

Operator learning for energy-efficient building ventilation control with computational fluid dynamics simulation of a real-world classroom

Yuexin Bian^a, Oliver Schmidt^b, Yuanyuan Shi^{a,*} 

^a Department of Electrical and Computer Engineering, University of California San Diego, La Jolla 92037, USA

^b Department of Mechanical and Aerospace Engineering, University of California San Diego, La Jolla 92037, USA

HIGHLIGHTS

- A novel data-driven neural operator learning framework for building ventilation control to optimize energy efficiency and indoor air quality.
- Propose an ensemble neural operator–transformer model to predict the spatiotemporal evolution of indoor CO₂ fields, achieving higher accuracy than individual neural operator models and a 250,000× speed-up over CFD simulations.
- Case studies show that our approach achieves energy savings compared to data-driven control with spatially averaged or deep learning–based reduced-order models, while still satisfying indoor air quality requirements.
- Release an open-access CFD-based building dataset with airflow and CO₂ fields for ventilation control benchmarking.

ARTICLE INFO

Keywords:

Building energy systems
Ventilation control
Energy efficiency
Indoor air quality
Neural operator learning

ABSTRACT

Energy-efficient ventilation control plays a vital role in reducing building energy consumption while ensuring occupant health and comfort. While Computational Fluid Dynamics (CFD) simulations provide detailed and physically accurate representations of indoor airflow, their high computational cost limits their use in real-time building control. In this work, we present a neural operator learning framework that combines the physical accuracy of CFD with the computational efficiency of machine learning to enable building ventilation control with the high-fidelity fluid dynamics models. Our method jointly optimizes the airflow supply rates and vent angles to reduce energy use and adhere to air quality constraints. We train an ensemble of neural operator transformer models to learn the mapping from building control actions to airflow fields using high-resolution CFD data. This learned neural operator is then embedded in an optimization-based control framework for building ventilation control. Experimental results show that our approach achieves significant energy savings compared to maximum airflow rate control, rule-based control, as well as data-driven control methods using spatially averaged CO₂ prediction and deep learning–based reduced-order models, while consistently maintaining safe indoor air quality. These results highlight the practicality and scalability of our method in maintaining energy efficiency and indoor air quality in real-world buildings.

1. Introduction

Buildings account for nearly 40 % of global energy consumption [1], with Heating, Ventilation, and Air Conditioning (HVAC) systems being among the primary contributors. Effective ventilation is critical not only for reducing energy use but also for maintaining indoor air quality, which directly impacts occupant health and comfort [2]. The urgency of this issue has grown in light of the COVID-19 pandemic, as

building system operators and public health agencies have increasingly emphasized the need for adaptive ventilation systems that respond to occupancy and pollutant levels throughout the indoor space [3].

Despite this urgency, most building HVAC systems still rely on fixed or rule-based ventilation strategies [4–6], which are often overly conservative. For instance, maximum fresh-air intake policies implemented at UC San Diego [6] in response to health concerns have led to 2–2.5× increases in energy usage. This highlights the critical trade-off between

* Corresponding author.

Email address: yyshi@eng.ucsd.edu (Y. Shi).

maintaining indoor air quality and minimizing energy consumption, and the need for intelligent control methods that balance these competing demands.

Recent advances in data-driven control approaches, including model predictive control (MPC) [7–10] and reinforcement learning (RL) [11–15], have shown strong potential for building ventilation management. However, most existing methods represent indoor air states—such as CO₂ concentration or temperature—using spatially aggregated variables, typically measured at a single point or averaged over an entire zone. For example, MPC and RL-based strategies often model CO₂ dynamics using ordinary differential equations (ODEs) [14,16] or neural networks [7,8,15] that predict future mean values based on current measurements and control inputs. While computationally efficient, these low-dimensional representations ignore spatial variations of indoor airflow velocity fields and pollutant distribution that arise from the ventilation layout, control actions, and occupancy patterns. As a result, such controllers fail to capture the localized effects of ventilation decisions. They may over-ventilate the entire space to maintain air quality at a single location, wasting energy, or overlook under-ventilated regions that compromise occupant health and comfort. This gap highlights the need for *spatiotemporal* indoor airflow modeling to accurately capture how ventilation control actions affect air quality—crucial for maintaining occupant health while enabling energy-efficient building HVAC operation.

To accurately model airflow dynamics and indoor air quality, Computational Fluid Dynamics (CFD) simulations are widely adopted in building ventilation research [17]. For example, Bianco et al. [18] employed CFD to assist in the design of ventilation units for buildings, while Gao et al. [19] used CFD to analyze airflow fields in an isothermal chamber under fixed ventilation rates. Bulinska et al. [20] modeled CO₂ distribution in a bedroom to inform optimal sensor placement, while Mou et al. [21] simulated airflow and CO₂ concentration in a seminar room, capturing complex 3D ventilation patterns. Similarly, Ning et al. [22] used CFD to evaluate the influence of supply outlet height on indoor air distribution in a bedroom. Extending to classroom environments, Mahmoud [23] applied CFD to analyze the dispersion of human-generated aerosols and CO₂, offering insights into ventilation design for indoor air quality.

Although CFD simulations can accurately model indoor airflow fields, solving the governing partial differential equations (PDEs) using numerical solvers, such as finite element and finite volume methods, is computationally expensive [24]. This high cost makes it impractical to embed CFD solvers directly into real-time control loops, where rapid decisions must be made in response to changing occupancy and environmental conditions. Consequently, most CFD-based studies [21–23] focus on offline analysis and system design rather than real-time control. Our prior work [25] attempted to bridge this gap using a differentiable PDE-based framework that optimizes control via adjoint methods. While

this approach enabled physically grounded optimization, it required repeatedly solving PDE-constrained optimization and was limited to a 2D setting with simplified geometry, highlighting a key scalability bottleneck.

To address these limitations, in this work, we propose a data-driven operator learning framework to approximate the input–output behavior of CFD simulations. Neural operators are designed to learn mappings between *infinite-dimensional* function spaces, enabling them to approximate the solution operator of PDEs across a family of initial and boundary conditions [26,27]. Previous studies [28,29] have used conventional neural networks to learn indoor air dynamics from CFD data. However, these models operate in finite-dimensional spaces and typically require fixed spatial grids, making them inflexible and often requiring retraining under new control or occupancy settings. In contrast, neural operators learn mappings between continuous functions, allowing predictions to be queried at arbitrary spatial locations within the room. This functional formulation captures how ventilation control actions shape the entire indoor airflow and air quality distribution, enabling flexible and efficient ventilation control across diverse operating conditions. Several neural operator architectures have demonstrated strong performance in modeling complex physical systems. For example, Deep Operator Networks (DeepONets) [27] and Fourier Neural Operators (FNOs) [26] enable efficient approximation of fluid dynamics with significantly reduced computational cost. These models have shown success in applications such as turbulent flow prediction [30], vehicle aerodynamics [31], and indoor airflow modeling [19]. More recently, General Neural Operator Transformers (GNOTs) [32] have extended these capabilities to irregular spatial domains and multiple input functions, making them well suited for modeling building airflow with complex geometries and diverse inputs (e.g., control actions, occupancy, and past room states).

1.1. Summary of contributions

In this work, we propose a data-driven operator learning framework for energy-efficient building ventilation control, validated in a real-world classroom using CFD simulations (Fig. 1). While prior studies [19,33] have applied neural operators for airflow prediction, our work advances this line in two key ways: (i) we propose an ensemble neural operator transformer model that enhances prediction accuracy and robustness, and (ii) we introduce the integration of operator learning into closed-loop control for building energy management. This is substantially more challenging, as performance must account for both predictive accuracy and reliable control under energy and air quality constraints.

We benchmark our approach against data-driven ventilation control methods using averaged models (e.g., MLPs) and deep learning-based reduced-order models (e.g., U-Nets) [34]. Unlike MLPs, neural operators capture full spatial airflow fields, and unlike reduced-order models, they

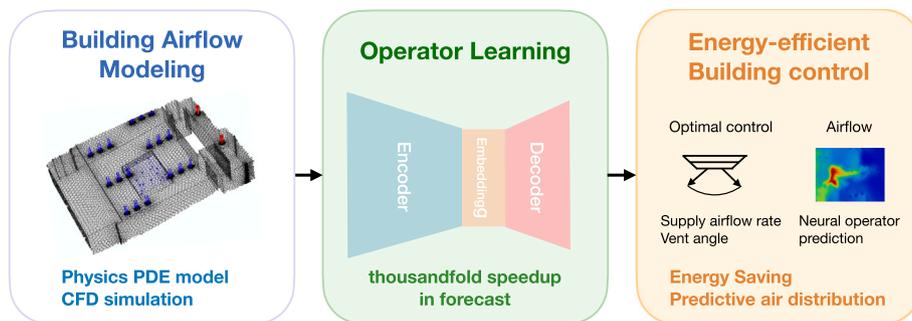


Fig. 1. Schematic of our data-driven operator learning framework for energy-efficient ventilation control. Computational fluid dynamics (CFD) simulations are used to model complex 3D airflow and CO₂ spatiotemporal dynamics. An ensemble of neural operator transformer models is trained to learn the mapping between ventilation control actions and airflow field evolution. Leveraging high-fidelity simulation data, our approach enables real-time optimization of ventilation strategies to minimize energy consumption while maintaining indoor air quality.

learn mappings between function spaces rather than pointwise values. This functional representation captures how control actions shape the entire airflow patterns and resulting indoor air quality. CFD-validated experiments demonstrate that these advantages enable neural operators to achieve more energy-efficient and healthier indoor environments compared to finite-dimensional alternatives. Our key contributions are summarized as follows:

- **Ensemble neural operator transformer:** We propose an ensemble neural operator transformer model to predict building airflow velocity and CO₂ fields under varying HVAC control actions and occupancy levels. This model improves prediction accuracy compared to individual neural operator models and achieves a remarkable speed-up of over 250,000× relative to CFD simulations, enabling accurate indoor air quality prediction with high computational efficiency.
- **Building control with operator learning:** We incorporate the learned neural operator into an optimization framework for energy-efficient ventilation control subject to air quality constraints. Compared to maximum airflow control, rule-based control, and data-driven control using averaged prediction or reduced-order models, our approach enables more effective closed-loop building ventilation control under varying occupancy, achieving lower energy use and significantly fewer CO₂ violations.
- **Open-source building CFD dataset:** We develop and release a high-fidelity, open-access building dataset derived from CFD simulations of a real-world classroom. Along with the dataset, we publish the underlying 3D room model and the files required to reproduce and extend the simulations. The dataset includes airflow velocity fields and CO₂ fields under diverse HVAC control actions and occupancy levels, providing a reproducible benchmark for training and evaluating machine learning models in building ventilation control applications.

