International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(Research Article)

Check for updates

Deep reinforcement learning for multi-criteria optimization in BIM-supported sustainable building design

Ruchit Parekh [1,*], Charles Smith [2] and Nathan Brown [3]

[1] *Department of Engineering, Hofstra University, New York, USA.*

[2] *Department of Computer Science, Hofstra University, New York, USA.*

[3] *Department of Sustainability and Renewable Energy, University of Colorado Boulder, USA.*

International Journal of Science and Research Archive, 2024, 13(01), 1030–1048

Publication history: Received on 11 August 2024; revised on 17 September 2024; accepted on 20 September 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.13.1.1775

Abstract

This paper introduces a multi-criteria optimization (MCO) framework tailored for sustainable building design, leveraging a deep reinforcement learning (DRL) methodology to adjust design parameters. The framework integrates three key components: developing an assessment index system, training a deep neural network (DNN) model, and generating an optimal solution set using the deep deterministic policy gradient (DDPG) model. The DRL approach was tested on a three-story educational building in Shanghai, where it demonstrated superiority over traditional genetic algorithms by optimizing building energy consumption, carbon emissions, and indoor thermal comfort simultaneously, improving overall building performance by 13.19%. The DDPG agent autonomously learns and enhances decision-making policies through iterative interactions with a Building Information Modeling (BIM) environment combined with DNN, offering advanced data-driven decision support for sustainable building design.

Keywords: Multi-criteria optimization; Sustainable building design; Deep reinforcement learning approach; DDPG Model; Building performance

1. Introduction

With China experiencing rapid economic development and urbanization, the construction of new buildings has surged significantly, resulting in a sharp increase in energy consumption and associated carbon emissions within the building sector. The building industry now accounts for over 46.5% of China's total energy use and approximately 6.0% of global energy demand. This sector is recognized as a critical area for energy conservation efforts, playing a pivotal role in China's objectives to peak carbon emissions by 2030 and achieve carbon neutrality by 2060. However, only 4% of the existing buildings have adopted energy-saving practices, which is inadequate to curb environmental degradation and reduce the carbon footprint. Given its dominant contribution to energy consumption, there is a pressing need to transform the "green building" concept, which focuses on near-zero or net-zero energy usage, into a widespread reality in Chinese cities. Such efforts not only promise significant energy savings but also offer substantial market potential.

The development of green buildings centers on creating environmentally friendly structures that prioritize energy efficiency, environmental impact, and indoor quality. These factors have become critical for designers and stakeholders working towards sustainable development goals. Sustainable building design, when implemented in the pre-construction phase, can prevent poor energy performance, potentially yielding over 30% energy savings by selecting the right design elements such as shape, materials, and type.

Currently, several international green building standards, including LEED (Leadership in Energy and Environmental Design), BREEAM (Building Research Establishment Environmental Assessment Method), and GRIHA (Green Rating for

* Corresponding author: Ruchit Parekh

Integrated Habitat Assessment), are used to assess and guide green building projects. LEED, in particular, is a widely accepted rating system in more than 150 countries and plays a central role in research aimed at optimizing sustainable building designs. Integrating LEED with Building Information Modeling (BIM) facilitates the certification process through seamless data exchange and accurate energy assessments.

In addition to international standards, China has its own green building regulations. The Ministry of Housing and Urban-Rural Development (MHURD) has introduced the ESGB (Evaluation Standard for Green Buildings), which classifies buildings into three-star ratings, scoring them based on six dimensions, including comfort, resource efficiency, and environmental impact. This study selects the comfort, resource, and environmental dimensions from the ESGB to improve building performance, aligning with the design requirements of engineering projects.

While these standards provide valuable guidance for green building design, their implementation often requires considerable human effort and expertise, making the process time-consuming, uncertain, and costly. Additionally, designers must rely on their experience to determine appropriate design values, which can lead to inaccuracies and inconsistent outcomes. Conventional design methods, especially in the early phases, depend heavily on existing standards and expert knowledge, which can be restricted by limited resources, preventing the consistent delivery of energy-efficient buildings.

To address the limitations previously mentioned, Building Information Modeling (BIM), as a digital tool, facilitates the integration and management of data throughout the project lifecycle, offering an opportunity to fully utilize design data for sustainable design and performance analysis. The BIM model contains comprehensive and detailed information related to design parameters, energy-using systems, and other elements, enabling optimization of design strategies at early stages. BIM tools allow for the evaluation of a building's energy efficiency by considering various external factors and internal components, such as orientation, walls, roofing, and HVAC systems. Acting as a central database, BIM can aggregate extensive data about a building's physical and functional attributes, taking on characteristics of "big data." This accumulation of data presents challenges for traditional data usage methods and experience-based decision-making in construction, as managing such vast amounts of information requires more advanced approaches. Furthermore, optimizing design schemes using BIM’s large datasets results in greater computational demands. Designers are increasingly seeking intelligent systems that enhance the efficiency and precision of building design optimization.

The challenge lies in fully leveraging this design data to make sustainable and objective decisions in BIM-based designs, which would elevate BIM to a higher level of intelligence and reliability. Significant research has focused on creating multi-factor building performance prediction models using various machine learning algorithms suited for different types of buildings, predictive objectives, and timeframes. For example, Liu et al. developed a random forest model for predicting building energy usage based on the building envelope's design, identifying key variables to optimize energy-efficient designs. Similarly, Toosi et al. created a life cycle sustainability assessment index to evaluate design options, utilizing machine learning to forecast energy performance for different design variable combinations. These studies highlight the benefits of fast predictions and simplified parameter structures, particularly in the early stages of design.

However, focusing solely on prediction is insufficient. A major drawback of prediction models is their lack of data-driven decision-making capabilities. These models cannot continuously adapt and improve to provide optimal, flexible strategies for complex and evolving environments. Moreover, they are unable to address multiple competing objectives simultaneously, making it difficult to find the best trade-offs for sustainable building designs. Addressing these optimization strategies is critical for identifying scenarios that enhance energy efficiency and reduce greenhouse gas (GHG) emissions. Current design methods tend to focus on individual goals like energy conservation, thermal efficiency, and carbon reduction, often ignoring the interactions between them. There is a pressing need to adopt multi-objective optimization (MOO) techniques that can simultaneously improve multiple building performance objectives.

Thus, to maximize the potential of the deep neural network (DNN)-based meta-model with predictive abilities, there is an urgent requirement for appropriate MOO methods to be implemented. These MOO-based decision support systems can autonomously identify optimal solutions during the early design phases, significantly reducing reliance on expert judgment and prior knowledge.

Some studies have explored the use of classical multi-objective optimization (MOO) algorithms for optimizing building design, such as the Non-dominated Sorting Genetic Algorithm (NSGA-II) and the Adaptive Geometry Estimation-based Multi-Objective Evolutionary Algorithm (AGE-MOEA). These algorithms follow NSGA principles to identify population-based Pareto front solution sets. However, they face two significant limitations. First, the process of updating populations or performing iterative searches is time-consuming, especially when dealing with large data sets, leading to low operational efficiency. Second, even slight changes in geometric properties, environmental factors, or additional optimization goals can cause these algorithms to struggle with adaptability, resulting in poor generalization performance. This exposes a research gap in the scalability and adaptability of current optimization algorithms when applied to BIM data for sustainable building design.

Deep reinforcement learning (DRL), a rapidly emerging field within machine learning, offers a promising solution by enabling the development of more effective strategies for sustainable building design. Unlike traditional MOO algorithms, DRL combines reinforcement learning with deep neural networks, creating intelligent agents that can autonomously learn from iterative interactions with the environment and adapt strategies through a trial-and-error process. This allows DRL to determine optimal policies by simply defining the desired objective and cost function. While research into DRL-based approaches for green building design is still in its early stages, this topic holds great potential and merits further investigation.

Building on the strong learning capabilities of deep neural networks (DNN) and DRL, this paper introduces a DRL-based MOO approach for BIM-supported green building design. As a powerful decision-making tool, DRL can simultaneously account for energy efficiency and indoor environmental factors to optimize the design parameters of the building envelope, openings, and HVAC systems, ultimately enhancing sustainability.

This research seeks to answer three key questions: (1) How can valuable information be extracted from the BIM model to prepare a dataset with a well-structured assessment index system? (2) How can the green building design decision-making process be mathematically framed using a DRL-based model, and how can appropriate algorithms be applied to yield accurate predictions and optimal strategies? (3) How can comparative experiments be designed to evaluate the robustness, stability, and generalization of the DRL approach across various conditions?

The case study presented demonstrates the effectiveness of DRL in generating optimal design solutions for real-world buildings, showcasing its potential in enhancing sustainable building design.

The key innovations of this research can be summarized as follows:

- Theoretical Value: This study introduces a DRL approach that combines the robust non-linear fitting capabilities of deep neural networks (DNN) with the decision-making strengths of reinforcement learning. This integration allows for a well-rounded and generalized model, enabling precise predictions and optimal decision-making in the context of BIM-based building design.
- Practical Value: The DRL-based multi-objective optimization (MOO) task is achieved through autonomous exploration, action selection, interaction with the environment, state observation, and reward collection. This method eliminates the need for cumbersome population updates and iterative searches, providing flexible solutions that balance multiple objectives. These solutions are both realistic and adaptable to changing physical conditions.

