, J. (2025). Demonstration and Evaluation of Model Predictive Control (MPC) for a Real-World Heat Pump System in a Commercial Low-Energy Building for Cost Reduction and Enhanced Grid Support. *Energies*, 18(6), 1434; https://doi.org/10.3390/en18061434

Xie, J., Ajagekar, A., & You, F. (2023). Attention Based Multi-Agent Reinforcement Learning for Demand Response in Grid-Responsive Buildings. In *2023 IEEE Conference on Control Technology and Applications (CCTA)*. IEEE, pp. 118-123; https://doi.org/10.1109/CCTA54093.2023.10253019

Yan, R., Ma, Z., Zhao, Y., & Kokogiannakis, G. (2016). A decision tree based data-driven diagnostic strategy for air handling units. *Energy and Buildings*, 133, 37-45; https://doi.org/10.1016/j.enbuild.2016.09.039

Yao, G., Chen, Y., Han, C., & Duan, Z. (2024). Research on the Decision-Making Method for the Passive Design Parameters of Zero Energy Houses in Severe Cold Regions Based on Decision Trees. *Energies*, 17(2), 506; https://doi.org/10.3390/en17020506

Yoon, Y., Amasyali, K., Li, Y., Im, P., Bae, Y., Liu, Y., & Zandi, H. (2024). Energy performance evaluation of the ASHRAE Guideline 36 control and reinforcement learning-based control using field measurements. *Energy and Buildings*, *325*, 115005. https://doi.org/10.1016/j.enbuild.2024.115005

Yu, L., Sun, Y., Xu, Z., Shen, C., Yue, D., & Jiang, T. (2021). Multi-Agent Deep Reinforcement Learning for HVAC Control in Commercial Buildings. *IEEE Transactions on Smart Grid*, *12*(1), 407–419. https://doi.org/10.1109/TSG.2020.3011739

Appendix A

Building and HVAC System Model Parameters

Table A1
HVAC System Component Model Parameters (Based on DOE Medium Office Reference Building, Climate Zone 5A)

Component Category	Component	Parameter	Value	Units
Chiller Plant	Water-Cooled Chiller	Nominal Capacity	211	kW
		Rated COP (Coefficient of Performance)	6.2	-
		Chilled Water Supply Temp. Range	6.7 - 12.0	°C
	Cooling Tower	Design Approach Temperature	4.0	°C
		Design Range Temperature	5.5	°C
		Fan Power at Design Airflow	7.5	kW
	Chilled Water Pump	Rated Flow Rate	9.1	L/s
	(Primary, Variable Speed)	Rated Head	180	kPa
		Motor Efficiency	0.90	-
	Condenser Water Pump	Rated Flow Rate	11.4	L/s
	(Constant Speed)	Rated Head	150	kPa
		Motor Efficiency	0.90	-
Air Handling Unit	AHU Supply Fan	Design Airflow Rate	7.55	m^3/s
	(Variable Speed)	Design Static Pressure	1120	Pa
		Total Fan Efficiency	0.65	-
		Motor Efficiency	0.92	-
	Chilled Water Cooling Coil	Design Capacity (Sensible)	155	kW
		Design Inlet Air Temperature (Dry-Bulb)	26.0	°C
		Design Inlet Water Temperature	6.7	°C
Zonal Equipment	VAV Terminal Units (x5)	Maximum Airflow Rate (Perimeter)	1.2	m^3/s
		Maximum Airflow Rate (Core)	1.55	m^3/s
		Minimum Airflow Fraction	0.30	-

Table B1
Controller Input (State/Terminals) and Output (Action) Variables

Type	Variable Name	Symbol	Description	Unit
Input	Outdoor Air Temperature	T_{OAT}	Current measured dry-bulb temperature of outside air.	°C
Input	Outdoor Air Temp. Forecast +1h	$T_{OAT,pred+1h}$	Predicted outdoor air temperature for the next hour.	°C
Input	Outdoor Air Temp. Forecast +2h	$T_{OAT,pred+2h}$	Predicted outdoor air temperature for two hours ahead.	°C
Input	Outdoor Air Temp. Forecast +4h	$T_{OAT,pred+4h}$	Predicted outdoor air temperature for four hours ahead.	°C
Input	Average Zone Air Temperature	$T_{zone,avg}$	Mean air temperature across all 5 conditioned zones.	°C
Input	Maximum Zone Air Temperature	$T_{zone,\max}$	Air temperature of the warmest conditioned zone.	°C
Input	Max Zone Temp. Deviation	$\Delta T_{zone,max\_dev}$	Maximum positive deviation from the cooling setpoint in any zone.	K or °C
Input	Current Supply Air Temperature	$T_{SAT,act}$	Current measured temperature of the air leaving the AHU.	°C
Input	Return Air Temperature	T_{Ret}	Temperature of air returning to the AHU from the zones.	°C
Input	Max VAV Damper Position	$V_{AV,\mathrm{damp,crit}}$	Position of the most-open VAV box damper.	%
Input	Time of Day	t_{day}	Current hour of the day (e.g., 0-23).	hour
Input	Day of the Week	d_{week}	Current day of the week (e.g., 1=Mon, 7=Sun).	-
Input	Demand Response Signal	$S_{DR}(t)$	Binary signal indicating an active DR event (0=No, 1=Yes).	-
Output	AHU Supply Air Temp. Setpoint	$T_{\mathrm{SAT,sp}}$	Target temperature for air leaving the AHU cooling coil.	°C
Output	AHU Duct Static Pressure Setpoint	$P_{\mathrm{DP,sp}}$	Target static pressure to be maintained in the main supply duct.	Pa
Output	Chiller Supply Water Temp. Setpoint	$T_{\mathrm{SWT,sp}}$	Target temperature for water leaving the chiller.	°C
Output	Economizer Control Parameter	u_{econo}	Control parameter for economizer (e.g., outdoor air fraction).	-

Table C.1
DRL Agent (Soft Actor-Critic) Hyperparameters

Category	Parameter	Value	Description
Algorithm Hyperparameters	Algorithm	Soft Actor-Critic (SAC)	State-of-the-art off-policy algorithm for continuous control, balancing exploration and exploitation via entropy maximization.
	Learning Rate (Actor & Critic)	3e-4	The step size for updating the neural network weights during training.
	Discount Factor (γ)	0.99	Determines the importance of future rewards. A value of 0.99 prioritizes long-term performance.
	Target Smoothing Coefficient (τ)	0.005	Controls the update speed of the target networks, promoting stable learning.
	Entropy Coefficient (α)	Auto-tuned	Automatically adjusted during training to balance reward maximization (exploitation) and entropy (exploration).
Neural Network Architecture	Actor/Critic Hidden Layers	2	The number of layers between the input and output layers for both networks.
	Neurons per Hidden Layer	256	The number of nodes in each hidden layer, defining the network's capacity.
	Activation Function (Hidden)	ReLU	Rectified Linear Unit, a standard non-linear activation for hidden layers.
	Activation Function (Output)	Tanh	Hyperbolic Tangent, used to bound the continuous actions to the [-1, 1] range.
Training Parameters	Total Timesteps	2,000,000	The total number of environment interactions used for training the agent.
	Replay Buffer Size	100,000	The number of past experiences (state, action, reward, next_state) stored for training.
	Batch Size	256	The number of experiences sampled from the replay buffer for each training update.
	Optimizer	Adam	An adaptive learning rate optimization algorithm used for training the networks.