2. Problem formulation

In this section, we introduce the governing dynamics of indoor airflow and CO₂ transport, and formulate the learning and control problems. The CFD simulation setup and dataset description are provided in Section 3.

2.1. Governing equations for CO₂ dynamics

Let $\Omega \subset \mathbb{R}^3$ be the spatial domain of interests, and let $t \in \mathbb{R}^+$ denote time. We define $C(\mathbf{x}, t)$ as the CO₂ concentration at spatial location $\mathbf{x} \in \Omega$ and time t . Let $m(t) = [m^r(t) \quad m^a(t)] \in \mathbb{R}^{12}$ be the control actions (airflow rates $m^r \in \mathbb{R}^6$ and airflow angles $m^a \in \mathbb{R}^6$) for the six groups of supply vents, and $n_p(t) \in \mathbb{R}$ denote the occupancy number. The distribution of CO₂ in an indoor environment follows the advection-diffusion equation [25,35],

$$\frac{\partial C(\mathbf{x}, t)}{\partial t} + \mathbf{u}(\mathbf{x}, t) \cdot \nabla C(\mathbf{x}, t) = D_{\text{eff}} \nabla^2 C(\mathbf{x}, t) + S(\mathbf{x}, t), \quad (1)$$

where

- $C(\mathbf{x}, t)$ is the CO₂ concentration (ppm) at location \mathbf{x} and time step t .
- $\mathbf{u}(\mathbf{x}, t)$ is the airflow velocity field obtained from CFD simulations, by solving the incompressible Navier-Stokes equations. The boundary conditions at the supply vents depend on the ventilation control $m(t)$, thereby allowing the control action to influence $\mathbf{u}(\mathbf{x}, t)$ throughout the domain. The numerical model for the airflow velocity field is presented in Appendix-A.
- D_{eff} is the diffusion coefficient for CO₂ in the air.
- $S(\mathbf{x}, t)$ represents the CO₂ source term, which models occupant-generated CO₂ on the designated occupancy surface. This source depends on the occupancy level $n_p(t)$ and we assume that the exhaled air rate is 6 L/min per person following [36].

2.2. Data-driven modeling with neural operator

Solving PDEs numerically is computationally expensive, making traditional CFD models impractical for real-time building control applications. To address this limitation, we aim to learn a neural operator \mathcal{G}_θ that efficiently maps historical CO₂ concentrations, control actions, and occupancy levels to future CO₂ concentration distributions. Formally, the neural operator learns the following mapping,

$$\mathcal{G}_\theta : (C(\mathbf{x}, \tau)_{\tau \in [t-H, t]}, m, n_p) \mapsto C(\mathbf{x}, \tau)_{\tau \in (t, t+T]} \quad (2)$$

where $C(\mathbf{x}, \tau)_{\tau \in [t-H, t]}$ is the historical CO₂ fields over the period $[t-H, t]$ and $C(\mathbf{x}, \tau)_{\tau \in (t, t+T]}$ is the predicted future CO₂ fields over the future interval $(t, t+T]$. Our forecasting approach incorporates historical CO₂ concentrations to account for temporal dependencies inherent to the system's physics (e.g., diffusion and advection dynamics). We assume that the control m and occupancy n_p remain *fixed* over $(t, t+T]$. In building control applications, the transient dynamics of airflow and CO₂ transport evolve rapidly over a short time horizon—typically within a few minutes. However, building control settings and occupancy levels generally remain constant during these intervals (e.g., a fixed HVAC setpoint over a 5-minute or 15-minute interval [6]). This assumption thus maintains consistency between the physical model and practical control timescales.

In practice, we rarely have continuous access to the CO₂ distribution over space and time. Instead, we rely on a discretized approximation of the underlying functions. For simplicity and consistency, we reuse the symbols t, H, T as discrete time indices. Let $\{x_i\}_{1 \leq i \leq N_x}$ be a spatial grid and $C_i^t \in \mathbb{R}$ be the CO₂ concentration at spatial grid point x_i and time step t . We collect the recent history of CO₂ fields over the last H time steps, along with the fixed parameters m and n_p . Define the discrete input set

$$\mathcal{A} = \{(x_i, C_i^{t-H:t})\}_{1 \leq i \leq N_x} \cup \{m\} \cup \{n_p\}, \quad (3)$$

although \mathcal{A} is finite-dimensional, it represents the sampled version of the underlying function $C(\mathbf{x}, \tau)$ over the history window $[t-H, t]$. From this discrete representation, the neural operator \mathcal{G}_θ produces predicted future concentrations $\hat{C}_{1 \leq i \leq N_x}^{t+1:t+T}$ (with short-hand notation \hat{C}), which approximate the continuous output function $\hat{C}(\mathbf{x}, \tau)$ for $\tau \in (t, t+T]$:

$$\hat{C}_{1 \leq i \leq N_x}^{t+1:t+T} = \mathcal{G}_\theta(\mathcal{A}). \quad (4)$$

2.3. Ventilation control optimization

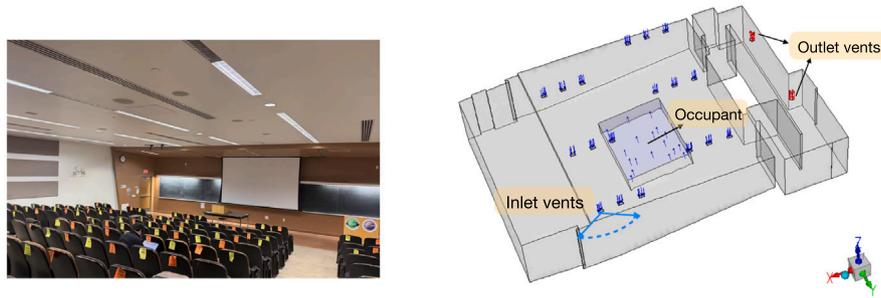
The learned neural operator \mathcal{G}_θ can be integrated into the building ventilation control optimization problem. The resulting optimization problem is formulated in (5).

$$\min_{m=[m^r, m^a]} w_1 \|\hat{C} - C_{\text{target}}\|_2^2 + w_2 \|m - m^{(0)}\|_2^2 + w_3 \|m^r\|_1, \quad (5a)$$

$$\text{s.t. } \hat{C} \leftarrow \mathcal{G}_\theta(\{(C_{1 \leq i \leq N_x}^{t-H:t})\} \cup \{m\} \cup \{n_p\}), \quad (5b)$$

$$\underline{m}^r \leq m^r \leq \overline{m}^r, \quad \underline{m}^a \leq m^a \leq \overline{m}^a. \quad (5c)$$

In the objective function (5a), the first term quantifies the deviation between predicted CO₂ concentrations \hat{C} and the desired CO₂ level C_{target} over a prediction horizon T . This term is commonly used in indoor air quality ventilation control studies, as seen in prior works [16,37]. The second term penalizes deviations of the optimized control actions from the previous control action $m^{(0)}$. This is important for real-world building management, where large deviations in control actions are typically avoided to maintain system stability and operational safety [25,38]. By encouraging control actions to remain close to $m^{(0)}$, we also reduce the risk of extrapolating beyond the model's training domain, thereby increasing the reliability of the predictions. The third term represents the energy consumption, measured through the L1 norm of the



(a) Picture of the studied room and geometrical model of the room.

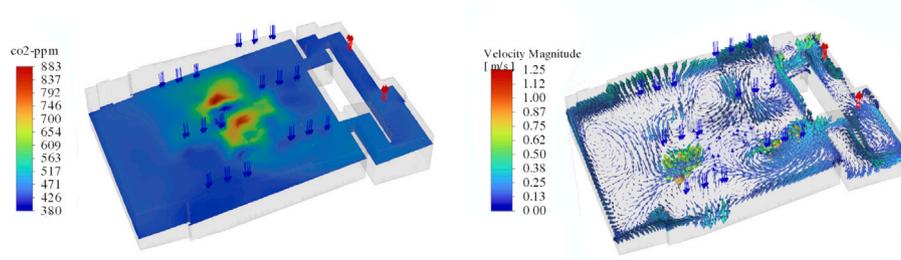
(b) Visualization of CO₂ field and airflow velocity field.

Fig. 2. (a) The picture of the studied room and the geometry of the CFD model: a classroom with a ventilation system including 18 inlet vents and 2 outlet vents on the ceiling. Occupants are modeled as a single rectangular cuboid with a prescribed CO₂ mass flux to represent CO₂ effects of varying occupancy levels. (b) Visualization of the CFD simulation results of CO₂ concentration and airflow velocity fields - one example from the developed CFD dataset.

ventilation rate [25]. The coefficients w_1, w_2, w_3 balance the relative importance of these objectives. Constraint (5b) describes that the predicted CO₂ concentrations are obtained via the trained neural operator model in Eqn (4), while constraint (5c) ensures that both mechanical ventilation rates m^r and vent angles m^a remain within their physical limits.

Remark 1. In this study, we do not explicitly model or control indoor temperature in the problem formulation. This choice is motivated by the fact that CO₂ concentration responds on a much faster time scale than temperature, which evolves more slowly because of the building's thermal inertia. By focusing on the faster dynamics-airflow and CO₂ transport, we are able to evaluate the proposed control strategy in a more responsive setting. We note that thermal effects are still captured, as our CFD simulations solve the energy Eq. (18) which models the transport of thermal energy within the airflow and includes occupant heat sources and boundary temperature conditions. Extending the formulation to include indoor temperature modeling and thermal control remains a valuable direction for future work, and our framework is compatible with such multi-objective settings.