The structure of the paper is as follows: Section 2 reviews relevant studies on reinforcement learning. Section 3 outlines the methodology of the proposed DRL approach, covering BIM-based simulations, DNN-based predictions, and DRL-based optimization. Section 4 applies this DRL framework to a real-world building project to validate its effectiveness in generating optimal decisions for green building design. Section 5 explores the stability and generalization of the DRL approach across different spatial and temporal contexts. Lastly, Section 6 summarizes the findings and proposes future research directions.

2. Literature Review

2.1. BIM-Based Green Building Design

Green building design focuses on creating environmentally responsible and resource-efficient buildings throughout their entire lifecycle, a concept that has gained significant traction in recent years. This design approach prioritizes the use of recycled materials, reduced water and energy consumption, and resource-efficient techniques to improve energy efficiency, reduce emissions, and enhance adaptability. These efforts not only improve the performance of buildings but also yield broader environmental, economic, and social benefits. Scholars have extensively researched the environmental, economic, and social impacts of green building design. Specifically, there is a close relationship between the economic and environmental effects, as these are influenced by building elements and materials. This research provides valuable insights into the current trends and future potential of green building practices.

For instance, Borkovskaya et al. developed a model to assess the environmental and economic lifecycle of buildings within the "green building" framework, aiming to promote sustainable development and enhance the competitiveness of environmentally friendly buildings. Moreover, the geographic distribution and influencing factors of green residential buildings (GRBs) in China have been studied, revealing variations in the distribution of green buildings across different regions and their correlation with local economic and social factors. This research underscores the importance of understanding how geographic and socio-economic factors affect the promotion of sustainable building practices.

Various green building assessment systems have also been developed to evaluate a building's environmental performance, energy efficiency, and sustainability, including popular tools such as LEED, GRIHA, and ESGB. These systems aim to align building practices with global sustainability trends, focusing on aspects such as energy consumption, material usage, and indoor environmental quality. For example, Pai et al. discussed the Whole Building Life Cycle Assessment (WBLCA) in the context of LEED, emphasizing its role in reducing the embodied energy and environmental impact of materials in new construction projects. Similarly, Singh et al. highlighted the GRIHA rating system's focus on evaluating the environmental efficiency of buildings over their entire lifecycle.

In China, the ESGB (Evaluation Standard for Green Buildings) has become one of the main assessment systems applied to engineering projects. Researchers have used the ESGB standard to explore energy-saving technologies and strategies in the construction industry, covering aspects such as thermal insulation materials, daylighting design, and building facades. These systems effectively link design efforts with sustainability principles. However, a significant challenge remains: the application of these standards requires substantial expertise and skills to accurately estimate the relevant design factors.

Olawumi et al. identified a research gap in the use of digital systems for assessing the sustainability performance of buildings. This gap led to the development of the green-BIM assessment (GBA) framework, which combines green building rating systems with the data-rich BIM model. The GBA framework offers a cost-effective way to assess sustainability and serves as a valuable tool for comparing different building designs and projects. Since BIM integrates multidisciplinary data for various analyses, it is an ideal platform to manage assessment systems and make informed design decisions early in the project lifecycle.

For instance, Nguyen et al. integrated LEED into the Revit API, enabling automated green building ratings and design processes. Similarly, Rehman et al. proposed a comprehensive BIM-based approach using the LEED system to optimize green building parameters for energy and water conservation, thereby achieving the sustainability objectives of a residential project. These examples highlight the increasing interest in utilizing BIM and green building evaluation systems to simplify green building design, streamline sustainability assessments, and reduce both time and costs. This trend points to the need for further exploration to enhance building designs with greater focus on sustainability, automation, and efficiency.

2.2. Reinforcement Learning Approach

Reinforcement learning (RL) is a core branch of machine learning, alongside supervised and unsupervised learning. It enables systems to interact with their environment, learn optimal strategies, and maximize an objective reward function through trial and error. In a typical RL framework, an agent takes actions, observes the resulting state of the environment, and adjusts its behavior based on the rewards it receives. This closed-loop system is particularly effective at mimicking human decision-making, where actions are chosen to maximize long-term benefits. RL has been widely applied in fields such as robotics and autonomous systems due to its capacity for improving decision-making over time.

However, traditional RL methods face challenges, particularly the "curse of dimensionality," where managing high-dimensional data becomes difficult. As the number of states and actions grows, the Q-table (used for decision-making) increases exponentially, making it unmanageable for large state spaces.

To overcome this limitation, Mnih et al. made a significant advancement in 2015 by integrating deep learning with RL, resulting in the creation of deep reinforcement learning (DRL). This approach revolutionized the field, enabling intelligent agents to autonomously learn complex tasks that were previously intractable. Classical RL systems require manually defined state representations, while DRL automates this process through deep learning. By using neural networks to approximate the Q-function, DRL can process state inputs and generate optimal actions, allowing for "end-to-end" task learning.

DRL effectively combines the perceptual learning abilities of deep learning with the decision-making capabilities of RL, enabling it to handle complex environments. It has demonstrated substantial advantages in domains such as autonomous driving, industrial automation, and robotics. DRL trains agents to learn continuously from their environment and adapt their decisions, achieving autonomous control and optimization.

Automating decision-making processes through DRL is especially beneficial in engineering fields, as it enhances operational efficiency, minimizes errors, and reduces risks when compared to manual processes. Furthermore, automation enables the system to explore vast datasets and learn adaptively from feedback, producing more insightful and flexible strategies that can guide projects with increased accuracy.

The deep Q-network (DQN) model, a widely used method in deep reinforcement learning (DRL), falls under the value-based approach. It combines the strengths of neural networks and Q-learning, allowing agents to execute high-level control in sequential decision-making tasks. DQN has demonstrated its ability to outperform human experts in devising optimal policies, which can be generalized across different scenarios. However, one significant limitation of the DQN model is that it works primarily in discrete action spaces, making it less effective for problems requiring continuous action domains. When applied to continuous action domains, DQN necessitates the use of discrete state and action variables, which can exponentially increase the number of possible actions as the action dimension grows.

To mitigate the computational challenges posed by this, policy gradient methods are often employed. These methods update policy parameters by continuously calculating the gradient of the total reward concerning the policy, enabling convergence toward an optimal solution. Policy gradient-based algorithms are particularly effective in optimizing the total reward of a policy, directly searching for the best policy within the policy space in an end-to-end fashion.

Lillicrap et al. introduced the deep deterministic policy gradient (DDPG) algorithm, which differs from DQN by being a policy-based method. DDPG combines the benefits of the deterministic policy gradient, DQN, and the actor-critic framework, using a deep learning-trained neural network to parameterize both the Q-function and policy.

Experimental results have confirmed the effectiveness of the DDPG algorithm, demonstrating its superior performance in continuous action spaces and its ability to find more optimal solutions compared to DQN. For instance, Wu et al. developed a DDPG-based agent that generated continuous control strategies for energy management in hybrid electric vehicles, significantly improving fuel efficiency. Similarly, Li et al. proposed a DDPG-based approach to manage energy in power-split hybrid electric buses, achieving substantial reductions in average fuel consumption. Gao and Yu, along with their respective teams, applied DDPG to optimize energy control in smart homes, reducing energy usage in HVAC systems while enhancing thermal comfort for occupants.

The key advantages of the deterministic policy gradient are its capacity to learn stochastic policies, its suitability for high-dimensional action spaces, and its ability to handle continuous action domains with effective convergence. As such, DDPG is a highly suitable choice for designing energy management systems across various real-world urban applications, including hybrid electric vehicles and HVAC systems. This makes DDPG particularly relevant in the construction industry, where it can be applied to building design to enhance energy performance, contributing to sustainable design efforts.

Although DRL-based building design is gaining traction, it remains in its early stages. Initial efforts have focused on applying DRL to optimize aspects such as building interior design, building plans, group layouts, and structural elements. DRL enables adaptive decision-making in these areas by interacting with different settings and responding to various energy demands and operational requirements. The model-free nature of DRL, coupled with its ability to learn optimal policies without the need for convexity assumptions, makes it a promising tool for transforming BIM-based green building design, particularly in handling dynamic and unpredictable energy needs.

Future research should aim to create a DRL framework that can simultaneously optimize energy efficiency, environmental sustainability, occupant comfort, and other critical factors. This would require carefully formulated design strategies for the building envelope, openings, and HVAC systems to enhance operational flexibility, energy performance, and system reliability. However, there are several key challenges associated with using DRL for real-world building control. These include the time-consuming and data-intensive training process, as well as the limited robustness and generalization capabilities of current algorithms.