3. The BEAR-CFD data

The CFD simulation is developed based on a real-world classroom located at the University of California, San Diego. The classroom is equipped with a ceiling-mounted ventilation system. The ventilation system includes 2 outlet vents and 18 inlet vents grouped into six zones, each allowing independent control of airflow rates and supply airflow angles to optimize energy efficiency while maintaining high indoor air quality. Fig. 2(a) illustrates the physical classroom and its corresponding 3D computer-aided design (CAD) model representation, and Fig. 2(b) presents an example of the CO₂ concentration and velocity field visualization. This section provides a detailed description of the simulation setup and the open-source dataset.

3.1. CFD simulation setup

3.1.1. Geometry

The simulated domain represents a mechanically ventilated classroom with dimensions of 19 m × 13 m × 3.5 m, see Fig. 2(a). The ventilation system consists of 18 rectangular inlet vents, each measuring 0.1349 m in width and 0.3048 m in height, and 2 ceiling-mounted outlet vents. Fresh air is supplied through the inlets at prescribed velocity and angle conditions, and exhausted through the outlets, forming a displacement ventilation flow pattern.

3.1.2. Occupant modeling

Occupants are collectively represented by a single, centrally positioned rectangular cuboid, visualized in Fig. 2(a, right). This abstraction aggregates the influence of multiple seated individuals into a compact volume that approximates their combined occupied zone [39,40].

CO₂ emissions from occupants are represented as a surface mass flux boundary condition and are modeled using the species transport equation [20,23]. To simulate varying occupancy levels (ranging from 10 to 80 people), we prescribe a total CO₂ mass flux of $n_p \times 0.00012$ kg/s on the cuboid surface, where n_p is the number of occupants. This corresponds to a constant breathing rate of 6 L/min per person, following the simplified exhalation model in [20]. A constant surface temperature is applied to the cuboid.

3.1.3. Mesh

The fluid domain is discretized into approximately 0.245 million tetrahedral elements to resolve airflow and CO₂ transport dynamics. Local mesh refinement is applied near critical regions, including the occupant cuboid, inlet and outlet vents, and wall boundaries, to capture steep gradients in velocity and scalar fields. Mesh quality is assessed using standard metrics: the minimum orthogonal quality is 0.172, and the maximum aspect ratio is 43.33, indicating an acceptable mesh for transient indoor airflow simulations.

Table 1
Boundary conditions.

Boundary surface	Boundary conditions
Inlet vents	velocity inlet, $C_{\text{inlet}} = 400$ ppm, $T_{\text{inlet}} = 20$ °C, $m^r \sim U[0.324, 3.24]$ m/s, $m^a \sim U[45, 135]^\circ$
Outlet vents	Pressure outlet
Walls	No-slip and adiabatic walls, $T_{\text{wall}} = 21$ °C
Occupant surface	mass flow inlet, $C_{\text{occupant}} = 40000$ ppm, $T_{\text{occupant}} = 25$ °C [20]

3.1.4. Numerical models and boundary conditions

Airflow and CO₂ dynamics within the classroom are simulated using the commercial CFD software ANSYS Fluent [41]. The incompressible Navier–Stokes equations are solved to capture airflow behavior, coupled with species transport equations to model the distribution of CO₂, O₂, H₂O, and N₂. Turbulence is modeled using the $k-\omega$ SST (Shear Stress Transport) model, which is well-suited for predicting indoor near-wall flow behavior [42]. The governing equations are discretized using the finite volume method, with second-order schemes for both momentum and species transport. The energy equation is also activated to solve for temperature.

The boundary surfaces include walls, inlet and outlet vents, and the occupant surface. A summary of boundary conditions is provided in Table 1. Inlets are modeled as velocity boundaries with prescribed temperature (T_{inlet}), CO₂ concentration (C_{inlet}), velocity magnitude (m^r), and velocity angle (m^a) [20,23]. To introduce airflow variability, the inlet velocities and angles for vent groups are randomly sampled from uniform distributions. A maximum velocity of 3.24 m/s corresponds to 10 air changes per hour (ACH), based on operational data. Outlets are modeled as pressure boundaries, and all walls are treated as no-slip, adiabatic surfaces with fixed temperature.

Remark 2. The CFD simulations used in this work follow standard modeling practices for indoor airflow [20,40,43], and are designed to provide consistent and reproducible training data for future research. The simulation setup assumes constant air density (incompressible flow), omitting the buoyancy effects. This simplification is reasonable in mechanically ventilated classrooms where forced ventilation dominates airflow. Nevertheless, we acknowledge that in densely occupied spaces, buoyancy forces may contribute to airflow patterns. Extending the CFD setup to incorporate buoyancy remains an important direction for future work, and our framework is readily adaptable to such enhanced simulations as well as to other CFD configurations and building control systems.

3.2. Bear-CFD dataset

Bear-CFD dataset is generated through a structured CFD simulation workflow designed to capture the spatiotemporal dynamics of indoor CO₂ levels under varying ventilation and occupancy conditions. The dataset is publicly available at <https://ucsdsmartbuilding.github.io/CFD-DATA.html>.

The simulations include both steady-state and transient flow cases. Specifically, we run 10 steady-state simulations under different boundary conditions. These steady-state results are then used as *initial conditions* for 300 transient simulations, each lasting 30 min of physical time and resolved at 10-second time steps. This initialization strategy ensures that each transient simulation begins from a realistic state shaped by prior control conditions. Without this step, starting from uniform or arbitrary states (e.g., same CO₂ everywhere) could introduce artificial transients unrelated to the actual control inputs. Simulation outputs are saved every 30 s, resulting in 60 time steps per transient case. Transient simulations capture the temporal evolution of airflow dynamics, reflecting unsteady effects influenced by varying boundary conditions. We leverage this transient data to train our neural operators.

To generate the dataset, for each simulation, key control parameters, including supply airflow rates (m_i^r) and airflow angles (m_i^a) for the i -th

Table 2
Data fields in the BEAR-CFD dataset pickle (.pkl) format.

Field	Description
HVAC surface	ndarray ($N_{\text{hvac}}, 3$), spatial coordinates of grid points on HVAC surfaces
CO ₂ -HVAC	ndarray (N_{hvac}, \bar{T}), CO ₂ concentration time series at HVAC surface
People surface	ndarray ($N_{\text{people}}, 3$), spatial coordinates of grid points on people surfaces
CO ₂ -People	ndarray ($N_{\text{people}}, \bar{T}$), CO ₂ concentration time series at people surface
steady case	int, Identifier for the initial steady-state condition used in the simulation
n_p	int, number of occupant
m_i^r	float, airflow rate (m/s) for i -th group of vents
m_i^a	float, angle (°) for i -th group of vents

group of inlet vents ($i = 1, \dots, 6$), as well as the occupant count (n_p)—are independently sampled from uniform distributions:

$$\begin{aligned} m_i^r &\sim U[\underline{m}^r, \overline{m}^r], m_i^a \sim U[45^\circ, 135^\circ], i \in [1, \dots, 6], \\ n_p &\sim U[10, 80], \end{aligned} \quad (6)$$

where m_i^r is the airflow rate for the i -th group of vents, bounded between $\underline{m}^r = 0.324$ m/s (10 % of maximum) and $\overline{m}^r = 3.24$ m/s; m_i^a is the airflow angle of the i -th group of vents, spanning 45° to 135°; and n_p is the number of occupants in the classroom. Each transient simulation follows this two-phase procedure: (1) Steady-state initialization: A random initial condition is selected from the 10 steady-state simulations. (2) Transient simulation: Control parameters (m_i^r , m_i^a) and occupancy n_p are sampled. The simulation runs for 30 min, with CO₂ and airflow fields recorded at 30-second intervals over $\bar{T} = 60$ time steps. The CO₂ concentrations were monitored at two critical planes:

- HVAC surface: A horizontal plane at 2.9-meter height near the ventilation inlets, capturing the supply and returning air quality.
- People surface: A horizontal plane at 1.6-meter height (average breathing height for standing adults), representing air quality at occupant exposure levels [44].

In our CFD simulation, the sitting surface is near the occupancy boundary, leading to potential inaccuracies from numerical artifacts and boundary effects. Thus, we focused on heights where airflow and CO₂ dispersion are more reliably captured.

The dataset is distributed in both ANSYS FLUENT’s native format (.cas and .dat files) and Python pickle (.pkl) format. The FLUENT files preserve the complete simulation environment and solution data, while the pickle format enables efficient programmatic access to the numerical results through standard Python libraries. The dataset includes spatiotemporal fields of CO₂ concentrations, airflow rates, vent angles, and occupancy, as detailed in Table 2.

4. Methodology

In this work, we propose a data-driven operator learning framework to model indoor air quality and optimize ventilation control for energy efficiency. The core component is an ensemble neural operator transformer architecture, \mathcal{G}_θ , shown in Fig. 3. In Fig. 3(a), the neural operator architecture enables fast building CFD simulations by processing multiple inputs: query points, supply airflow rates, airflow angles, and historical CO₂ concentrations. The trained operator learning model uses these inputs to predict future CO₂, which are then utilized to determine optimal ventilation control parameters by solving the optimization problem defined in Eq. (5). In the following subsections, we will first explain the concept of operator neural transformer, then describe the control algorithm.

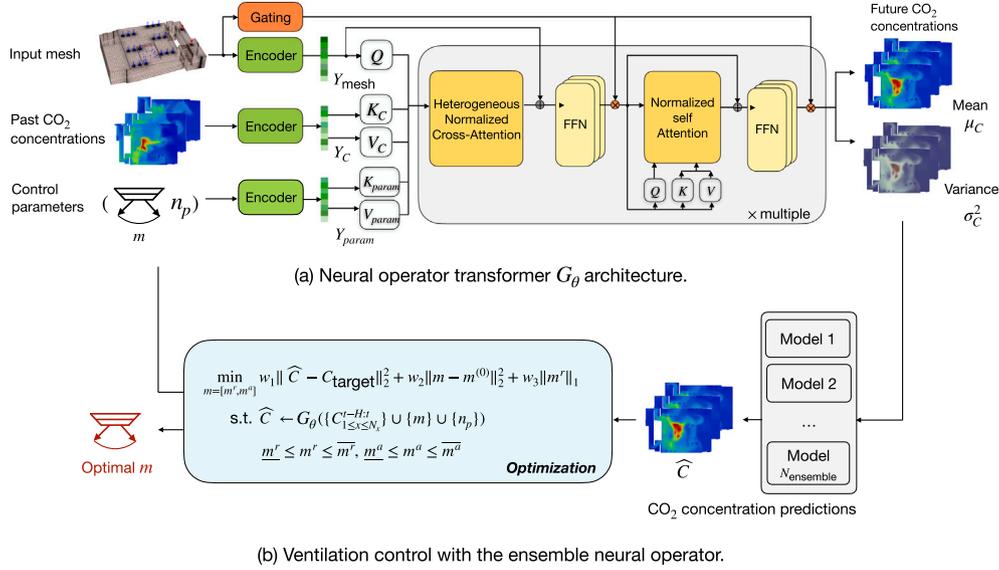


Fig. 3. Overview of the proposed data-driven operator learning framework for energy-efficient ventilation control. The framework consists of two phases: (1) the learning phase, where neural operator transformers are trained to map past air field data and ventilation control parameters to future air field evolution, and (2) the control phase, where the trained ensemble neural operator is integrated into an optimization framework to solve the ventilation control problem. This approach enables real-time optimization of airflow supply rates and vent angles while maintaining air quality standards.