The deep deterministic policy gradient (DDPG)-trained agent, known for its strong generalization abilities and adaptability to unfamiliar environments, offers a potential solution. The optimal-policy search process within DDPG must be conducted to overcome these challenges. Therefore, our research aims to develop a stable and efficient DRL approach that incorporates the DDPG model. By leveraging the interaction between the agent, actions, and the environment, this approach will deliver accurate predictions and reliable optimizations, ultimately guiding decision-making towards effective building design strategies that align with sustainability goals and occupant needs.

3. Methodology

This paper introduces a DRL-enabled approach aimed at automating and enhancing the design and analysis process for green building development within the BIM framework. Serving as a decision-making tool, this approach optimizes green building designs by improving energy efficiency, environmental sustainability, and occupant comfort.

As outlined in the flowchart in Figure 1, the methodology is divided into three key stages:

- Stage 1: The first step is to identify building performance objectives and relevant design parameters using BIM technology and building performance simulation software. This process results in a well-defined assessment index system that lays the foundation for optimization.
- Stage 2: A deep neural network (DNN) model is constructed with an appropriate architecture to capture and learn the complex non-linear relationships between design parameters and multiple building performance objectives.
- Stage 3: A deep reinforcement learning (DRL) algorithm, specifically the deep deterministic policy gradient (DDPG) method, is applied directly to the trained DNN model. This enables effective interactions between the system and its environment, producing optimal design strategies that significantly improve overall building performance.

In summary, the proposed DRL approach automatically explores and optimizes green building designs, minimizing environmental impact while enhancing occupant comfort. This process not only brings environmental benefits but also aligns with broader sustainability goals, offering tangible societal advantages.

3.1. Integration of BIM Model and Simulation Tool

To construct an effective prediction and optimization model, the first critical task is the preparation of an assessment index system. As depicted in Figure 2, this system is based on the Evaluation Standard for Green Buildings (ESGB) and consists of three primary objectives and twelve influential variables. The process begins by creating a detailed 3D BIM model of a teaching building using Autodesk Revit 2021. This model provides an enhanced digital representation, containing both geometric and physical information that is essential for the analysis.

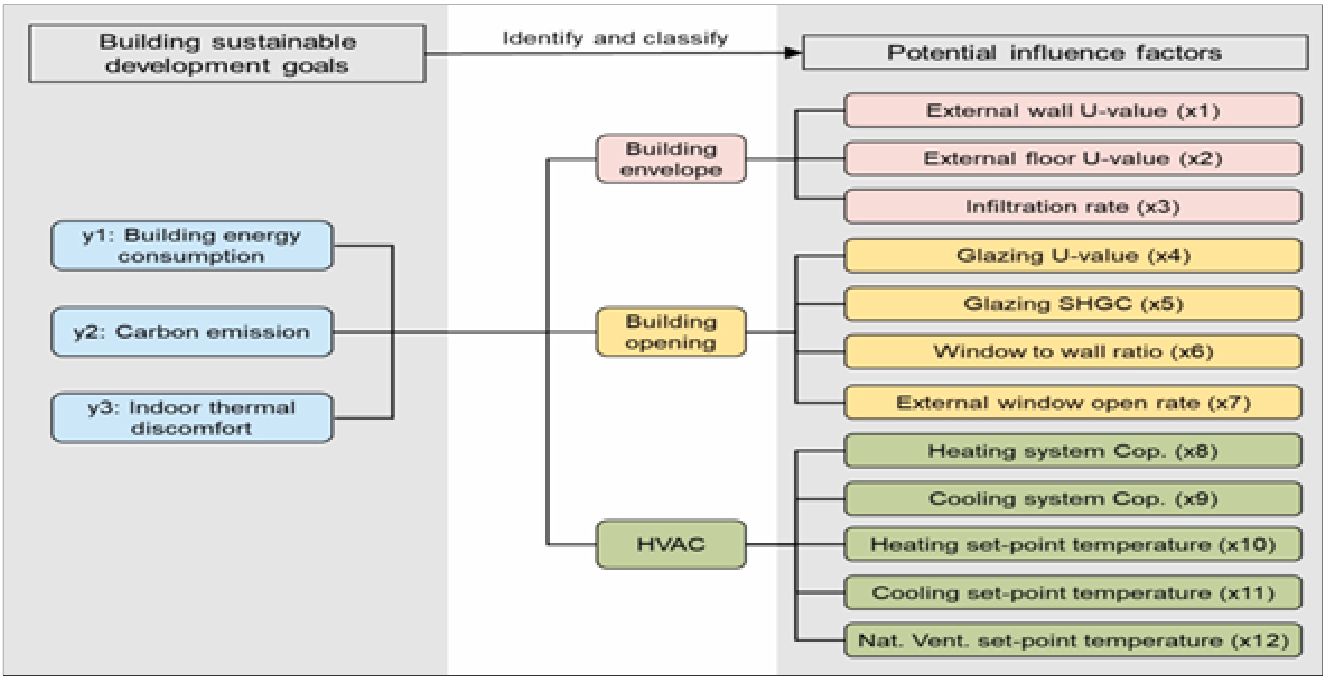

Figure 1 Flowchart of the proposed DRL approach for building performance prediction and optimization in BIM

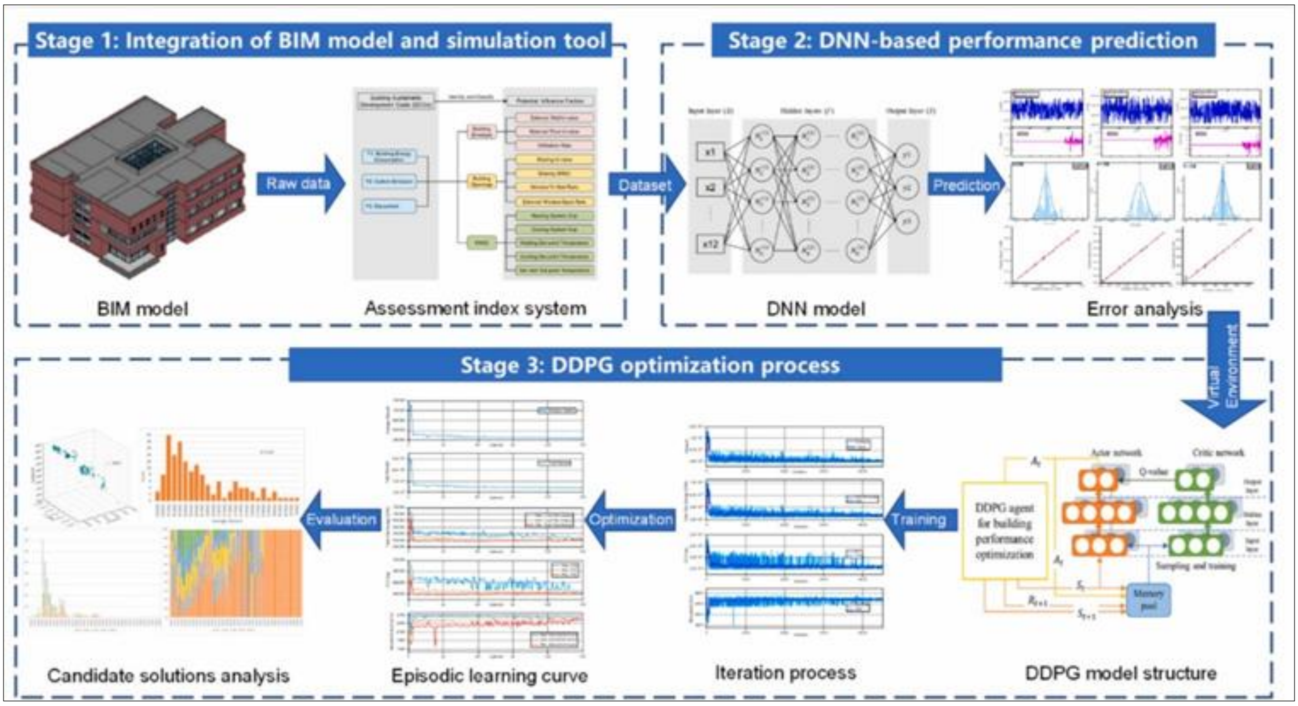

Figure 2 Assessment index system for building performance prediction and optimization. (Note: SHGC: solar heat gain coefficient, CoP.: coefficient of performance

Within the BIM framework, all relevant design parameters—ranging from geometric aspects like building layout and dimensions to physical characteristics such as thermal performance and HVAC system efficiency—are easily accessible and modifiable.

The twelve selected parameters, organized into three categories (building envelope, building openings, and HVAC systems), play a pivotal role in determining building performance. Their adaptability at the preliminary design stage is essential for meeting sustainable design objectives. The assessment focuses on three main performance targets:

- Building Energy Consumption (y1): This is measured in kilowatt-hours (kWh) per month, representing the total energy used by the building to maintain a comfortable indoor environment and meet essential needs.
- Carbon Emissions (y2): This indicates the total greenhouse gases produced over the building's entire lifecycle, serving as a key measure for decarbonization efforts.
- Indoor Thermal Discomfort (y3): This reflects the level of discomfort experienced by building occupants due to suboptimal indoor temperatures. High levels of discomfort can negatively impact occupants' well-being.

By focusing on these three targeted objectives—energy efficiency (y1), environmental sustainability (y2), and occupant comfort (y3)—the model seeks to optimize building performance for sustainability while enhancing overall indoor conditions.

The BIM model created in Revit can be saved in the green building extensible markup language (gbXML) format, which is then imported into energy simulation software, DesignBuilder (version 7.1.3.015), for building performance simulation and orthogonal testing. The gbXML format is particularly effective for preserving important building parameter data and can be easily integrated with major building energy simulation and analysis tools. Its high level of interoperability eliminates the need to repeatedly construct analysis models, making it an ideal bridge between Revit and DesignBuilder for this study.

DesignBuilder, a dynamic simulation software with a built-in EnergyPlus engine, accounts for various factors that influence building performance. It includes five core modules: a visual modeling module, an EnergyPlus dynamic thermal simulation module, an HVAC module, a computational fluid dynamics (CFD) module, and a sunshine environment module. With its intuitive graphical interface, DesignBuilder offers a cost-effective and efficient solution to accurately simulate building performance across three main objectives—building energy consumption (y1), carbon emissions (y2), and indoor thermal discomfort (y3).

In this study, orthogonal tests are conducted within DesignBuilder, involving these three objectives and twelve key variables. Ultimately, the process generates a sample dataset with 384 sets of building performance results across various scenarios, forming the foundation for data-driven optimization in green building design.

To improve the accuracy of input-output mapping in a neural network, hyperparameter tuning is essential. In this case, Bayesian optimization (BO) is used to efficiently search for optimal configurations of two key hyperparameters: the number of layers in the deep neural network (DNN) and the number of nodes in the hidden layers. Unlike the traditional grid search method, which can be slow when dealing with multidimensional parameters, BO strategically selects the next set of hyperparameters in a more informed and time-saving manner. It tracks previous evaluations to build a probabilistic model, also known as a surrogate, which enhances the efficiency of hyperparameter selection.

The steps in BO can be summarized as follows:

- Create a surrogate probability model of the objective function.
- Select hyperparameters that are expected to perform best based on the surrogate model.
- Apply these hyperparameters to the actual objective function.
- Update the surrogate model with the new results.
- Repeat steps 2–4 until the maximum number of iterations is reached.

Through this iterative process, the optimal combination of hyperparameters is identified, resulting in the minimization of the network's loss function.

3.2. DDPG Optimization Process

Building on the established DNN meta-model, the three key objectives related to building performance—energy efficiency, carbon emissions, and occupant comfort—can be accurately estimated and optimized through optimal design parameter configurations. To handle this as a multi-objective optimization (MOO) task, a DRL-based framework with two primary components is developed.

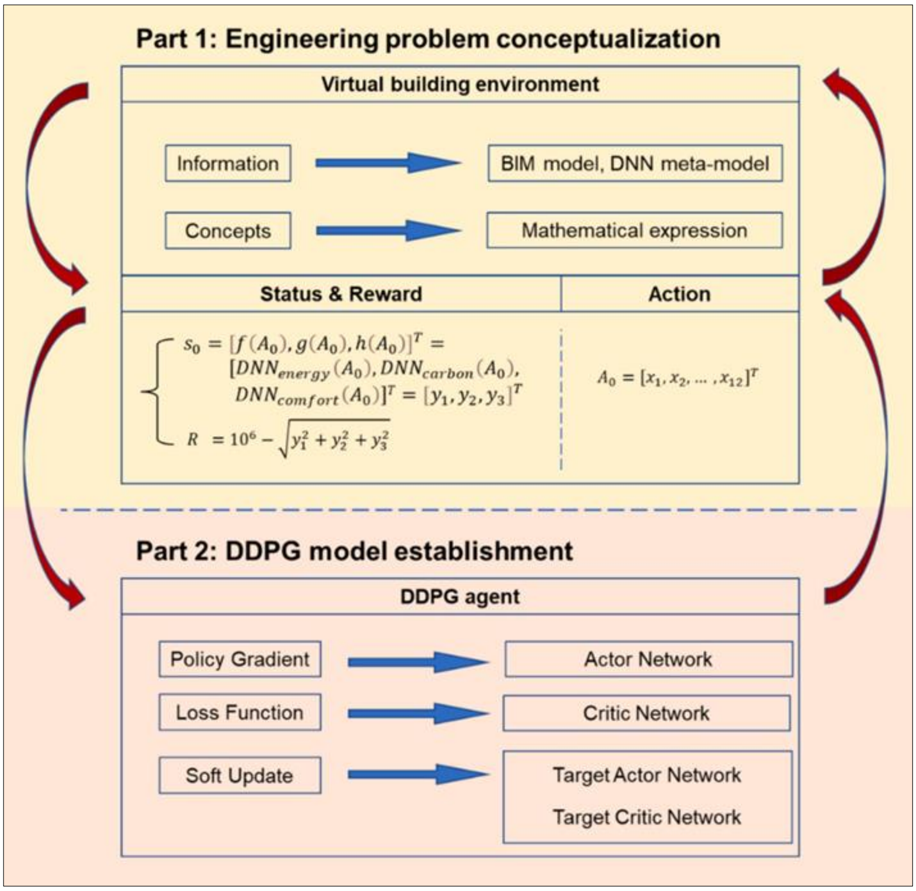

Figure 3 DDPG Optimization Process

- Component 1: This involves conceptualizing real-world engineering problems by translating the MOO task into mathematical expressions. This step specifies the environment, the set of executable actions, and the state functions that guide the optimization process.
- Component 2: This focuses on creating the DDPG model. A stable and efficient DDPG agent is developed to interact with the pre-configured environment, where it can iteratively refine its action strategies using an experience buffer. This iterative learning process enables the DDPG model to effectively evaluate the comprehensive performance of the building based on its design.

The core concept behind this DRL-based framework is the continuous acquisition of critical information from the environment, allowing the agent to adaptively adjust its responses. This dynamic policy generation process is aimed at maximizing the reward function, which is aligned with the building's real-world physical characteristics.

In essence, the optimization goal is to solve the decision-making challenge for green building design by facilitating interaction between the agent, actions, and the environment. Through this agent-action-environment interaction, the approach is expected to significantly improve overall building performance.

3.2.1. Engineering Problem Conceptualization

To address the multi-objective optimization (MOO) problem, a DRL-based framework with a focus on reliability, flexibility, and efficiency is proposed. By continuously receiving feedback in the form of states and rewards from the environment, the DRL agent can self-update and adapt the building design strategy to the current conditions. Specifically, the agent—comprising neural networks—learns how to solve the problem through trial and error, leveraging the data provided by the environment. It acts as a virtual decision-maker, selecting actions based on calculated rewards generated through its interaction with the environment. In turn, the virtual environment updates its state based on the agent's actions, creating a dynamic feedback loop that reflects real-world engineering principles.

In this scenario, the virtual environment is constructed using the BIM model and the DNN meta-model, allowing for rapid feedback on building performance as well as the configuration of design parameters. This structure enables the derivation of optimal strategies related to the building envelope, openings, and HVAC systems, serving as key benchmarks for green building design. By integrating these components, the framework allows for the simultaneous optimization of the three main objectives: energy consumption, carbon emissions, and occupant comfort.

Mathematically, the DRL-based framework is characterized by four elements, denoted as (A, S, P, R):

- A: Actions taken by the agent.
- S: States representing the current environment.
- P: State transition probabilities, defining how the environment reacts to specific actions.
- R: Rewards, which guide the agent in making decisions that maximize performance aligned with the building's design objectives.

This system empowers the agent to make informed and adaptive decisions, continuously optimizing the building design strategy for sustainable development.

Action (A)

Actions taken by the agent are intended to transition the current state to a new state. Different actions yield different outcomes, and the goal is to select the optimal actions from the action space \(a \in A \), aiming to maximize rewards in a given scenario. In the context of the virtual environment, the building design parameters act as the agent's actions within the DRL framework. There are twelve building design parameters in total, which can be adjusted within a defined range of values. Finding the optimal combination of these parameters requires numerous trials and iterative feedback loops, a task well-suited for DRL's capacity to generate action functions efficiently.

The state represents the current condition of the environment within which the agent operates. As an interface between the agent's perception and the environment, the state \(s \in S \) provides the necessary information for the agent to make decisions. This allows for effective exploration of the broad range of potential decisions within the search space. Based on the state, the agent can plan its subsequent actions, encapsulating the current scenario and its objectives.