4.1. Ensemble neural operator transformer

Existing operator learning approaches, while effective in many applications, often struggle with limited training data. To address this limitation, we enhance the General Neural Operator Transformer (GNOT) [32] by ensemble learning [45]. The network architecture of the proposed model together with the control is illustrated in Fig. 3. In the following subsections, we will describe the input encoding and ensemble learning for GNOT.

4.1.1. Input encoding

The model takes the input mesh, historical CO₂ concentrations, and control parameters (including ventilation control actions and the number of occupants), as inputs. To accommodate these heterogeneous inputs, a general encoder, highlighted in green in Fig. 3, is employed to transform them into the feature embedding $Y \in \mathbb{R}^{N \times n_e}$, where N denotes an arbitrary number of input elements and n_e is the embedding dimension. The model employs simple multilayer perceptrons (MLPs), denoted as f_{w1}, f_{w2}, f_{w3} , to map each type of input to its corresponding embedding.

- **Input mesh:** A MLP maps the mesh points to query embeddings $Y_{\text{mesh}} = (f_{w1}(x_i))_{1 \leq i \leq N_x} \in \mathbb{R}^{N_x \times n_e}$.
- **Past CO₂ concentrations:** At time t , we have $\{x_i, c_i\}_{1 \leq i \leq N_x}$, where $c_i = C_i^{t-H:t}$ represents the historical CO₂ levels at position x_i . A MLP encodes both positions and concentrations to produce the feature $Y_C = (f_{w2}(x_i, c_i))_{1 \leq i \leq N_x} \in \mathbb{R}^{N_x \times n_e}$.
- **Control parameters:** A MLP encodes control parameters $[m, n_p] \in \mathbb{R}^{13}$ into embeddings $Y_{\text{param}} = f_{w3}([m, n_p]) \in \mathbb{R}^{1 \times n_e}$.

4.1.2. Ensemble Learning of GNOT

As shown in Fig. 3(a), GNOT begins by encoding input features and updating them using a heterogeneous normalized cross-attention layer, followed by a normalized self-attention layer to refine representations. To effectively capture spatial heterogeneity, GNOT incorporates a geometric gating mechanism that leverages the query point coordinates to compute a weighted combination of expert feed-forward networks (FFNs). The model stacks N such attention blocks to produce the final predictions.

While GNOT is originally designed to output only the mean of the prediction, we enhance its robustness by introducing an ensemble-based extension. Instead of training GNOT to minimize mean squared error or relative error, we modify it to predict a probability distribution for future CO₂ concentrations, characterized by the mean (μ_C) and variance (σ_C^2). Specifically, the model predicts:

$$\mu_C, \sigma_C^2 = G_\theta(A), \quad (7a)$$

$$\mu_C = \mathbb{E}[\hat{C}_{1 \leq i \leq N_x}^{t+1:t+T}], \quad \sigma_C^2 = \text{Var}[\hat{C}_{1 \leq i \leq N_x}^{t+1:t+T}], \quad (7b)$$

where $[\mu_C]_i^t, [\sigma_C^2]_i^t$ represent the mean and variance at location i and time step t . The model is trained using the Negative Log-Likelihood (NLL) loss, defined as:

$$\mathcal{L} = \frac{1}{N_x T} \sum_{i=1}^{N_x} \sum_{k=1}^T \left(\frac{\log(2\pi[\sigma_C^2]_i^{t+k})}{2} + \frac{(C_i^{t+k} - [\mu_C]_i^{t+k})^2}{2[\sigma_C^2]_i^{t+k}} \right) \quad (8)$$

where $C_i^{t+k} \in \mathbb{R}$ is the true value at x_i and time $t+k$. By minimizing the NLL loss, the model learns to jointly optimize the mean and variance predictions, effectively mitigating overfitting [45].

We train an ensemble of neural operator transformers, with the final prediction obtained by averaging outputs from multiple independently trained models. Let $\mu_C^{(n)}$ denote the mean prediction of the n -th model. The ensemble prediction is then computed as,

$$\mu_C^{\text{ensemble}} = \frac{1}{N_{\text{ensemble}}} \sum_{n=1}^{N_{\text{ensemble}}} \mu_C^{(n)}, \quad (9a)$$

$$\mu_C^{(n)}, (\sigma_C^2)^{(n)} = G_{\theta_n}(A), \quad (9b)$$

where N_{ensemble} represents the number of models in the ensemble, θ_n is the model parameters for the n -th trained neural operator. This ensemble approach not only improves prediction accuracy but also enhances the reliability of ventilation control, as demonstrated in our experiments.

4.2. Control algorithm

With the learned ensemble neural operator transformer model, we are ready to solve the building control problem in (5). Recall that the

Algorithm 1 Algorithm for solving (5).**Require:** Neural operator transformers \mathcal{G}_θ ,1: Control inputs $C_{1 \leq i \leq N_x}^{t-H:t}$, n_p **Ensure:** $m^{(0)}$ \triangleright initial control actions2: **for** $ite = 0, 1, \dots, \text{MaxIte}$ **do**3: **obtain** future CO₂ predictions \hat{C} (9)(11)4: **evaluate** loss $\mathcal{L}(m)$ (10)5: **compute** the gradient $\nabla \mathcal{L}(m)$ 6: **update** the control vector with step size η :

$$m^{(ite+1)} \leftarrow m^{(ite)} - \eta \nabla \mathcal{L}(m)$$

7: **project** the updated m to satisfy the box constraints (5c)8: **end for**9: **Return** m

objective function of the building control problem is defined as:

$$\mathcal{L}(m) = w_1 \|\hat{C} - C_{\text{target}}\|_2^2 + w_2 \|m - m^{(0)}\|_2^2 + w_3 \|m\|_1 \quad (10)$$

where the first term $\|\hat{C} - C_{\text{target}}\|_2^2$ is defined as,

$$\|\hat{C} - C_{\text{target}}\|_2^2 = \frac{1}{N_x T} \sum_{k=1}^T \sum_{i=1}^{N_x} (\hat{C}_i^{t+k} - C_{\text{target}_i}^{t+k})^2$$

and w_1, w_2, w_3 are weighting coefficients. \hat{C} represents the predicted CO₂ concentrations generated by the ensemble neural operator (9):

$$\hat{C} = \mu_C^{\text{ensemble}}. \quad (11)$$

To solve the building control problem in (5), we leverage the differentiability of the neural operator \mathcal{G}_θ and propose a gradient-based method to update the control actions for ventilation. To ensure that the control vector m remains within feasible bounds (5c), we use a projected gradient descent method. Specifically, after computing the gradient of the objective function with respect to m , the control vector is updated and then clipped to satisfy the predefined bounds. Our optimization procedure is summarized in Algorithm 1.

5. Numerical experiments

In this section, we evaluate the performance of the proposed framework through two experiments: (1) learning experiments to assess the accuracy of the ensemble neural operator, and (2) control experiments to evaluate the effectiveness of our data-driven ventilation control framework. The source code, input data, and trained models from all experiments are available on GitHub.¹ All experiments are conducted on NVIDIA GeForce RTX 2080 Ti GPUs.

5.1. Learning results

We train our ensemble neural operator transformer to predict CO₂ levels on the people surface. The number of query points is $N_x = 7462$, and we select $F = 12$ and $T = 6$, meaning that the model utilizes data from the past 12 time steps (equivalent to 6 min) to forecast CO₂ levels over the next 6 time steps (3 min). The dataset is divided into an 80 % training set and a 20 % testing set to evaluate model performance. We train $N_{\text{ensemble}} = 5$ independent models and compute the mean of their predictions as the final output. We use the AdamW optimizer with a cyclical learning rate schedule, and each model is trained for 200 epochs.

To evaluate model performance, we use the average l_2 relative error as the primary metric. Let $C^{(d)}, \widehat{C}^{(d)} \in \mathbb{R}^{N_x \times T}$ denote the ground

truth and the predicted *mean* future CO₂ concentrations for the d -th sample, respectively, and let D represent the total dataset size. The error is defined as:

$$l_2 = \frac{1}{D} \sum_{d=1}^D \frac{\|\widehat{C}^{(d)} - C^{(d)}\|_2}{\|C^{(d)}\|_2}. \quad (12)$$

Fig. 4 shows the training convergence and test performance of the five neural operator transformers over 200 epochs. The left plot presents the NLL loss (8), while the right shows the l_2 relative error on both training and test sets. Table 3 lists the l_2 error metrics for the five trained models (Model 1 to Model 5) and the ensemble model, evaluated on both the training and testing datasets. Notably, the ensemble model achieves the lowest errors, with a training l_2 error of 5.9 % and a testing l_2 error of 10.90 %. The ensemble model consistently outperforms the individual models, leading to more robust and accurate predictions. The performance improvement can be attributed to the ensemble model's ability to mitigate potential overfitting of individual models by averaging out their prediction errors.

Fig. 5 shows the ground truth CO₂ concentration fields with predictions from the ensemble model and Model 3 (best-performing individual model), along with their relative errors for three test cases. Overall, the neural operator framework demonstrates strong capability in predicting complex spatial CO₂ distributions under varying control parameters. The ensemble model achieves lower relative errors and more accurately captures spatial patterns compared to Model 3. In Section 5.4, we further demonstrate its effectiveness in activating control actions and optimizing air quality.