The state function is described as follows:

- The state transition probability represents the likelihood that the agent will transition from one state to another within a defined environment. It determines how the environment responds to a particular action. Typically, a state transition probability matrix is used, where each row outlines the probabilities of transitioning from one state to various potential successor states. However, in a model-free DRL approach, this state transition probability \(P \) is unknown and can be excluded from direct consideration in the optimization process.
- Once appropriate action and state functions are determined for efficient communication between the agent and the virtual environment, a performance evaluation is necessary to refine the optimization process. This is achieved through the formulation of a reward function, which acts as a guiding principle for the generation of actions and updates to states. The reward function is central to decision-making, directing the agent toward more favorable outcomes based on continuous interaction with the environment.
- In the case of building design optimization, as the agent adjusts building design parameters, it seeks to find the optimal combination that enhances building performance relative to the baseline scenario. The reward function plays a crucial role in evaluating the agent's decisions and steering the performance improvements towards the desired outcomes.
- Given that each optimization objective (such as energy consumption, carbon emissions, and occupant comfort) has a unique potential for improvement, defining the reward function presents a challenge due to inherent trade-offs. To address this, the reward function is based on the Euclidean distance from the origin, which serves as a criterion for decision-making.
- The underlying principle is that during the DRL training process, the reward function aims to minimize the Euclidean distance between the current performance (solution coordinates) and the origin (representing the optimal solution). The reward function is expressed as:
- This formulation ensures that as the agent learns and iterates, the reward function converges towards the optimal value, driving the building design to achieve the best possible performance.

3.2.2. DDPG Model Establishment

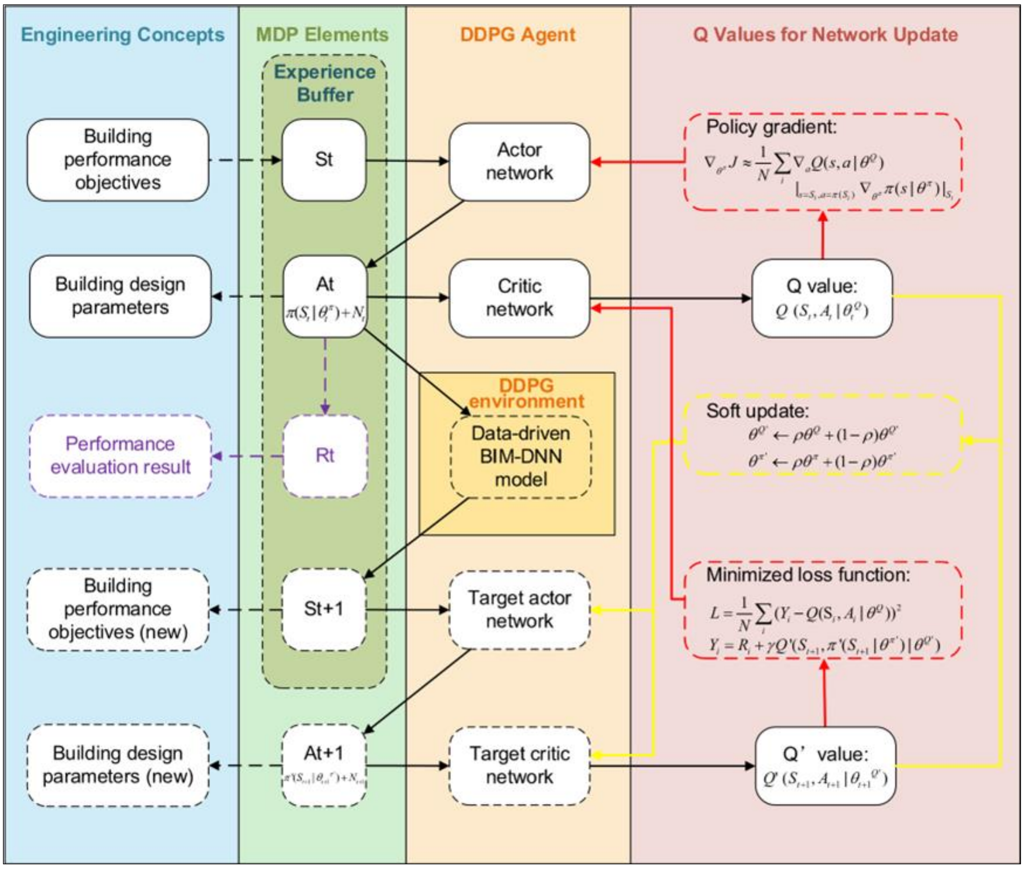

Figure 4 DDPG Model Establishment

The Deep Deterministic Policy Gradient (DDPG) algorithm is part of the actor-critic family of reinforcement learning (RL) algorithms. It consists of two distinct neural networks: the actor network and the critic network, which work together to simultaneously learn the Q-function and the policy. The DDPG model leverages the deterministic policy gradient, making it particularly suitable for environments with continuous action spaces.

- Actor Network: This network receives the current state as input and outputs the corresponding action. Its role is to map states to actions by optimizing the policy to maximize long-term rewards.
- Critic Network: The critic network, on the other hand, takes both the state and action as inputs and outputs a policy gradient, which guides the actor network in improving its decisions. A higher value of the policy gradient indicates a better match between the current state and action, leading to more effective optimization.

By combining the advantages of both value function and policy parameterization, the DDPG algorithm provides a powerful framework for learning optimal policies in continuous action domains. The state is fed into the actor network to determine the best possible action, while the critic network evaluates this action to inform further policy updates.

In summary, a DDPG agent functions as an actor-critic RL agent, continuously seeking the optimal policy that maximizes the reward. The DDPG agent's structure is visually represented in Figure 5, which shows the integration of two target networks—one for the actor and one for the critic. The agent learns by updating the actor and critic networks at each time step, utilizing a circular experience buffer to store past experiences, and introducing small perturbations to actions selected by the policy to explore new strategies.

This process is formalized in the pseudo-code of the DDPG algorithm (Algorithm 1), which outlines the key steps for training the DDPG agent. It involves continuously updating the actor and critic networks based on feedback from the environment, with the goal of improving the agent's decision-making and ultimately achieving optimal building design performance.

In DDPG, the deterministic policy gradient is incorporated into the actor component, guiding the update of the policy network parameters according to Eq. (5):

While directly updating the critic network using these equations can cause significant fluctuations, DDPG addresses this issue by introducing the concept of target networks, similar to DQN. Instead of directly copying weights, the target network is updated using a soft update method, which smooths the transition and improves the stability of the algorithm. The target networks for both the critic and actor are updated as follows:

A notable limitation of the Deterministic Policy Gradient (DPG) algorithm is its limited ability to explore the environment effectively. To address this, the DDPG algorithm incorporates random noise into the action selection process, which introduces a degree of randomness and enhances policy exploration. This exploration process is defined by the equation:

In this case study, the DDPG model is tasked with controlling the building parameters within a physically realistic range. The goal is to iteratively optimize and adjust the green building's performance through the learning mechanism built into the DDPG algorithm. The process begins with the actor network, which evaluates the current performance of the building based on its objectives and devises an action function that aims to maximize the Q value by combining the current state functions. These actions generate a new set of building parameter combinations.

The resulting building parameters, along with the performance metrics, are passed to the critic network. The critic network then computes a new Q value, which is used to update the actor network through a gradient descent process. The agent uses these newly generated parameter combinations to interact with the environment, which is represented by the DNN models, to update the building's performance metrics. This updated information is transmitted to the target actor network to determine the next step's building parameters.

The state function and action function values are then processed by the target critic network, resulting in a new target Q value, along with the reward derived from the environment interaction. The critic network is updated by minimizing the error between the expected and actual Q values. Both the target actor and target critic networks are updated incrementally through a soft update method to ensure stability.

Throughout the iterative process, the DDPG model stores four key pieces of information:, which represent the current state of building performance, the current combination of building parameters, the subsequent state, and the reward value. These items are stored in an experience buffer. When the model needs to be updated, data can be randomly sampled from the experience buffer to improve learning efficiency and performance.

Table 1 Detail settings of basic parameters for BIM-based building performance simulation

Building Parameter	Value
Mean outdoor temperature (°C)	20
Indoor design temperature (°C) Equipment power density (W/m^2)	25 11.77
Indoor lighting power (W/m2-100lx) Occupancy density (people/m^2)	50.11
Fresh air volume (vs/person)	8

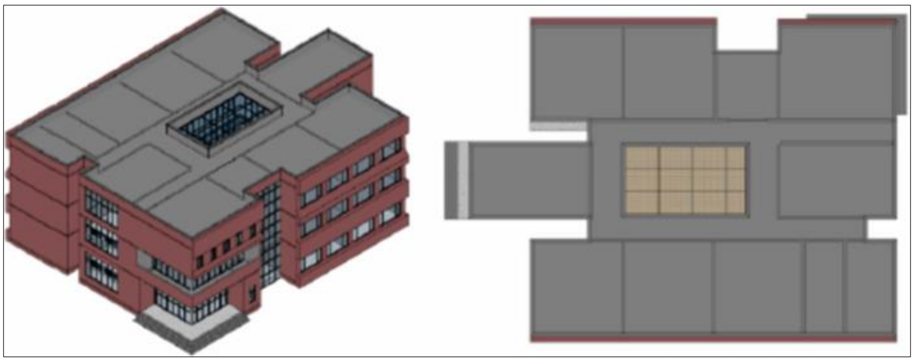

Figure 5 BIM model of the teaching building

4. Case Study

4.1. Setting of BIM Model

In this case study, a series of experiments based on the developed DRL framework is conducted using a three-story teaching building, as shown in Figure 6. The building is located in Shanghai, China, and is constructed primarily with a reinforced concrete frame. It covers a total area of approximately 2381.8 m^2 with a height of 14.1 meters. The building consists of 27 rooms serving various functions, including classrooms, conference rooms, lounges, restrooms, and stairwells. It is oriented north-south and exposed to Shanghai's hot-summer and cold-winter climate, along with the sun's annual light cycle. The building operates entirely on electricity, utilizing a centralized HVAC system to maintain indoor temperatures.