5.2. Ventilation control results

We now evaluate the control performance under three occupancy scenarios: high ($n_p = 75$), medium ($n_p = 45$), and low ($n_p = 15$). According to ASHRAE Standard 62.1 [46], classrooms typically require 4–6 air changes per hour (ACH) to ensure adequate ventilation and indoor air quality. For all scenarios in this study, the initial control actions are initialized at 5 ACH (within the recommended range) for supply vents, with inlet angles fixed at a 90° downward orientation to align with conventional HVAC configurations. We evaluate the following control strategies:

- **Max Control:** The airflow rate is set to its maximum value, with the inlet angle fixed at 90° downward.
- **Baseline Control:** The control actions are determined based on the ASHRAE standard as 5 ACH with 90° downward angle, serving as a baseline for comparison.
- **Rule-based Control:** The airflow rate is set proportionally to the occupancy level, calculated as the number of occupants divided by the maximum occupancy, with 90° downward angle.
- **Data-driven (DL-Avg):** We employ a neural network to predict the average CO₂ concentration at the center of the occupancy surface. Ventilation control actions are optimized based on (5) using the prediction.
- **Data-driven (DL-ROM):** We implement a state-of-the-art Deep Learning-based Reduced-Order Modeling approach (DL-ROM) [34] to approximate the spatiotemporal CO₂ distribution. The resulting surrogate model is integrated into our ventilation control optimization framework (5) to optimize control actions.
- **Ours:** We use the proposed ensemble neural operator transformer to model the spatiotemporal dynamics of CO₂ concentrations. The learned neural operator is then used within the same ventilation control optimization framework in (5) to optimize control actions.

For our optimization solver, we choose $C_{\text{target}} = 400$ (ppm) to ensure that the CO₂ concentration loss is more broadly activated across the spatial domain, enabling more effective optimization with respect to the control actions. We choose $w_1 = 1, w_2 = 0.4$ and $w_3 = 0.12$ through

¹ <https://github.com/alwaysbyx/BuildingControlCFD>

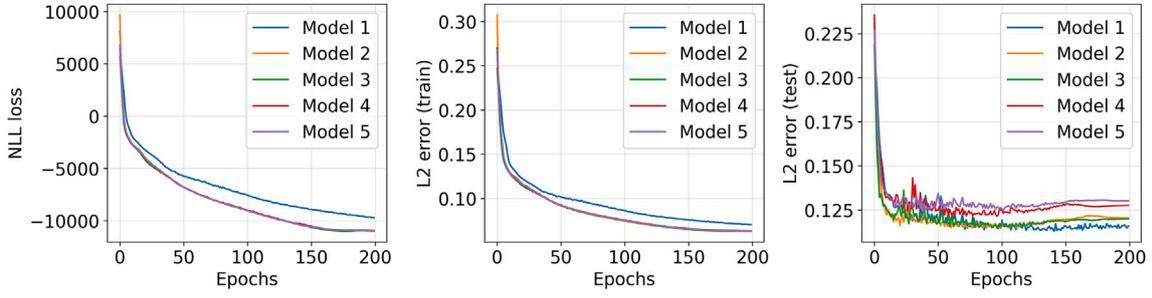


Fig. 4. Training loss (NLL loss) during training (left) and the l_2 error for the training (middle) and test sets (right). The l_2 error (12) is computed based on the ground truth and the model's mean prediction output.

Table 3

The l_2 error for five independently trained neural operator transformer models (Model 1 to Model 5) and their ensemble.

	Model 1	Model 2	Model 3	Model 4	Model 5	Ensemble
l_2 error (Train)	6.35 %	6.33 %	6.33 %	6.35 %	6.33 %	5.9 %
l_2 error (Test)	12.09 %	11.83 %	11.82 %	12.74 %	13.01 %	10.90 %

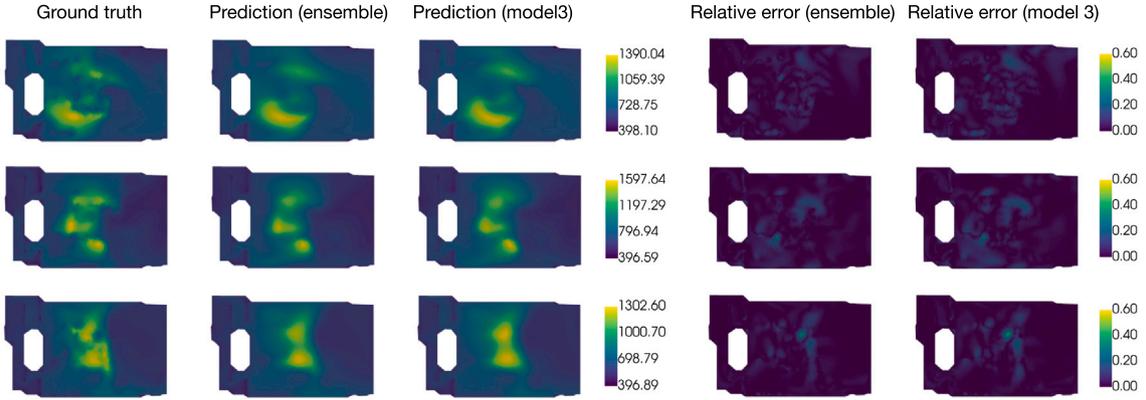


Fig. 5. Operator Learning: Visualization of the ground truth, corresponding predictions from the ensemble model and Model 3, and the relative errors between the ground truth and predictions at the final time step.

empirical tuning to balance air quality, energy consumption, and control smoothness. These weights are used consistently across all control experiments.

We validate the three control strategies using CFD simulations. Performance is measured using three metrics: mean CO_2 (ppm), peak CO_2 (ppm), and CO_2 violation (%), defined as $\frac{\text{Peak CO}_2 - 1200}{1200}$. As suggested by [47], 1200 ppm is the maximum acceptable CO_2 level for human health. For each metric, we record both the temporal average over the simulation period and the value at the final time step. Energy consumption is modeled as the percentage of the maximum power required to operate the ventilation system at its maximum airflow rate, providing a *normalized* measure of energy usage relative to the system's peak capacity:

$$E(m) = \frac{1}{6} \sum_{i=1}^6 m_i^r / \overline{m^r} \times 100 \%, \quad (13)$$

where m_i^r represents the ventilation rate for the i -th vent, and $\overline{m^r} = 3.24$ (m/s) is the maximum ventilation rate.

The results are summarized in Table 4. The experimental results demonstrate the effectiveness of the proposed control strategy (Ours) in balancing air quality and energy consumption in the first two cases. In Case 3, under the low-occupancy scenario, the optimization favors main-

taining higher air quality, leading to slightly higher energy consumption than the baseline control. Compared to the Max control strategy, our method achieves comparable air quality performance while reducing energy usage by 34–56 %. Compared to the baseline control, our approach effectively lowers CO_2 levels and adheres to air quality constraints with slight energy increase. Rule-based control suffers from CO_2 violations (e.g., 35.2 % final-step violation in Case 3) due to its reliance on well-chosen static operation schedules. Unlike rule-based systems, which require manual fine-tuning of thresholds for different room layouts and occupancy levels, our method autonomously adapts to operating conditions and reduces energy consumption by 12–28 % in Cases 1–2 while maintaining safe CO_2 levels.

In addition, our method outperforms data-driven control (DL-Avg) and (DL-ROM) by maintaining good energy efficiency without incurring any violations of the air quality constraints. While the DL-Avg approach can optimize ventilation locally, it has two key limitations: (1) it primarily predicts average CO_2 concentrations rather than spatial variations, and (2) it cannot effectively map control actions to real-time CO_2 distributions across different zones. While the DL-ROM approach captures spatial-temporal airflow dynamics, it is less effective at modeling the relationship between control inputs and the resulting air distributions compared to our method. For completeness, we present the learning results of DL-ROM in Appendix B.

Table 4

Comparison of CO₂ metrics (mean, peak, and violation percentages) and energy consumption across different control strategies. “Average” refers to the temporal mean over the simulation period, while “Final Step” refers to the CO₂ level at the end of the simulation. Our approach achieves significant ventilation energy savings compared to the other control approaches while maintaining acceptable CO₂ violation levels.

Case	Control Strategy	Mean CO ₂ (ppm)		Peak CO ₂ (ppm)		CO ₂ > 1200 ppm (%)		Energy Consumption (%)
		Average	Final Step	Average	Final Step	Average	Final Step	
1	Max Control	565.6	534.0	895.7	854.2	0.00	0.00	100.0
	Baseline Control	616.1	612.3	1203.0	1143.6	1.32	0.00	50.0
	Rule-based Control	567.4	533.5	898.1	807.9	0.00	0.00	93.8
	Data-driven(DL-Avg)	604.4	595.2	1021.6	971.6	0.00	0.00	83.3
	Data-driven(DL-ROM)	626.2	635.8	1300.1	1326.7	8.34	10.56	44.6
	Ours	600.7	587.9	1060.2	1004.0	0.00	0.00	65.8
2	Max Control	546.7	512.2	800.0	741.3	0.00	0.00	100.0
	Baseline Control	599.9	595.9	1097.3	1108.4	0.00	0.00	50.0
	Rule-based Control	593.7	584.6	1057.6	1059.3	0.00	0.00	56.2
	Data-driven(DL-Avg)	584.7	562.1	912.7	914.6	0.00	0.00	83.0
	Data-driven(DL-ROM)	602.7	600.1	1062.2	1066.2	0.00	0.00	50.8
	Ours	605.0	600.0	1044.1	881.4	0.00	0.00	43.8
3	Max Control	532.3	497.6	754.0	692.0	0.00	0.00	100.0
	Baseline Control	585.2	578.0	1034.2	1013.9	0.00	0.00	50.0
	Rule-based Control	650.7	712.7	1361.4	1623.1	15.67	35.26	18.8
	Data-driven(DL-Avg)	576.9	566.0	946.1	978.3	0.00	0.00	66.1
	Data-driven(DL-ROM)	590.3	590.2	1121.6	1208.4	0.13	0.70	49.4
	Ours	589.4	585.4	984.4	949.8	0.00	0.00	55.2

5.3. Multi-step ventilation control results

To assess the proposed operator learning model in realistic control scenarios, we extend the single-step optimization in Section 5.2 to a multi-step planning horizon. Specifically, a 30-minute ventilation horizon is considered with 10 planning steps, each optimizing control over the next 6 time steps (3 min), yielding 60 control steps in total. The initial step ($t = 0$) is initialized using historical CO₂ data from Scenario 1 in Section 5.2, while subsequent steps use the predicted CO₂ levels by recursively rolling out the learned operator model with previously optimized controls.