Given the size of the building, three key performance indicators are particularly relevant for the subsequent data-driven analysis: energy consumption, carbon emissions, and indoor thermal discomfort. These factors are critical to assessing building performance and sustainability.

The goal of this BIM-based green building design task is to develop an optimized design scheme that addresses these three indicators, creating a more sustainable and resilient built environment. By focusing on these aspects, the aim is to reduce the building's negative environmental impacts while improving occupant well-being.

Before proceeding with the prediction and optimization tasks, a detailed BIM model containing physical property information—such as geometric data and material thermal properties—was created using Autodesk Revit 2021. This model includes all relevant details about the building's structure, including walls, floor slabs, doors, and windows. The BIM model was then imported into the DesignBuilder software (version 7.1.3.015) using the gbXML format for energy simulation.

Although the initial BIM model contains extensive attribute data, additional refinements were made to ensure the accuracy of the energy simulations. Essential settings, such as geometric parameters and building elements, were calibrated based on project requirements and engineering codes to ensure precise simulations. Table 1 outlines these critical settings, which contribute to the fidelity and reliability of the analysis.

Based on the BIM model and the basic parameter settings, building energy simulations were carried out in DesignBuilder using orthogonal experiments. This structured approach ensures that the data generated is both robust and representative, providing an accurate depiction of building performance.

The simulation involves adjusting twelve predefined building design parameters (as shown in Figure 2) in DesignBuilder, resulting in outcomes for three primary building performance metrics: energy consumption (y1), carbon emissions (y2), and indoor thermal discomfort (y3). Each simulation, with a unique set of design parameters, is independently executed in DesignBuilder. It was observed that different combinations of these parameters produce varying results in terms of building performance.

The range of values derived from the simulation for the three objectives is as follows:

- Energy consumption (y1): [558,128.3 kWh, 756,775.8 kWh]
- Carbon emissions (y2): [282,091.5 kg, 340,015.7 kg]
- Indoor thermal discomfort (y3): [2261.4 hours, 3659.5 hours]

This process yielded a dataset containing 384 sets of building performance simulation results, which serves as the foundation for training the DNN meta-model and applying the DRL framework.

Additionally, the Euclidean distance between the three objectives is calculated to evaluate the proximity of each set of parameters to the optimal solution. The parameter set with the smallest Euclidean distance is selected as the baseline, allowing for a quantitative comparison of building performance before and after optimization. This comparison will help measure the extent of improvement achieved through optimization.

4.2. Meta-Model Development for Building Performance Prediction

A DNN meta-model is developed based on the 384 simulation results to understand the complex relationship between twelve design parameters and the three target objectives (energy consumption, carbon emissions, and indoor thermal discomfort). The dataset obtained from the BIM-based simulation process is divided into a training set and a test set at a ratio of 3:1. This random partitioning is done to eliminate any potential order-related biases. A total of 288 data points are randomly selected for regression fitting and model parameter determination, allowing the DNN model to learn from diverse data samples. The remaining 96 data points (25% of the total) are used as an evaluation test set to assess the prediction performance on unseen data. A fixed random seed value of 100 is used to ensure the reproducibility of the data partitioning.

To prevent common issues such as overfitting and training instability, early stopping and dropout techniques are applied during model training. Hyperparameter tuning is critical for improving model accuracy, and Bayesian Optimization (BO) is employed to automatically find the optimal configuration for the DNN model. The results of this tuning process, including the final hyperparameter values, are provided in Table 2.

Two primary evaluation metrics—R^2 (coefficient of determination) and MAE (mean absolute error)—are used to assess the DNN meta-model's performance in both the training and test sets. R^2 measures the goodness of fit, while MAE provides insight into the average deviation between predicted and actual values, expressed as a percentage. In both the training and test sets, R^2 values are close to 1, and the MAE remains below 0.05, indicating the absence of overfitting or underfitting. These metrics demonstrate that the DNN model is capable of producing accurate and reliable predictions for building performance.

For visual clarity, the prediction results and corresponding errors are illustrated in Figures 7 and 8. The close alignment between the predicted values (blue line) and the actual simulation results (red line) confirms the DNN model's high degree of accuracy in predicting all three objectives under various combinations of design parameters. Furthermore, Figure 8 shows the distribution of prediction errors, which follows a normal distribution curve. Almost all error values fall within 2σ, indicating effective error control during model training.

In conclusion, the well-constructed DNN meta-model provides accurate predictions, which are essential for further integration with the DRL framework. This ensures that the DNN model can effectively support the DRL-based optimization process, resulting in better decision-making for green building design.

4.3. Building Performance Optimization via DRL Approach

4.3.1. DDPG Parameter Setting

The DNN meta-model, developed earlier, effectively captures the relationship between building performance and design parameters across various scenarios, offering reliable predictions. The next step is to integrate this meta-model into the Deep Reinforcement Learning (DRL) framework to address the multi-objective optimization (MOO) task. This task aims to find an optimal green building design solution that balances three key objectives: building energy consumption (y1), carbon emissions (y2), and indoor thermal discomfort (y3).

The DDPG model defines an efficient agent that continuously interacts with and learns from the virtual environment, which is simulated using BIM and DNN models. During the DDPG training process, parameters must be continuously adjusted to expand the optimization range and enhance the performance of the DRL framework. The DDPG parameter settings, listed in Table 4, help improve the generalizability and effectiveness of the DRL-based approach. Below is an explanation of the key parameter settings:

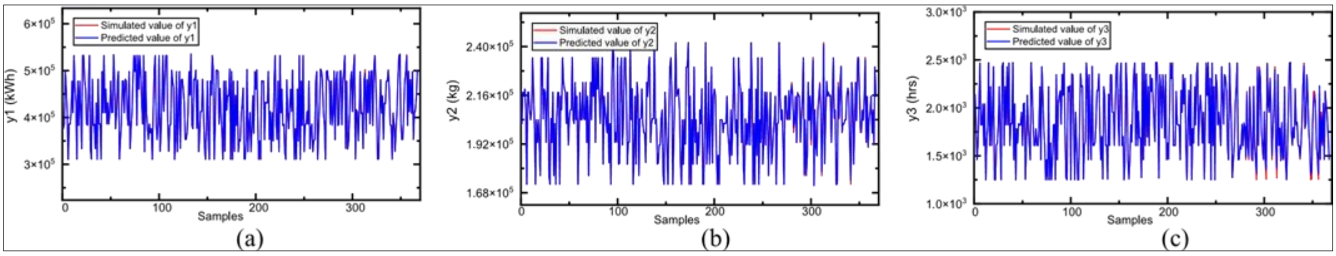

Figure 6 DRL Approach

Discount Factor for Reward Function

The discount factor controls how cumulative rewards are calculated during the DDPG training. A smaller discount factor reduces the impact of future rewards, while a value of 1 ensures that the effect of the reward does not diminish over time. This allows the agent to consider long-term benefits during optimization iterations.

Learning Rate, Network Layers, and Batch Size

The learning rate determines the step size during gradient updates. If the learning rate is too large, it may fail to converge to the optimal gradient; if too small, the convergence rate will be slow. Typically, the actor and critic networks require fewer layers than the DNN, generally between 2–3 layers, as excessive layers can destabilize the algorithm. The batch size also plays an important role in training speed—larger batch sizes lead to faster convergence in the DDPG algorithm.

Upper and Lower Bounds of Building Parameters

Since the DDPG algorithm optimizes a set of 12 building parameters, it is crucial to define the upper and lower bounds for these parameters. These limits are derived from the original dataset and reflect the practical constraints and physical significance of the building design elements.

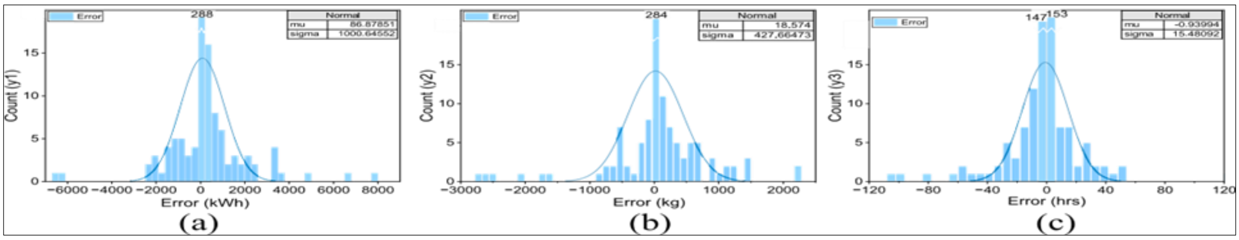

Figure 7 Bounds of Building Parameters

Training Steps and Episodes

To ensure the DDPG model has enough time to learn, a sufficient number of training steps and episodes are needed. These extended interactions with the environment allow the agent to gather ample data regarding the building's key variables. Through trial and error, the DDPG agent gradually improves its optimization paths. In this case, 150 training episodes with a maximum of 30 steps per episode are set to provide enough learning opportunities.

Exploration through Noise Addition

The DDPG algorithm trains a deterministic policy in an off-policy manner, meaning the agent consistently follows a single policy. However, this can limit exploration at the start of training. To overcome this, noise is added to the action function values during training. Gaussian noise with a mean of 0 and variance of 1 is used to introduce randomness in the early stages. A weight discount factor of 0.9998 is applied to gradually reduce the noise variance during iterations, balancing exploration and exploitation, and ultimately improving the global solution optimization efficiency.

These carefully tuned parameters enable the DDPG agent to efficiently explore the design space and generate optimized building designs that meet the objectives of energy efficiency, reduced emissions, and occupant comfort in a balanced and automated manner.

4.3.2. DDPG Learning Curve

During the training process of the DDPG model, the agent iteratively generates the corresponding state function based on the initial action function. It then evaluates the optimization effects on the building performance objectives—energy consumption (y1), carbon emissions (y2), and indoor thermal discomfort (y3)—and reflects these results via the reward function. This process is executed on a workstation equipped with an Intel Core i9-10920X CPU and an NVIDIA GeForce RTX 3080 GPU. The iterations continue until the stopping criterion is reached. The fluctuations and convergence of the reward function, along with the performance objectives for each step, are illustrated in Figure 9, representing the DDPG learning curve.

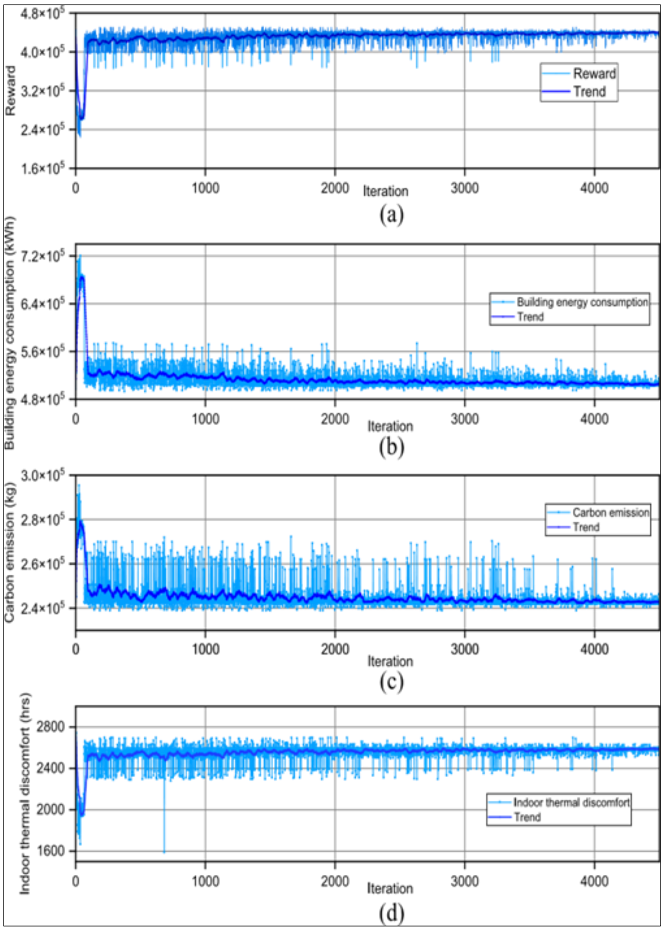

Figure 8 Learning Curve

The DDPG model undergoes two distinct phases: learning and convergence. - Initial Learning and Exploration Phase (First 200 Iterations):

In this phase, substantial optimization of the reward function occurs, with its value dropping from approximately 75,000 to 60,000. The unstable nature of this phase is likely due to the exploration of different optimization paths, as the model searches for potential candidates for optimization. The large fluctuation in Gaussian noise variance generated by the DDPG agent during this phase indicates that the agent is engaging in random exploration, trying to identify the best solutions across various optimization directions.

- Convergence Phase (After 200 Iterations):

After around 200 iterations, the DDPG model enters a convergence phase. During this phase, fluctuations in the reward function gradually diminish, and the model's performance stabilizes. This signifies that the strategy has been sufficiently optimized and no further improvements are being made. By the end of the training process, the DDPG model can return optimal design decisions for green building performance.

To visualize the optimization trends more clearly, the arithmetic mean of the rewards and performance objectives over 30 iterations is calculated in a moving average manner. The trend line generated from this approach confirms the stable convergence of the DDPG model, demonstrating the effectiveness of the training. As the learning curve stabilizes, it reflects the DDPG agent's capacity to consistently make better decisions and generate more efficient design solutions for green building performance.

Learning Curve Analysis for Building Performance Objectives

The learning curve of the three objectives—building energy consumption (y1), carbon emissions (y2), and indoor thermal discomfort (y3)—can also be visualized by plotting the episode number on the x-axis, as shown in Figure 10. The DDPG model continuously iterates through cycles of learning until it either reaches the desired state or the maximum number of steps per episode. In this investigation, a total of 150 episodes were conducted, with each episode consisting of a maximum of 30 steps, yielding 4500 iterations.

In Figure 10, the three trendlines represent different aspects of the learning process:

- The orange trendline shows the average value derived from the 30 steps in each episode.
- The blue trendline indicates the maximum value.
- The red trendline represents the minimum value.

These trendlines help to visualize the fluctuation range and iterative convergence of each objective throughout the learning process. As the training progresses, all three objectives display a tendency to stabilize during the later episodes, indicating convergence.

For example, in the case of building energy consumption (y1) (Figure 10a), the blue and red lines represent the fluctuation interval for energy consumption during episodic learning. Initially, the interval spans a range from 200,000 to 50,000, but as the learning process continues, this interval narrows, with uncertainty decreasing. This reduction in fluctuation is a key indicator that the DDPG model is approaching the optimal solution for minimizing energy consumption.

The orange mean curve for energy consumption remains closer to the red minimum curve rather than the blue maximum curve, signifying that the training process effectively minimizes the occurrence of outliers on the upper end of the spectrum. This shows that the model is gradually refining its predictions and focusing more on achieving lower energy consumption values.

The learning curves for carbon emissions (y2) exhibit a similar pattern to energy consumption, demonstrating the model's ability to simultaneously optimize multiple objectives. However, the learning curve for indoor thermal discomfort (y3) displays a different trend, reflecting the trade-offs inherent in balancing the three objectives in a multi-objective optimization (MOO) task.

Overall, the fluctuation magnitude and direction of convergence for each objective provide insight into how the DDPG model manages the complex trade-offs in the green building design optimization process.

5. Conclusion

This study introduces a deep reinforcement learning (DRL)-based multi-criteria optimization (MCO) framework that successfully integrates Building Information Modeling (BIM) and a deep neural network (DNN) model to optimize sustainable building design. By leveraging the deep deterministic policy gradient (DDPG) algorithm, the framework effectively balances energy consumption, carbon emissions, and indoor thermal comfort, demonstrating superior performance in a case study of a teaching building in Shanghai, with a 13.19% improvement over traditional genetic algorithms. This approach not only offers advanced data-driven decision-making capabilities but also presents a promising solution for automating green building design to meet sustainability goals, enhancing operational efficiency, and improving occupant well-being.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] X. Li, Y. Pan, L. Zhang, J. Chen, Dynamic and explainable deep learning-based risk prediction on adjacent building induced by deep excavation, Tunn. Undergr. Space Technol. 140 (2023) 105243, https://doi.org/10.1016/j.tust.2023.105243.

[2] H.-J. Zhao, W. Liu, P.-X. Shi, J.-T. Du, X.-M. Chen, Spatiotemporal deep learning approach on estimation of diaphragm wall deformation induced by excavation, Acta Geotech. 16 (11) (2021) 3631–3645, https://doi.org/10.1007/s11440-021- 01264-z.

[3] Y. Shen, Y. Pan, BIM-supported automatic energy performance analysis for green building design using explainable machine learning and multi-objective optimization, Appl. Energy 333 (2023) 120575, https://doi.org/10.1016/j. apenergy.2022.120575.

[4] H.H. Hosamo, M.S. Tingstveit, H.K. Nielsen, P.R. Svennevig, K. Svidt, Multiobjective optimization of building energy consumption and thermal comfort based on integrated BIM framework with machine learning-NSGA II, Energ. Buildings 277 (2022) 112479, https://doi.org/10.1016/j.enbuild.2022.112479.