Fig. 6 illustrates the performance of the proposed control strategy in comparison to other control methods. The occupancy profile (top) shows significant variations over time. Our strategy dynamically adjusts ventilation in response to occupancy levels. At the beginning, when occupancy is high, the controller increases airflow to mitigate CO₂ accumulation. Although occupancy drops shortly after, the controller maintains a moderately high ventilation rate to clear the residual CO₂ from the earlier crowded period. In contrast, during later low-occupancy intervals, lower accumulated CO₂ allows the system to reduce airflow further, resulting in greater energy savings.

Over the control period, total energy consumption for Max, Baseline, Rule-based, Data-driven (DL-Avg), Data-driven (DL-ROM), and our method is 100 %, 50 %, 63.1 %, 55.3 %, 57.6 %, and 50.6 %, respectively. In terms of CO₂ violations, Max control yields zero violations; Baseline, Rule-based, and DL-ROM incur 17, 10, and 7 violations; DL-Avg results in 37 violations and ours incurs only 1 violation. The relatively poor performance of DL-Avg stems from its reliance on predicting only the average CO₂ concentration, without capturing the dynamic spatial shifts of peak concentrations over time. As a result, the control decisions fail to address localized high-risk areas, leading to both higher energy consumption and more CO₂ violations.

In addition, we plot the time series of control actions for the six groups of vents, under our approach in Fig. 7. Notably, the airflow rates for vent groups 2 and 6 exhibit a strong correlation with occupancy, increasing during higher occupancy periods to enhance ventilation. Additionally, the controller consistently assigns the highest airflow to vent group 1, while groups 3, 4, and 5 maintain relatively low flow rates. This likely reflects spatial differences in ventilation effectiveness, with the optimization learns to prioritize vents that more effectively contribute to CO₂ removal.

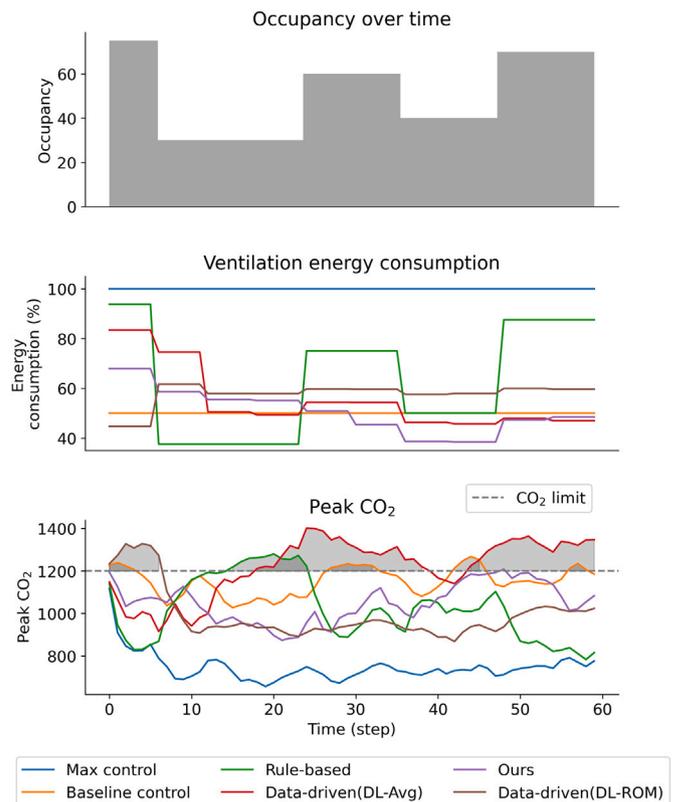


Fig. 6. Comparison of occupancy profile, energy consumption, and peak CO₂ concentration over a 60-step period. The top panel illustrates dynamic occupancy changes, the middle panel shows corresponding energy consumption as a percentage of the maximum, and the bottom panel depicts CO₂ concentration levels with the threshold limit (1200 ppm) indicated by a dashed line.

5.4. Effectiveness of the ensemble model

In this subsection, we demonstrate the effectiveness of the ensemble model in control compared to individual models, serving as an ablation study. As shown in Section 5.1, the ensemble model achieves the lowest prediction error among all models. Building on these results, we

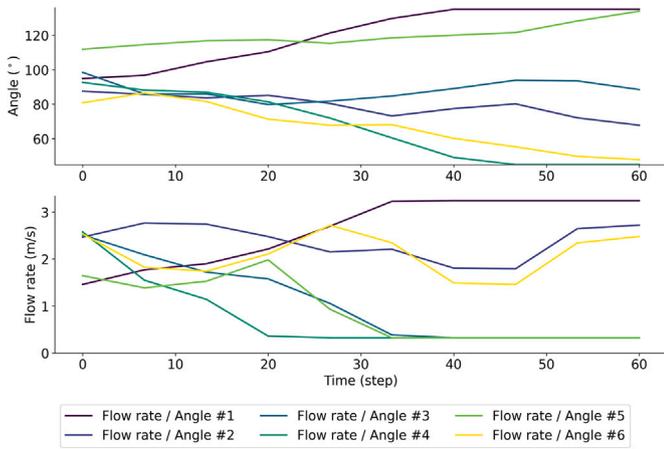


Fig. 7. Time series of control actions of our approach. Top plot shows the airflow angles and bottom shows airflow rates for 6 group of inlet vents under dynamic occupancy levels.

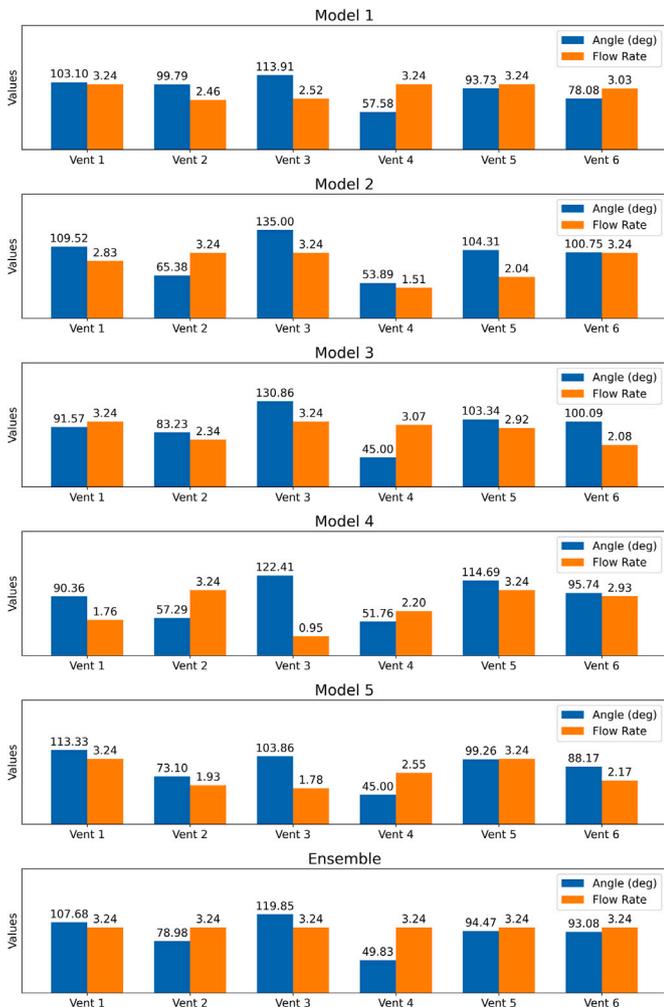


Fig. 8. Control actions (flow rates and angles) for individual models (Model 1 to Model 5) and the ensemble model across six inlet vents for one case. The ensemble model consistently achieves maximum airflow rates for all inlet vents.

further illustrate that the ensemble model, which aggregates predictions from multiple independently trained neural operators, consistently outperforms individual models in the control stage.

Table 5
Runtime of the operator learning pipeline.

Stage	Time Cost
Learning	
CFD data generation	3.5 CPU hours per simulation, totaling 1100 CPU hours
Neural operator training	16 GPU hours
Control	
CFD simulation (Sec 5.2)	1253.7 s
CFD simulation (Sec 5.3)	3.5 hours
Inference (ours)	<0.005 s

To gain more insights from the optimization results, we set the coefficient for energy consumption $w_3 = 0$ in the building control problem (5). This allows the control strategy to prioritize air quality without considering energy costs. We optimize the control actions for a case in which the control parameters are randomly selected, using both the individual models and the ensemble model. The resulting control actions are illustrated in Fig. 8. We observe that the ensemble model consistently reaches maximum airflow rates for all inlet vents, which align with the optimal action in this test case as the energy cost coefficient $w_3 = 0$. In contrast, the individual models exhibit variability in their control actions, with some models failing to achieve the maximum airflow rates across all vents.

5.5. Runtime analysis

To evaluate the computational efficiency of our proposed operator learning for building ventilation control, we analyze the runtime associated with each stage of the algorithmic pipeline, including the cost of generating CFD data, training time of the neural operator models, and performing inference during control deployment.

Table 5 summarizes the runtime of each component. The CFD simulations are conducted using ANSYS Fluent in parallel 6 cores, while neural operator training was performed using 2 RTX 2080 Ti GPUs. We observe that CFD simulation takes 1,253.7 seconds to compute the transient flow over six time steps (3 min), whereas the neural operator transformer completes the same task in just 0.005 s. This represents a remarkable speed-up of approximately 250,000 times compared to the CFD simulation. Although the learning phase requires moderate computational resources (the training of the five transformer models takes 16 GPU hours in total), the resulting neural operator enables rapid inference and real-time building control with the high-fidelity indoor fluid dynamics models.

6. Conclusion and future work

In this work, we propose a novel operator learning framework for energy-efficient ventilation control. Our approach involves training an ensemble of neural operator transformer models to learn the mapping from past CO₂ fields, occupancy levels, and control actions to future CO₂ fields. The ensemble model demonstrates superior predictive performance compared to individual models. We further integrate the learned neural operator into building ventilation optimization to optimize control actions. Using CFD simulations, we validate the proposed approach achieves substantial energy savings compared to maximum airflow control, rule-based control, and data-driven methods based on average CO₂ predictions and reduced-order modeling. In addition, compared to baseline control, our method maintains similar energy consumption while significantly reducing CO₂ violations. We open-source the CFD data to facilitate further research in developing machine learning models for building ventilation control with PDE-based models.

Promising future directions include extending the framework to more complex building environments, such as multi-zone systems. In addition, real-world experiments with integrated sensing and actuation systems are planned to validate the framework’s practical performance.

CRediT authorship contribution statement

Yuxin Bian: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation. **Oliver Schmidt:** Writing – review & editing, Software, Methodology, Investigation. **Yuanyuan Shi:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors gratefully acknowledge Ling Zhong for the framework visualization design. This work is supported by a Schmidt Sciences AI2050 Early Career Fellowship, NSF grant ECCS-2442689, and DOE grant DE-SC0025495.

Appendix A. Airflow dynamics modeling

Indoor air is modeled as an incompressible fluid consisting of four species, namely: oxygen (O₂), carbon dioxide (CO₂), water steam (H₂O) and nitrogen (N₂) [35], where each gas species (O₂, CO₂, H₂O, N₂) is assigned a constant density. The numerical model is based on the Navier-Stokes equations for incompressible flow, incorporating continuity, momentum, energy conservation, and turbulence model equations, along with species transport equations.

The general form of the continuity equation is:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \quad (14)$$

where ρ is fluid density and \mathbf{u} is the velocity field. In our simulations, ANSYS Fluent solves this equation under the incompressible flow assumption by setting the density ρ as constant.

The conservation equations for air constituents govern the transport and distribution of individual gas species within the airflow.

$$\frac{\partial Y_i}{\partial t} + \nabla \cdot (Y_i \mathbf{u}) = -\nabla \cdot \frac{1}{\rho} \mathbf{j}_i, \quad (15)$$

where i denotes three air constituents, namely, O₂, CO₂ and H₂O, Y_i is the mass fraction of the i -th air constituent, and ρ is density of fluid. The mass flux of the i -th constituent can be calculated $\mathbf{j}_i = -D_{\text{eff}} \nabla Y_i$, where D_{eff} is the effective diffusion coefficient which includes turbulence effects. The mass fraction of N₂ was calculated from the sum of mass fractions of all air species which should be equal to unity. CO₂ concentration can be converted from mass fraction of CO₂ with

$$C(\mathbf{x}, t) = Y_{\text{CO}_2}(\mathbf{x}, t) \cdot 10^6 \cdot \frac{\text{molecular weight of CO}_2}{\text{molecular weight of air}} \quad (16)$$

where molecular weight of CO₂ = 44.01g/mol and molecular weight of air is 28.97g/mol. The momentum equation describes the motion of air as an incompressible fluid, governed by the Navier-Stokes equations:

$$\frac{\partial(\rho \mathbf{u})}{\partial t} + \nabla \cdot (\rho \mathbf{u} \mathbf{u}) = -\nabla p + \rho \mathbf{g} + \nabla \cdot (\mu \nabla \mathbf{u}) - \nabla \cdot \boldsymbol{\tau}_t, \quad (17)$$

where p is the pressure field, \mathbf{g} is a vector of gravitational acceleration, μ is a molecular dynamic viscosity and $\boldsymbol{\tau}_t$ is a turbulence tensor. The energy conservation equation governs the transport of thermal energy within the airflow,

$$\frac{\partial(\rho e)}{\partial t} + \nabla \cdot (\rho e \mathbf{u}) = \nabla \cdot (k_{\text{eff}} \nabla T) - \nabla \cdot \left(\sum_i h_i \mathbf{j}_i \right), \quad (18)$$

where e is a specific internal energy, k_{eff} is an effective heat conductivity, T is fluid temperature and h_i refers to a specific enthalpy of fluid. The turbulence model equations approximate the effects of small-scale turbulent eddies

$$\tau_{t,ij} = \mu_t \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) - \frac{2}{3} \rho \kappa \delta_{ij},$$

where μ_t is a turbulent viscosity, κ is a turbulent kinetic energy and δ_{ij} is Kronecker's delta.

The domain boundary ∂Z encompasses all surfaces, including walls, ventilation interfaces, and occupant boundaries. We further define Ω_{supply} as the inlet vent boundary, Ω_{occupant} as the occupant surface boundary, Ω_{return} as the outlet vent boundary. The airflow boundary conditions are then given by:

$$\mathbf{u}(\mathbf{x}, t) = m_i^a(t) \begin{bmatrix} \sin(m_i^a(t)) \\ 0 \\ -\cos(m_i^a(t)) \end{bmatrix}, \forall \mathbf{x} \in \Omega_{\text{supply},i}, \quad (19a)$$

$$\mathbf{u}(\mathbf{x}, t) = \frac{v_{\text{occupant}}}{A_{\text{occupant}}} \cdot N_{\text{occupant}} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \forall \mathbf{x} \in \Omega_{\text{occupant}}, \quad (19b)$$

$$\mathbf{n} \cdot \nabla \mathbf{u} = 0, \forall \mathbf{x} \in \mathcal{Z}_{\text{return}}, \quad (19c)$$

$$\mathbf{u}(\mathbf{x}, t) = 0, \forall \mathbf{x} \in \partial \Omega \setminus (\Omega_{\text{supply}} \cup \Omega_{\text{occupant}} \cup \Omega_{\text{return}}). \quad (19d)$$

(19a) specifies that the airflow velocity at i -th group of supply vents is determined by its corresponding airflow rate $m_i^a(t)$ and direction angle $m_i^a(t)$. (19b) relates the airflow rate to the number of occupants, where the exhaled air rate v_{occupant} is set to 6 L/min per person [36], and A_{occupant} denotes the occupant boundary area. Constraint (19c) sets the Neumann boundary conditions at the return vent and constraint (19d) applies Dirichlet conditions to all other boundaries by setting the airflow velocity to zero [25].

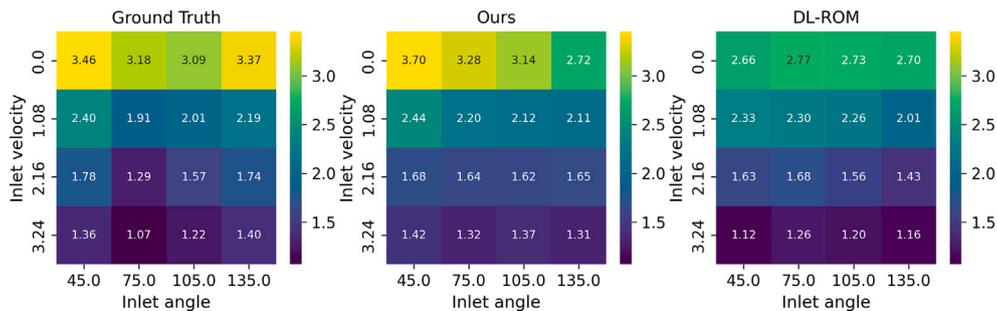


Fig. 9. Visualization of control objective values $\|\hat{C} - C_{\text{target}}\|_2^2$ across a grid of control inputs (airflow rate and vent angle), evaluated using (left) ground truth CFD simulations, (middle) our neural operator ensemble, and (right) DL-ROM [34].

Table 6

l_2 error and Control Landscape Error (CLE) for each model, with the best performance highlighted in bold.

Model	l_2 error (train dataset)	l_2 error (test dataset)	Control Landscape Error(CLE)
Ours	5.9 %	10.9 %	0.17
DL-ROM	8.4 %	9.6 %	0.26

Appendix B. Learning results of deep learning-based reduced-order modeling of CFD data

We compare the proposed neural operator learning approach against a state-of-the-art deep learning-based reduced-order modeling method for predicting the spatial-temporal dynamics of CO₂ concentration and building ventilation control. Specifically, we follow the work [34], which employed 3D autoencoder and U-Net architectures to perform nonlinear reduced-order modeling of CFD dynamics. We then integrate this learned surrogate model into our ventilation control optimization (5) to optimize control actions.

Apart from l_2 error, we propose a new metric, Control Landscape Error (CLE), to quantify discrepancies between ground-truth CFD results and model predictions (our neural operator or DL-ROM [34]) under different control conditions. We generate 16 unique control combinations by uniformly sampling airflow angles from 45° to 135° and rates from 0 to 3.24 m/s, applied identically across all six vents. For each, we simulate the future CO₂ distribution and use it to evaluate model accuracy across varying control inputs. We include a representative visualization in Fig. 9 to complement this quantitative evaluation. The control objective includes the term $\|\hat{C} - C_{\text{target}}\|_2^2$, which measures the deviation of the predicted CO₂ concentration from the target value. We visualize this objective value (with CO₂ concentrations standardized using Z-score normalization) for the ground truth CFD simulations, our neural operator ensemble, and the DL-ROM baseline. Our model captures the control-performance landscape more accurately, especially in lower-velocity regimes (that saves energy), whereas DL-ROM tends to over-smooth the variations, particularly at the extremes of the action space. This result demonstrates that our neural operator preserves the input-output sensitivity of the system, which is crucial for robust control optimization.

The model performance is reported in Table 6, where CLE is defined as

$$\text{CLE} = \frac{1}{N} \sum_{i=1}^N \left| \left\| C_{i,\text{gt}} - C_{\text{target}} \right\|_2^2 - \left\| C_{i,\text{model}} - C_{\text{target}} \right\|_2^2 \right|, \quad (20)$$

where $C_{i,\text{gt}}$ and $C_{i,\text{model}}$ denote the ground truth and model-predicted CO₂ fields for the i -th control input, and N is the number of evaluated inputs. Here, we generate 2 test samples (requiring approximately 32 CPU hours in total), each containing 16 control input pairs, resulting in $N = 32$ evaluations. While DL-ROM achieves a slightly lower l_2 error on the test set, our operator-based model attains a lower CLE, indicating better accuracy in capturing the control-response relationship. This highlights a key strength of our method: although the global l_2 error is comparable, our model more faithfully preserves the sensitivity of CO₂ outcomes to varying control inputs.

Data availability

I have included links in the manuscript to share the research data.

References

- [1] Chua KJ, Chou SK, Yang W, Yan J. Achieving better energy-efficient AIR conditioning—a review of technologies and strategies. *Appl Energy* 2013;104:87–104.
- [2] Chenari B, Carrillo JD, Silva MGD. Towards sustainable, energy-efficient and healthy ventilation strategies in buildings: a review. *Renewable Sustain Energy Rev* 2016;59:1426–47.
- [3] Hosseinloo AH, Nabi S, Hosoi A, Dahleh MA. Data-driven control of Covid-19 in buildings: a reinforcement-learning approach. In: *IEEE transactions on Automation science and Engineering*; 2023.
- [4] Shi S, Miyata S, Akashi Y. Event-driven model-based optimal demand-controlled ventilation for multizone vav systems: enhancing energy efficiency and indoor environmental quality. *Appl Energy* 2025;377:124683.
- [5] REHAV. Covid 19 guidance. <https://www.rehva.eu/activities/covid-19-guidance>.
- [6] Bian Y, Fu X, Liu B, Rachala R, Gupta RK, Shi Y. Bear-data: analysis and applications of an open multizone building dataset. In: *Proceedings of the 10th ACM International Conference on Systems for energy-efficient buildings, cities, and transportation*; 2023. p. 240–3.
- [7] Drgoña J, Arroyo J, Figueroa IC, Blum D, Arendt K, Kim D, Ollé EP, Oravec J, Wetter M, Vrabie DL, et al. All you need to know about model predictive control for buildings. *Annu Rev Control* 2020;50:190–232.
- [8] Chen Y, Shi Y, Zhang B. Optimal control via neural networks: a convex approach. In: *International Conference on Learning representations*; 2018.
- [9] Gao Y, Miyata S, Akashi Y. Energy saving and indoor temperature control for an office building using tube-based robust model predictive control. *Appl Energy* 2023;341:121106.
- [10] Su W, Ai Z, Liu J, Yang B, Wang F. Maintaining an acceptable indoor AIR quality of spaces by intentional natural ventilation or intermittent mechanical ventilation with minimum energy use. *Appl Energy* 2023;348:121504.
- [11] Du Y, Zandi H, Kotevska O, Kurte K, Munk J, Amasyali K, Mckee E, Li F. Intelligent multi-zone residential HVAC control strategy based on deep reinforcement learning. *Appl Energy* 2021;281:116117.
- [12] Yang T, Zhao L, Li W, Wu J, Zomaya AY. Towards healthy and cost-effective indoor environment management in smart homes: a deep reinforcement learning approach. *Appl Energy* 2021;300:117335.
- [13] Wang H, Chen X, Vital N, Duffy E, Razi A. Energy optimization for HVAC systems in multi-vav open offices: a deep reinforcement learning approach. *Appl Energy* 2024;356:122354.
- [14] Shang W, Liu J, Wang C, Li J, Dai X. Developing smart AIR purifier control strategies for better iaq and energy efficiency using reinforcement learning. *Build Environ* 2023;242:110556.
- [15] Li C, Cui C, Li M. A proactive 2-stage indoor co2-based demand-controlled ventilation method considering control performance and energy efficiency. *Appl Energy* 2023;329:120288.
- [16] Li B, Wu B, Peng Y, Cai W. Tube-based robust model predictive control of multi-zone demand-controlled ventilation systems for energy saving and indoor AIR quality. *Appl Energy* 2022;307:118297.
- [17] Tian W, Han X, Zuo W, Sohn MD. Building energy simulation coupled with CFD for indoor environment: a critical review and recent applications. *Energy Build* 2018;165:184–99.
- [18] Bianco N, Fragnito A, Iasiello M, Mauro GM. A CFD multi-objective optimization framework to design a wall-type heat recovery and ventilation unit with phase change material. *Appl Energy* 2023;347:121368.
- [19] Gao H, Qian W, Dong J, Liu J. Rapid prediction of indoor airflow field using operator neural network with small dataset. *Build Environ* 2024;251:111175.
- [20] Bulińska A, Popiołek A, Buliński Z. Experimentally validated CFD analysis on sampling region determination of average indoor carbon dioxide concentration in occupied space. *Build Environ* 2014;72:319–31.
- [21] Mou J, Cui S, Khoo DWY. Computational fluid dynamics modelling of airflow and carbon dioxide distribution inside a seminar room for sensor placement, measurement. *Sensors* 2022;23:100402.
- [22] Ning M, Mengjie S, Mingyin C, Dongmei P, Shiming D. Computational Fluid Dynamics (CFD) modelling of AIR flow field, mean age of AIR and co2 distributions inside a bedroom with different heights of conditioned AIR supply outlet. *Appl Energy* 2016;164:906–15.
- [23] Mahmoud MMA, Bahl P, D. A. Aquino A, MacIntyre C, Bhattacharjee S, Green D, Cooper N, Doolan C, de Silva C. A numerical framework for the analysis of indoor AIR quality in a classroom. *J Build Eng* 2024;92:109659.
- [24] Michoski C, Milosavljević M, Oliver T, Hatch DR. Solving differential equations using deep neural networks. *Neurocomputing* 2020;399:193–212.
- [25] Bian Y, Fu X, Gupta RK, Shi Y. Ventilation and temperature control for energy-efficient and healthy buildings: a differentiable PDE approach. *Appl Energy* 2024;372:123477.
- [26] Li Z, Kovachki NB, Azizzadenesheli K, Bhattacharya K, Stuart A, Anandkumar A, et al. Fourier neural operator for parametric partial differential equations. In: *International Conference on Learning representations*; 2021.
- [27] Lu L, Jin P, Pang G, Zhang Z, Karniadakis GE. Learning nonlinear operators via Deeponet based on the universal approximation theorem of operators. *Nat Mach Intell* 2021;3(3):218–29.
- [28] Zhou Q, Ooka R. Neural network for indoor airflow prediction with CFD database. In: *Journal of Physics: conference series*, vol. 2069. IOP Publishing; 2021. p. 012154.
- [29] Warey A, Kaushik S, Khalighi B, Cruse M, Venkatesan G. Data-driven prediction of vehicle cabin thermal comfort: using machine learning and high-fidelity simulation results. *Int J Heat Mass Transfer* 2020;148:119083.
- [30] Wang Y, Li Z, Yuan Z, Peng W, Liu T, Wang J. Prediction of turbulent channel flow using Fourier neural operator-based machine-learning strategy. *Phys Rev Fluids* 2024;9(8):084604.
- [31] Li Z, Kovachki N, Choy C, Li B, Kossaiji J, Otta S, Nabian MA, Stadler M, Hundt C, Azizzadenesheli K, et al. Geometry-informed neural operator for large-scale 3d PDES. *Adv Neural Inf Process Syst* 2024;36.
- [32] Hao Z, Wang Z, Su H, Ying C, Dong Y, Liu S, Cheng Z, Song J, Zhu J. Gnot: a general neural operator transformer for operator learning. In: *International Conference on machine Learning*, PMLR; 2023. p. 12556–69.

- [33] Ding X, Zhang H, Zhang W, Zhang W, Xuan Y. A Fourier neural operator-based method for rapid prediction of 3d indoor airflow dynamics. In: *Building Simulation*, Springer; 2025. pp. 1–17.
- [34] Pant P, Doshi R, Bahl P, Farimani AB. Deep learning for reduced order modelling and efficient temporal evolution of fluid simulations. *Phys Fluids* 2021;33(10).
- [35] Bulińska A, Buliński Z. A CFD analysis of different human breathing models and its influence on spatial distribution of indoor AIR parameters. *Comput Assist Methods Eng Sci* 2017;22(3):213–27.
- [36] He Y, Chu Y, Zang H, Zhao J, Song Y. Experimental and CFD study of ventilation performance enhanced by roof window and mechanical ventilation system with different design strategies. *Build Environ* 2022;224:109566.
- [37] Sha X, Ma Z, Sethuvenkatraman S, Li W. Online learning-enhanced data-driven model predictive control for optimizing HVAC energy consumption, indoor AIR quality and thermal comfort. *Appl Energy* 2025;383:125341.
- [38] Zhang S, Ai Z, Lin Z. Novel demand-controlled optimization of constant-AIR-volume mechanical ventilation for indoor AIR quality, durability and energy saving. *Appl Energy* 2021;293:116954.
- [39] Chen Y, Shi Y, Zhang B. Modeling and optimization of complex building energy systems with deep neural networks, In: 2017 51st Asilomar Conference on Signals, systems, and Computers. IEEE; 2017. p. 1368–73.
- [40] D'Alicandro AC, Mauro A. Experimental and numerical analysis of co2 transport inside a university classroom: effects of turbulent models. *J Build Perform Simul* 2023;16(4):434–59.
- [41] Manual U. Ansys fluent 12.0. Theory Guide 2009;67.
- [42] Abuhegazy M, Talaat K, Anderoglu O, Poroseva SV. Numerical investigation of aerosol transport in a classroom with relevance to Covid-19. *Phys Fluids* 2020;32(10).
- [43] Weekly K, Bekiaris-Liberis N, Jin M, Bayen AM. Modeling and estimation of the humans' effect on the CO 2 dynamics inside a conference room, In: *IEEE transactions on control systems technology*, vol. 23. 2015. p. 1770–81 (5).
- [44] ASHRAE. Standard 55—thermal environmental conditions for human occupancy, Tech. rep., Atlanta, GA: ASHRAE Inc; 1992.
- [45] Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. *Adv Neural Inf Process Syst* 2017;30.
- [46] ASHRAE. Standard 62.1—ventilation for acceptable indoor AIR quality, Tech. rep., Atlanta, GA: ASHRAE Inc; 2010.
- [47] Zhang W, Wu W, Norford L, Li N, Malkawi A. Model predictive control of short-term winter natural ventilation in a smart building using machine learning algorithms. *J Build Eng* 2023;73:106602.