[5] S. Sharma, V. Chahar, A comprehensive review on multi-objective optimization techniques: past, present and future, Arch. Comput. Methods Eng. 29 (2022) 3, https://doi.org/10.1007/s11831-022-09778-9.

[6] M. Feurer, K. Eggensperger, E. Bergman, F. Pfisterer, B. Bischl, F. Hutter, Mind the gap: measuring generalization performance across multiple objectives, in: B. Cr´emilleux, S. Hess, S. Nijssen (Eds.), Advances in Intelligent Data Analysis XXI, Springer Nature, Switzerland, Cham, 2023, pp. 130–142, https://doi.org/10.1007/ 978-3-031-30047-9_11.

[7] K. Li, T. Zhang, R. Wang, Deep reinforcement learning for multiobjective optimization, IEEE Trans. Cybernet. 51 (6) (2021) 3103–3114, https://doi.org/ 10.1109/TCYB.2020.2977661.

[8] Y. Pan, J. Qin, L. Zhang, W. Pan, J.J. Chen, A probabilistic deep reinforcement learning approach for optimal monitoring of a building adjacent to deep excavation, Comput. Aided Civ. Inf. Eng. (2023), https://doi.org/10.1111/ mice.13021.

[9] O.A. Olubunmi, P.B. Xia, M. Skitmore, Green building incentives: a review, Renew.

[10] Sust. Energ. Rev. 59 (2016) 1611–1621, https://doi.org/10.1016/j.rser.2016.01.028.

[11] Parekh, Ruchit. Constructing Wellness: Harnessing AI for a Sustainable and Healthy Future. Elsevier, 2024.

[12] K. Roy, A. Dani, V. Say, Z. Fang, J. Lim, The circular economy of steel roofing and cladding and its environmental impacts—a case study for New Zealand, Sustainability 14 (2022), https://doi.org/10.3390/su142416832.

[13] V. Borkovskaya, Environmental and economic model life cycle of buildings based on the concept of “green building”, Appl. Mech. Mater. 467 (2013) 287–290, https://doi.org/10.4028/www.scientific.net/AMM.467.287.

[14] K. Guo, Y. Yuan, Geographic distribution and influencing factor analysis of green residential buildings in China, Sustainability 13 (2021) 12060, https://doi.org/ 10.3390/su132112060.

[15] D.T. Doan, A. Ghaffarianhoseini, N. Naismith, T. Zhang, A. Ghaffarianhoseini, J. Tookey, A critical comparison of green building rating systems, Build. Environ. 123 (2017) 243–260, https://doi.org/10.1016/j.buildenv.2017.07.007.

[16] V. Pai, H. Elzarka, Whole building life cycle assessment for buildings: a case study ON HOW to achieve the LEED credit, J. Clean. Prod. 297 (2021) 126501, https:// doi.org/10.1016/j.jclepro.2021.126501.

[17] R. Singh, K. Singh, R. Vyas, “Green building”: a step towards environmental and economic construction, Int. J. Environ. Clim. Change (2019) 391–401, https://doi. org/10.9734/ijecc/2019/v9i730124.

[18] L. Chen, X. Gao, G. Shitao, Z. Li, Regionalization of green building development in China: a comprehensive evaluation model based on the catastrophe progression method, Sustainability 12 (2020) 5988, https://doi.org/10.3390/su12155988.

[19] A. Alwisy, S. BuHamdan, M. Gül, Criteria-based ranking of green building design factors according to leading rating systems, Energ. Buildings 178 (2018) 347–359, https://doi.org/10.1016/j.enbuild.2018.08.043.

[20] T. Olawumi, D.D. Chan, Green-building information modelling (Green-BIM) assessment framework for evaluating sustainability performance of building projects: a case of Nigeria, Archit. Eng. Des. Manag. 17 (2021) 458–477, https:// doi.org/10.1080/17452007.2020.1852910.

[21] Parekh, Ruchit. Blueprint for Sustainability: LEED Implementation in Commercial Projects. Elsevier, 2024.

[22] T. Nguyen, S.H. Toroghi, F. Jacobs, Automated green building rating system for building designs, J. Archit. Eng. 22 (4) (2016) A4015001, https://doi.org/ 10.1061/(ASCE)AE.1943-5568.0000168.

[23] Parekh, Ruchit, and Charles Smith. "Innovative AI-driven software for fire safety design: Implementation in vast open structure." *World Journal of Advanced Engineering Technology and Sciences* 12.2 (2024): 741-750.

[24] R.S. Sutton, A.G. Barto, Reinforcement learning: an introduction, IEEE Trans. Neural Netw. 9 (5) (1998) 1054, https://doi.org/10.1080/ 15623599.2022.203341910.1109/TNN.1998.712192.

[25] G.H. Erharter, T.F. Hansen, Z. Liu, T. Marcher, Reinforcement learning based process optimization and strategy development in conventional tunneling, Autom.

[26] Constr. 127 (2021) 103701, https://doi.org/10.1016/j.autcon.2021.103701.

[27] T.T. Nguyen, N.D. Nguyen, S. Nahavandi, Deep reinforcement learning for multiagent systems: a review of challenges, solutions, and applications, IEEE Trans. Cybernet. 50 (9) (2020) 3826–3839, https://doi.org/10.1109/ TCYB.2020.2977374.

[28] V. Mnih, K. Kavukcuoglu, D. Silver, A.A. Rusu, J. Veness, M.G. Bellemare, A. Graves, M. Riedmiller, A.K. Fidjeland, G. Ostrovski, Human-level control through deep reinforcement learning, Nature 518 (7540) (2015) 529–533, https:// doi.org/10.1038/nature14236.

[29] J. Xu, H. Zhang, J. Qiu, A deep deterministic policy gradient algorithm based on averaged state-action estimation, Comput. Electr. Eng. 101 (2022) 108015, https://doi.org/10.1016/j.compeleceng.2022.108015.

[30] S. Nageshrao, H.E. Tseng, D. Filev, Autonomous highway driving using deep reinforcement learning, in: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), 2019, pp. 2326–2331, https://doi.org/10.1109/ SMC.2019.8914621.

[31] C. Li, P. Zheng, Y. Yin, B. Wang, L. Wang, Deep reinforcement learning in smart manufacturing: a review and prospects, CIRP J. Manuf. Sci. Technol. 40 (2023) 75–101, https://doi.org/10.1016/j.cirpj.2022.11.003.

[32] D.F. Leguizamo, H.-J. Yang, X.Y. Lee, S. Sarkar, Deep reinforcement learning for robotic control with multi-Fidelity models, IFAC-PapersOnLine 55 (37) (2022) 193–198, https://doi.org/10.1016/j.ifacol.2022.11.183.

[33] Q. Fu, Z. Han, J. Chen, Y. Lu, H. Wu, Y. Wang, Applications of reinforcement learning for building energy efficiency control: a review, J. Build. Eng. 50 (2022) 104165, https://doi.org/10.1016/j.jobe.2022.104165.

[34] R.S. Sutton, D. McAllester, S. Singh, Y. Mansour, Policy gradient methods for reinforcement learning with function approximation, in: Proceedings of the 12th International Conference on Neural Information Processing Systems, MIT Press, Denver, CO, 1999, pp. 1057–1063.

[35] T. Lillicrap, J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra, Continuous control with deep reinforcement learning, CoRR (2015), https://doi. org/10.48550/arXiv.1509.02971.

[36] Y. Wu, H. Tan, J. Peng, H. Zhang, H. He, Deep reinforcement learning of energy management with continuous control strategy and traffic information for a series- parallel plug-in hybrid electric bus, Appl. Energy 247 (2019) 454–466, https://doi. org/10.1016/j.apenergy.2019.04.021.

[37] Parekh, Ruchit. "Automating the design process for smart building technologies." *World Journal of Advanced Research and Reviews* 23 (2024): 1213-1234.

[38] Energy 255 (2019) 113762, https://doi.org/10.1016/j.apenergy.2019.113762.

[39] G. Gao, J. Li, Y. Wen, DeepComfort: energy-efficient thermal comfort control in buildings via reinforcement learning, IEEE Internet Things J. 7 (9) (2020) 8472–8484, https://doi.org/10.1109/JIOT.2020.2992117.

[40] L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, Z. Sun, L. Zhang, Y. Zhang, T. Jiang, Deep reinforcement learning for smart home energy management, IEEE Internet Things J. 7 (4) (2019) 2751–2762, https://doi.org/10.1109/JIOT.2019.2957289.

[41] H. Wang, W. Liang, L.-F. Yu, Scene mover: automatic move planning for scene arrangement by deep reinforcement learning, ACM Trans. Graph. 39 (6) (2020), https://doi.org/10.1145/3414685.3417788. Article 233.

[42] L. Mandow, J.-L. P´erez-de-la-Cruz, A.B. Rodríguez-Gavilan, M. Ruiz-Montiel, ´ Architectural planning with shap grammars and reinforcement learning: