atmosphere

MDPI

Article

Hybrid Digital Twin Framework for Real-Time Indoor Air Quality Monitoring and Filtration Optimization

Valentino Petrić [1], Dejan Strbad [2,3], Nikolina Račić [2,4], Tareq Hussein [5,6,*], Simonas Kecorius [7,8], Francesco Mureddu [2] and Mario Lovrić [9,10,*]

1 Faculty of Chemical Engineering and Technology, University of Zagreb, 10000 Zagreb, Croatia
2 The Lisbon Council, 1040 Brussels, Belgium
3 Ascalia d.o.o., 40000 Čakovec, Croatia
4 Institute for Medical Research and Occupational Health, 10000 Zagreb, Croatia
5 Institute for Atmospheric and Earth System Research (INAR/Physics), University of Helsinki, FI-00014 Helsinki, Finland
6 Environmental and Atmospheric Research Laboratory (EARL), Department of Physics, School of Science, The University of Jordan, Amman 11942, Jordan
7 Institute of Epidemiology, Helmholtz Zentrum München, 85764 Neuherberg, Germany
8 Environmental Science Center, University of Augsburg, 86135 Augsburg, Germany
9 Institute for Anthropological Research, 10000 Zagreb, Croatia
10 Faculty of Food Technology Osijek, Josip Juraj Strossmayer University of Osijek, 31000 Osijek, Croatia
* Correspondence: tareq.hussein@helsinki.fi (T.H.); mario.lovric@inantro.hr (M.L.)

Abstract

This study presents a hybrid digital twin system designed for real-time indoor air quality (IAQ) monitoring and filtration optimization within a residential environment. Using a network of low-cost sensors, physics-based simulations, and machine learning models, the system dynamically replicates the indoor environment to enable continuous assessment and optimization of key pollutants, including particulate matter, volatile organic compounds, and carbon dioxide. The system architecture integrates mass balance and decay models, computational fluid dynamics simulations, regression models, and neural network algorithms, all evaluated under both filtering and non-filtering conditions. A graphical user interface allows users to interact with the system, test air purifier placements, and visualize air quality dynamics in real time. The results demonstrate that, within this system, simpler models, such as linear regression, outperform more complex architectures under data-limited conditions, achieving test-set coefficients of determination ranging from 0.97 to 0.99 across multiple IAQ parameters. At the same time, the hybrid modelling approach enhances interpretability and robustness. Overall, this digital twin system contributes to smart building management by offering a scalable, interpretable, and cost-effective solution for proactive IAQ control and personalized decision-making.

Keywords: computational fluid dynamics; digital twin; hybrid modelling; low-cost sensors; machine learning; smart buildings; indoor air quality

Academic Editor: Klaus Schäfer

Received: 27 December 2025
Revised: 3 February 2026
Accepted: 5 February 2026
Published: 10 February 2026

1. Introduction

Indoor air quality (IAQ) is increasingly recognized as a key driver of personal exposure and health risk, as people spend considerable time indoors [1]. A wide range of indoor air pollutants, particularly particulate matter (PM) and gaseous compounds, originate from activities such as cooking, heating, cleaning, and human occupancy, while outdoor pollution may also infiltrate indoor environments [2–4]. Indoor microenvironments exhibit

firm spatial and temporal heterogeneity: emissions are episodic, ventilation rates fluctuate with behavior and weather, and geometry and furnishings shape airflow and deposition [5]. As a result, concentrations of PM, volatile organic compounds (VOCs), and carbon dioxide (CO_2) can change rapidly within and between rooms, complicating the monitoring and understanding of pollutant dispersion [4,6]. High indoor (and outdoor) concentrations of these pollutants are associated with various health effects. Elevated levels of CO_2 have been linked to impaired cognitive function, increased risk of headaches, fatigue, and long-term health deterioration [7]. Similarly, exposure to PM, particularly $PM_{2.5}$ (aerodynamic particle diameter less than 2.5 μm) and ultrafine fractions (diameter < 100 nm), is connected to respiratory and cardiovascular morbidity and mortality: short-term peaks exacerbate asthma and chronic obstructive pulmonary disease and can trigger arrhythmias and ischemic events, while long-term exposure increases risks of chronic bronchitis, reduced lung function, stroke, and lung cancer via pathways involving oxidative stress and systemic inflammation [8–11]. Traditional IAQ monitoring with reference-grade instruments is accurate but expensive and geographically sparse, limiting continuous, room-level coverage [12]. Low-cost sensors (LCS) enable dense, distributed networks. Yet, performance is often constrained by measurement noise, cross-sensitivities to temperature/humidity, aging-related drift, and inter-unit variability, making results highly context and calibration-dependent [13–16]. As a result, deployments face a trade-off between accuracy and scale; achieving both demands sensor fusion, online calibration/drift correction, and physics-guided quality control. This gap motivates a hybrid digital twin that combines dense LCS data with mechanistic models and machine learning (ML) to deliver robust, scalable IAQ monitoring and control.

Digital twins (DTs) create dynamic virtual replicas of physical systems that are continuously updated with real-time sensor data, enabling predictive modeling, resource optimization, and enhanced decision-making [17,18]. In contrast to traditional static simulations, DTs support bidirectional data exchange, enabling not only monitoring but also informed adjustments to the physical system [19]. This capability is particularly relevant in the built environment, where DTs are used for applications ranging from occupancy detection and Heating, Ventilation, and Air Conditioning (HVAC) optimization to building energy management and IAQ control [20]. Studies indicate that applying DT technologies in buildings can lead to energy savings typically reported in the range of 20–30%, while also reducing operational costs and improving indoor environmental quality [20–23].

Relevant is the emergence of hybrid DTs that integrate physics-based models with data-driven ML and deep learning (DL) algorithms. This hybridization enables accurate, real-time predictions while preserving the interpretability and domain knowledge embedded in physical laws [24–26]. Physics-informed ML models, such as those using state-space methods and Gated Recurrent Units, can capture nonlinear, temporally dependent behavior in sensor data with improved accuracy and generalizability compared to purely ML-based approaches [27,28]. Notably, recent studies have demonstrated that CO_2-based occupancy detection models grounded in physical principles outperform conventional ML models by eliminating the need for room-specific customization, thereby enhancing scalability across diverse environments [29]. In addition to occupancy modeling, Computational Fluid Dynamics (CFD) plays a foundational role in IAQ assessment, providing a better understanding of airflow and pollutant dispersion in 2D and 3D spaces [30]. On the one hand, standalone CFD models are computationally intensive, limiting their real-time use; however, they can still serve as valuable tools for describing airflow patterns to stakeholders and for validating simplified physics-based models. In this context, physics-informed approaches bridge the gap between purely data-driven models and detailed numerical simulations, preserving scalability and interpretability. To address these challenges, this

study integrates low-cost sensing technologies with physical and data-driven modeling. Emphasis is placed on sensor selection, calibration, and performance evaluation, including accuracy, reliability, and drift mitigation through manufacturer calibration and data-driven validation. The framework incorporates physical models, such as mass-balance formulations for indoor pollutant dynamics and CFD simulations, to examine airflow and particle transport. These models are coupled with ML to improve the interpretability and robustness of IAQ predictions. Within this framework, CFD is used to visualize airflow and pollutant dispersion and to support ventilation optimization by identifying stagnation zones, assessing ventilation effectiveness, and evaluating alternative ventilation and air purification strategies. This enables informed optimization of window ventilation and air filter placement, particularly in naturally ventilated environments. To ensure robust and generalizable predictions, multiple ML approaches are evaluated. Linear regression (LR) provides a transparent and reliable baseline under limited data; tree-based ensembles, such as Random Forests (RF), capture nonlinear relationships with high stability; and neural networks, including LSTM models, model temporal dependencies in IAQ data. Comparing these methods within a unified framework enables a systematic assessment of robustness, interpretability, and predictive performance under realistic sensing constraints.

This study aims to develop and evaluate a DT framework for investigating IAQ in residential environments. The proposed hybrid DT integrates four complementary components: LCS networks, CFD simulations, physical models, and ML techniques. This work focuses on a residential indoor environment under natural ventilation, aiming to capture temporal and spatial variability in key IAQ parameters and to evaluate the impact of air purification strategies. By employing robust, scalable, and interpretable modeling approaches, the proposed system overcomes key limitations of prior studies, including low spatial resolution, high calibration demands, and limited real-time applicability. Through bidirectional interaction between physical data and models, the DT provides a foundation for future autonomous and adaptive IAQ management, supporting sustainable and intelligent building operation while contributing to environmental and public health objectives.

2. Materials and Methods

All experimental, modeling, and data-analysis methods employed in this study follow established practices reported in the scientific literature. The overall methodology and the specific phases of the study, ranging from experimental setup to DT deployment, are summarized in the flowchart provided in Figure S1 of the Supplementary Information (SI). No formal ISO or EN measurement standards were explicitly applied during the experimental campaign.

2.1. Experimental Setup and Data Collection

The study was conducted in a bedroom located on the first floor of an occupied family home (Figure 1), situated in a suburban neighborhood approximately 20 km from Zagreb, Croatia. Detailed photographs of the study site and a geographical overview of the location are provided in Figure S2 of the SI. The floor plan shown in Figure 1 is a schematic that illustrates the relative position of the test room within the house. Ventilation was characterized using measured window-opening events and experimentally derived air-exchange rates rather than geometry-based airflow modeling. Furthermore, the north–south orientation of the house was not considered a controlling variable in this study, since outdoor air quality was directly monitored using an outdoor sensor representative of the air entering the room, independent of wind direction or solar exposure. The house is situated at the crossroads of a main road and a dead-end street. Nearby, there are two factories: one

producing pharmaceuticals and the other laser-cutting metal sheets. Traffic in the area is significantly lighter compared to busy city center streets.

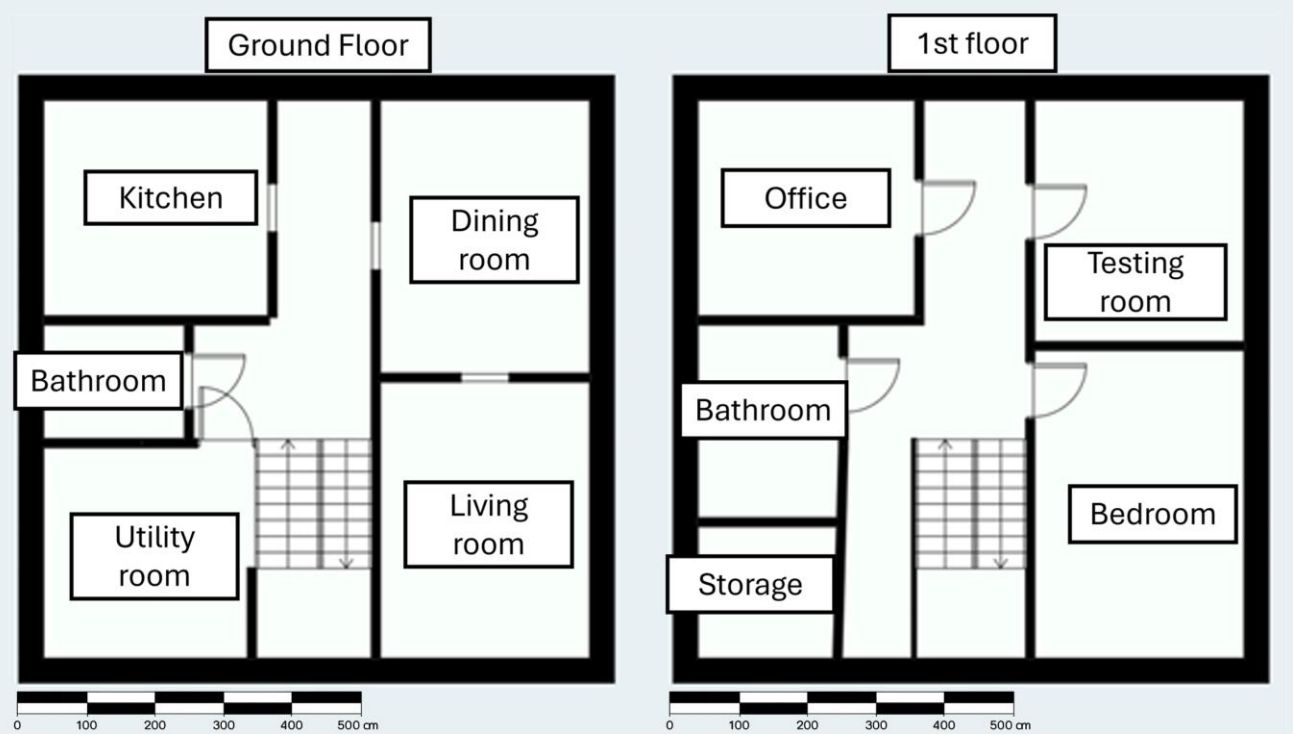

Figure 1. Study site: 2D plan of the house showing the ground floor layout and the location of the test room, which is situated on the first floor.

For IAQ data collection, six LCS were used: four from Wings ICT (Wings ICT Solutions, Athens, Greece) and two from LabService Analytica (LabService Analytica srl, Anzola Emilia (BO), Italy). The accuracy, calibration procedures, and measurement uncertainty of these sensors have been documented in previous studies [31]. Briefly, the Wings devices exhibit typical accuracies of ±10–15% for PM ($PM_{2.5}$ and PM_{10}), ±50 ppm for CO_2, ±0.5 °C for temperature, and ±3% for relative humidity, while the LabService Analytica sensors show comparable or slightly improved performance, with PM uncertainties below ±10% and CO_2 accuracy within ±30–50 ppm. In addition to these LCS, Xiaomi door and window sensors (Xiaomi Inc., Beijing, China) were used to detect window openings and closings. A Basler (Basler AG, Ahrensburg, Germany) camera was employed to monitor street traffic outside. All data from the sensors and the camera were captured using a Raspberry Pi (Raspberry Pi Ltd., Cambridge, UK), a small single-board computer designed for embedded applications. Using Bluetooth and custom software developed in Python 3.10 (Python Software Foundation, Wilmington, DE, USA), the Raspberry Pi collected and processed data, enabling integration with sensor and video inputs.

Outdoor environmental conditions were monitored using a Wings outdoor device (Wings ICT Solutions, Athens, Greece), which measures gaseous pollutants (NO_2, O_3, SO_2, CO, NO, and CO_2), PM (PM_1, $PM_{2.5}$, PM_{10}), meteorological variables (temperature, relative humidity, and atmospheric pressure), and noise levels; a complete list of measured parameters is provided in Table S4.

The indoor sensors were positioned using a Cartesian room-based coordinate system (X, Y, Z), with the origin at the lower corner of the room at floor level. Six sensors were installed at two horizontal locations in the room: near the window (X = 115 cm, Y = 137 cm) and near the center (X = 310 cm, Y = 225 cm). At each horizontal location, sensors were

deployed at three heights to capture vertical gradients: 13 cm, 100 cm, and 145 cm above the floor. The complete set of sensor coordinates was therefore given in Table S1. The test room was unfurnished and remained unoccupied throughout the measurement campaign to minimize uncontrolled indoor emission sources. The room was equipped with a split central heating system; however, the heating was not operated during the study period to avoid introducing additional thermal or airflow disturbances unrelated to window ventilation or air purification. Indoor temperature and relative humidity were therefore not actively controlled and were governed by passive factors, including solar gains, heat exchange with the building envelope, outdoor temperature variations, and air exchange through window opening and natural infiltration. During periods when the room was closed, indoor PM concentrations primarily reflected residual particles from prior ventilation events, background infiltration through the building envelope, and minor resuspension from interior surfaces rather than ongoing indoor activities. No intentional indoor pollutant-generating activities were conducted during the study.

Sensors from both providers recorded measurements every minute from 23 November 2024 to 19 December 2024. The exact sensor placements are detailed in Table S1 of the SI. All indoor and outdoor sensors used in this study were factory-calibrated by the respective manufacturers prior to shipment. Calibration was performed using the manufacturer's standard procedures and reference instrumentation under controlled laboratory conditions. No additional on-site calibration was conducted before deployment. The expected accuracy ranges and measurement uncertainties of the sensors have been previously validated and are summarized in the SI (Tables S2 and S3). Only one window was used during the study. The second window was used for traffic observation, with a camera mounted on it aimed at the street to monitor traffic conditions. An outdoor Wings sensor was placed on the balcony above the window to monitor outdoor air quality representative of the air entering the room. This location was selected due to architectural constraints and the absence of suitable power outlets or secure mounting options at window height. The balcony position provided stable installation conditions and unobstructed airflow, and the small vertical offset was not expected to affect the representativeness of the measured outdoor concentrations significantly.

All the data collected in this study are presented in Table S4 in the SI, along with the exact measurement units and parameters. Once measurements began, only a few activities were conducted in the room due to space limitations. These included opening the window, conducting experiments related to the mass balance model (explained in Section 2.3, and repositioning the air-filtering unit. One objective of the DT was to evaluate how the filtering unit's position affects IAQ. To purify the room air, we used the iTherapy air purifier (Kristalna ideja, Model 100, Split, Croatia), which has an intake velocity of 2.8 m/s, an outflow velocity of 1.5 m/s, an outflow area of 62.1 cm^2, and an airflow rate of 33.53 m^3/h. The air purifier features, as per the producer's description, a five-layer filtration system: a pre-filter to remove $PM_{2.5}$, dust, and fibers; an antibacterial layer targeting allergens and viruses; nano silver membranes for capturing 0.3-micron particles; a cold catalytic layer for decomposing harmful chemicals; and a vitamin C composite layer to neutralize odors. The filtering unit was placed inside the room and turned on after the first week of measurements, allowing baseline air quality to be captured without its influence. This approach ensured that any changes observed after activation could be attributed to the filter's impact. The air-filtering unit was repositioned daily and tested at 14 predefined locations in the room at two installation heights (50 cm and 150 cm above the floor). All filter placements were defined using the same room-based Cartesian coordinate system (X, Y, Z) as for sensor deployment, and the full list of coordinates is provided in Table S5 of the SI. The filter material was replaced twice during the measurement cycle as part of a separate analysis for future tasks.

2.2. CFD Simulations in Indoor Spaces

The CFD methodology adopted in this study follows established best practices for indoor airflow and pollutant dispersion modeling, as documented in the IAQ literature. The airflow and particle transport within the experimental room were simulated using SimScale (version 2.0 SimScale GmbH, Munich, Germany), a cloud-based CFD platform. SimScale utilizes an OpenFOAM backend for its numerical solvers. In this study, the "simscaleBuoyantBoussinesqSimpleFoam" solver was employed, which applies the Boussinesq approximation to account for buoyancy-driven air motion due to temperature gradients. The simulation time was set to 5000 s, with a time step (deltaT) of 5 s, and results were written every 100-time steps. This temporal configuration allowed the observation of transient airflow development and steady-state behavior. The airflow was modeled as a Newtonian fluid with constant physical properties defined as follows: dynamic viscosity (ν): 1.5295×10^{-5} m^2/s, reference density: 1.1965 kg/m^3, thermal expansion coefficient (β): 0.00343 K^{-1}, reference temperature: 293.15 K (20 °C), laminar Prandtl number (Pr): 0.713, turbulent Prandtl number (Prt): 0.85, and specific heat capacity (Cp_0): 1201.29 J/kg·K. These parameters correspond to typical indoor air conditions and were used to accurately model convective heat transfer and buoyancy effects. The turbulence was modeled using the Reynolds-Averaged Navier–Stokes (RANS) approach.

The k–ω SST (Shear Stress Transport) model was selected for its robustness in resolving both near-wall and free-shear-layer turbulence. The computational domain was discretized using a three-dimensional unstructured finite-volume mesh generated in SimScale. The final grid consisted of approximately 13.25 million control volumes. The mesh was created using the Standard meshing algorithm with automatic sizing, curvature-based refinement, physics-based meshing, and automatic boundary layers, employing a hex-dominant core. Mesh quality metrics were within acceptable ranges, with a maximum non-orthogonality of 84.6 (average 7.8) and a maximum skewness of 2.5 (average 0.1). This grid resolution was selected to capture key indoor airflow structures relevant for the qualitative visualization of airflow and particle transport. While the k–ω SST turbulence model was employed and near-wall resolution followed recommended best practices, the y^+ parameter was not explicitly evaluated and is therefore not reported. A formal grid-independence study was not performed in this work. The CFD simulations were used primarily for qualitative visualization and comparative assessment of airflow and particle transport patterns rather than for fully grid-converged quantitative prediction. Confidence in the numerical grid was based on high mesh resolution, acceptable mesh quality metrics, physics-based meshing with automatic boundary layers, and the use of the k–ω SST turbulence model, following established best-practice guidelines for indoor airflow simulations. The CFD model represents a naturally ventilated indoor environment without a mechanically forced air supply or exhaust. No explicit inlet or outlet velocity boundary conditions were prescribed. Airflow within the room was driven by buoyancy forces from temperature gradients and by localized momentum sources from the air-filtering unit. All solid surfaces, including the walls, floor, and ceiling, were modeled as no-slip boundaries. The window opening was treated as an opening-type boundary condition, allowing bidirectional flow in response to the local pressure field, thereby enabling natural air exchange with the outdoor environment without imposing fixed inflow or outflow rates. The resulting flow fields were also consistent with experimentally observed ventilation behavior, supporting the suitability of the selected grid for the intended purpose. Following the airflow simulations, a Lagrangian particle tracking phase was conducted to further analyze the influence of the simulated flow field on particle dispersion and deposition. In this stage, particle traces were generated during post-processing using SimScale's Particle Trace filter to visualize airflow pathways and transport patterns within the room. The traces represent massless tracer points that

follow the computed velocity field and were seeded at user-defined locations within the fluid domain. These particle traces were used solely for qualitative visualization and do not represent physically injected particles with prescribed mass, number, or size distribution. The motion of each particle was governed by aerodynamic drag and buoyancy forces, with one-way coupling assumed, meaning the particles followed the flow field without influencing it. Because the SimScale particle model does not account for particle-wall interactions, such as adhesion or rebound, wall-friction effects were neglected. Consequently, particle velocity magnitudes near solid surfaces may appear higher than expected because frictional deceleration and deposition are not accounted for. Despite this limitation, the simulation reliably describes bulk-flow particle interactions and enables identification of regions characterized by strong recirculation, turbulence, and particle suspension, which are essential for understanding pollutant behavior in indoor environments.

The CFD simulations were not used to compute a formal optimization metric; instead, they served as a qualitative tool to assess ventilation effectiveness and spatial optimization by comparing airflow structures and particle transport across different air purifier placements.

2.3. Physical Models

In this study, we incorporated the indoor aerosol model (IAM) based on the mass-balance equation. This was applied to indoor PM concentrations and to the decay of CO_2 concentration [5]. The mathematical form of the IAM is defined in Equation (1) [32] where the change rate of the indoor concentration I is governed by the outdoor concentration O, the penetration factor P, and the ventilation rate λ. The model also accounts for particle loss due to ventilation (i.e., λI) and deposition (i.e., $\lambda_d I$). Finally, S represents an effective indoor source term that accounts for all indoor contributions not explicitly modeled by ventilation and deposition processes, including human activity, resuspension, or other unmodeled disturbances:

$$\frac{dI}{dt} = P\lambda O - \lambda I - \lambda_d I + S. \quad (1)$$

To determine the model parameters [33], two experiments were performed using candles as sources of indoor pollution. In the first experiment, candles were burned for two hours in a closed room. They were then extinguished, and the window was opened. This setup allowed for the estimation of λ. In the second experiment, the same procedure was followed, but the room remained closed for 24 h to determine λ_d. A visual explanation of the method is provided in Figure S3. The candle experiments were conducted solely for parameter calibration purposes. During the decay phases used to estimate λ and λ_d the indoor source term was set to $S = 0$ after extinguishing the candles. These calibration experiments were independent of the main measurement campaign.

Once the relevant time-series data were collected using LCS, LR was applied to estimate the ventilation rate and the deposition rate λ and λ_d as described by [33]. Assuming steady-state concentrations indoors and outdoors, the penetration factor P can be estimated as shown in Equation (2):

$$P = \frac{(\lambda + \lambda_d)I}{\lambda \times O}. \quad (2)$$

These values were subsequently included as features in the training of ML models. This analysis was conducted for both PM_{10} and $PM_{2.5}$. Figure S4 presents the results of the IAM simulations, showing temporal changes in air exchange and the impact of indoor sources. Notably, large spikes in the indoor source term or significant changes in dI/dt correspond to events such as candle burning, indicating PM_{10} production from combustion [5].

While the mass balance model is well suited for modeling aerosols, CO_2 behavior is better captured using a modified and simplified version for the IAM, as shown in Equation (3) [34]:

$$C_t = C_{ext} + (C_0 - C_{ext}) \times e^{-nt} \quad (3)$$

Here, C_t is the CO_2 concentration at time t, C_0 is the initial concentration, C_{ext} is the outdoor (steady state) concentration and n is the CO_2 decay constant that dictates the air change rate (i.e., ventilation rate). To determine the value of n, an experiment like the previous one was conducted. Candles were burned until extinguished, and the window was then opened to allow ventilation. CO_2 concentration decay was tracked, and the constant n was calculated using Equation (3), with all other variables known. This decay model is incorporated into the dataset used for training gas prediction models.

2.4. Hybrid Integration of Physical and Data-Driven Models

The proposed DT follows a hybrid modeling strategy in which physical and data-driven components play complementary roles. Physical models are first used to describe fundamental IAQ processes and to inform data interpretation. Mass-balance equations are used to derive physically meaningful quantities, such as ventilation rates, deposition rates, penetration factors, and effective indoor source terms, which are then incorporated as engineered features in ML models. In parallel, CFD simulations are employed to resolve airflow structures and particle transport patterns within the room. While CFD results are not directly coupled to the ML models, they provide qualitative insight into ventilation effectiveness, spatial variability, and the influence of air purifier placement. This information supports the interpretation of sensor data and helps contextualize model predictions. The ML models are trained on measured sensor data augmented with features derived from the physical models, allowing the hybrid framework to retain physical interpretability while benefiting from the predictive capability of data-driven approaches. Embedding physically meaningful constraints and features into the learning process enhances robustness by reducing sensitivity to measurement noise, sensor bias, and limited data availability. Interpretability is improved by linking model inputs and outputs to well-understood physical processes governing ventilation, pollutant transport, and air purification. At the same time, computational efficiency is maintained by combining lightweight physical models and offline CFD analysis with relatively low-complexity ML models, enabling near–real-time operation suitable for scalable DT applications.

2.5. Data Preparation

All data were consolidated into a single dataframe, which served as the basis for subsequent preprocessing and analysis. We implemented physical models to construct a hybrid modelling framework. The process began with applying the mass balance model to PM_{10} and $PM_{2.5}$. Using measured indoor and outdoor concentration time series and previously estimated parameters (λ, λ_d and P), Equation (1) was rearranged to compute the effective indoor source term S. Although the room was unoccupied and no intentional indoor emission activities occurred during the study, S captures residual indoor contributions from unmodeled effects, such as particle resuspension, airflow disturbances, or the operation of the air-filtering unit. This calculation was performed exclusively during periods when the window was open, to account for air exchange. When the window was closed, the term representing outdoor air concentration was omitted, as small air infiltration was assumed. The decay method model was applied, which is only relevant under conditions of active air exchange between the indoor and outdoor environments. Indoor pollutant concentrations were measured at three time points: 1, 5, and 10 min after the window was opened. These intervals were selected because the indoor environment

reached a steady state within 10 min, beyond which concentration changes were negligible. This steady state represents environmental equilibrium, and the indoor concentration C changes when the window is opened. The model was further refined by incorporating the air-filtering unit's operational status and spatial position. Euclidean distances were computed between the filtering unit and each sensor based on their spatial coordinates. A binary operational factor normalized these distances: 1 when the unit was active and 0 when inactive. When the unit was not operational, the distance term was omitted because it did not affect sensor measurements.

In addition to feature engineering, data preprocessing included outlier removal using a sliding 3-day window. Observations falling below the 1st percentile or above the 99th percentile were discarded. Missing values were imputed using the IterativeImputer from scikit-learn, which treats the imputation task as a sequence of regression problems, iteratively modeling each feature with missing values as a function of the others to improve accuracy [35]. Furthermore, lag features were introduced for all measurements from both indoor and outdoor LCS to capture the temporal dependence of current readings on previous values. To assess the influence of temporal patterns such as the day of the week and hour of the day, two sinusoidal features were created: day_sin and hour_sin. These features employed a cyclic representation of time, with hour_sin ranging from midday to midnight and day_sin ranging from Monday to Sunday, capturing daily and weekly periodicity [36]. Finally, the dataset was divided into two subsets based on the operational status of the air-purifying unit: one in which it was active and one in which it was not.

2.6. Model Training

In this study, 70 models were developed using Python 3.10 and its associated libraries, primarily scikit-learn (version 1.6) [35] and PyTorch (version 2.1.2) [37]. For each of the seven parameters monitored: PM_{10} (aerodynamic particle diameter less than 10 µm), $PM_{2.5}$, temperature, relative humidity (RH), pressure, total volatile organic compounds (TVOC), and CO_2, two distinct datasets were employed: one corresponding to scenarios with the filtering unit operational and one without it. Five regression models were evaluated: Linear Regression (LR), Random Forest Regression (RF), XGBoost, Multi-Layer Perceptron (MLP), and Long Short-Term Memory (LSTM). The five regression models were selected to provide a representative comparison across different modeling paradigms. Linear models were included as interpretable baselines, while tree-based and ensemble models were selected for their ability to capture nonlinearities and feature interactions common in IAQ data. This combination allows evaluation of the trade-off between model simplicity, interpretability, and predictive performance within the proposed hybrid modeling framework.

Data preprocessing included feature scaling for all models except RF and XGBoost, which are not sensitive to feature scaling. Model hyperparameters were optimized via grid search implemented with the scikit-learn library. Cross-validation was employed during hyperparameter tuning to ensure robust model evaluation; therefore, a separate validation set was not constructed. The datasets were split into training and test sets at 80/20, with 80% of the data allocated to training and 20% to testing.

For the dataset without the filtering unit, the training set comprised 24,733 data points, and the test set 6184. In contrast, the dataset with the filtering unit contained fewer observations: 5291 for training and 1323 for testing. The datasets differ because we recorded the exact times when the filtering unit was on and off. The reduced volume of data in the filtering scenario reflects periods when the unit was operating, while the rest of the time it was off. Such disparities can significantly influence model performance, particularly complex algorithms that require larger datasets for practical training. The final input to each model included outdoor sensor measurements, all preprocessed features (e.g., lag

variables and temporal features), and outputs from mass-balance models. The model's targets were indoor sensor readings, resulting in six distinct output variables.

2.7. Prophet Model Setup

To assess the feasibility of time-series forecasting using only sensor data, the Prophet algorithm, developed by Facebook's research team [38], was applied. Prophet decomposes a time series into trend, seasonality, and holiday components, allowing flexible modeling of periodic or event-driven variations. Its robustness to missing data and outliers makes it suitable for sensor-based applications. A separate Prophet model was trained for each indoor sensor using default settings, mirroring the setup of other ML models.

2.8. System Architecture and Deployment

The DT is implemented as a modular, scalable web-based application, deployed to a Kubernetes cluster to ensure reliability, dynamic scalability, and maintainability. Its architecture is structured around two main components: the Backend Service and the Frontend Service, which interact seamlessly to provide a responsive and robust user experience. The backend service encapsulates DT's core functionality. It is built on the FastAPI framework and manages data processing pipelines, integrates with external systems, and orchestrates computational logic. The backend is containerized and deployed as pods within the Kubernetes cluster, with horizontal pod autoscaling activated to dynamically allocate resources based on CPU usage, memory consumption, and API request load. Services are exposed through Kubernetes-native Service objects, enabling both intra-cluster communication and secure external access [39]. The frontend service delivers the user interface and is responsible for displaying model outputs and managing user interactions. Static files (HTML, CSS, JS) are hosted via a reverse proxy, such as Nginx. This service is also containerized and runs in its own pods, enabling isolated scaling and deployment. It communicates directly with the backend via API endpoints, ensuring synchronized data flow between user inputs and model predictions. The application relies on Kubernetes-native internal networking and service discovery to ensure efficient communication between backend and frontend components. External traffic is routed via an ingress controller, which distributes traffic based on defined path or domain-based rules, supporting both performance optimization and secure access [40]. A Continuous Delivery (CD) pipeline automates the entire application lifecycle from building and testing to deployment, ensuring consistent, rapid updates. Integration with version control systems enables traceable and reproducible deployments, accelerating iteration cycles while maintaining system stability. This architecture demonstrates the system's readiness for real-world deployment, not only as a research prototype but as a production-level solution with robust DevOps integration and comprehensive performance observability.

2.9. Model Evaluation

The trained models were evaluated using three standard performance metrics: Coefficient of Determination (R^2), Root Mean Squared Error ($RMSE$) and Mean Absolute Error (MAE). R^2, also known as the goodness of fit measure, quantifies the proportion of variance in the dependent variable explained by the independent variables in the model. It is computed as shown in Equation (4), where y_i denotes the measured value, $\hat{y}_i$ represents the corresponding model prediction, $\overline{y}_i$ is the mean of the measured values, and n is the total number of samples. Higher R^2 values indicate better predictive performance.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2} \quad (4)$$

Since R^2 alone may not fully capture model accuracy, *RMSE* and *MAE* were also utilized, as defined in Equations (5) and (6), respectively. Both metrics are error-based and should ideally be as low as possible. *RMSE* penalizes larger errors more heavily, making it suitable when larger deviations are especially undesirable.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (5)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (6)$$

3. Results and Discussion

In this section, we present the created tool, use cases, and simulation results.

3.1. The Created Tool

The GUI provides an intuitive and interactive environment for configuring simulation scenarios, visualizing results, and optimizing IAQ interventions. On the left side of the screen, users can define environmental parameters such as:

- Indoor and outdoor air quality data;
- Nearby traffic conditions (categorized into cars, trucks, buses, and motorcycles);
- The operational status of the window, which represents the sole ventilation mechanism in the simulation.

A central feature of the system is the ability to manually specify the air purification unit's position within the 3D-scanned model of the experimental room. Users can enter coordinates (X, Y, Z) to place the air purification device, with the origin (0, 0, 0) at the corner where the small window meets the door. This functionality enables flexible testing of multiple spatial configurations to assess how placement affects predicted indoor pollutant concentrations. Once the input parameters are configured, users can initiate the simulation by clicking the "Run" button. The simulation output is displayed on the right side of the interface, presenting two sets of predicted values: one for the "Filter" scenario (air purification unit active) and one for the "No Filter" scenario (air purification inactive) (Figures S6 and S7). This dual-output view facilitates direct comparison of the filtration unit's effectiveness under various spatial and environmental conditions.

To enhance user engagement and knowledge transfer, a chatbot is integrated above the results section. The chatbot automatically receives simulation outputs and provides personalized recommendations to improve IAQ based on the user's configured scenario. The GUI includes three main tabs that support different visualization modes:

- The first tab displays the 3D-scanned room model, allowing users to visualize the layout and potential placements for air purification units.
- The second tab (Figure S5) shows the CFD simulation of air fluctuation patterns.
- The third tab (Figure S6) presents the CFD simulation of particle distribution dynamics.

These CFD visualizations complement the numerical outputs by providing an intuitive understanding of airflow and pollutant dispersion in relation to filter placement.

Overall, this case demonstrates the practical utility of the air purification DT for evidence-based IAQ management. By enabling users to configure multiple air purification unit positions, run simulations, and directly observe the predicted impact on pollutant concentrations, the platform serves as a valuable decision-support tool for researchers, building managers, and environmental engineers.

3.2. Comparison of Environmental Parameters Regarding Air Purification Using the Digital Twin

Figure 2 presents a statistical comparison of indoor environmental parameters across the entire measurement campaign, separated into periods when the air-filtering unit was inactive and when it was active. Each box plot represents the distribution of all recorded values under the respective operating conditions. The median values shown in the figure were calculated across all six sensors, ensuring that the results reflect the overall indoor environment rather than a single device's readings. To quantify the improvement achieved by the air purifier, percentage reductions were calculated using the median values shown in Figure 2, comparing periods when the purifier was inactive and active. The results clearly indicate that the air purification system had a notable effect on particulate matter concentrations, as expected, since the primary goal of air purification is to remove particulate pollutants. Other environmental variables exhibited more moderate or negligible changes.

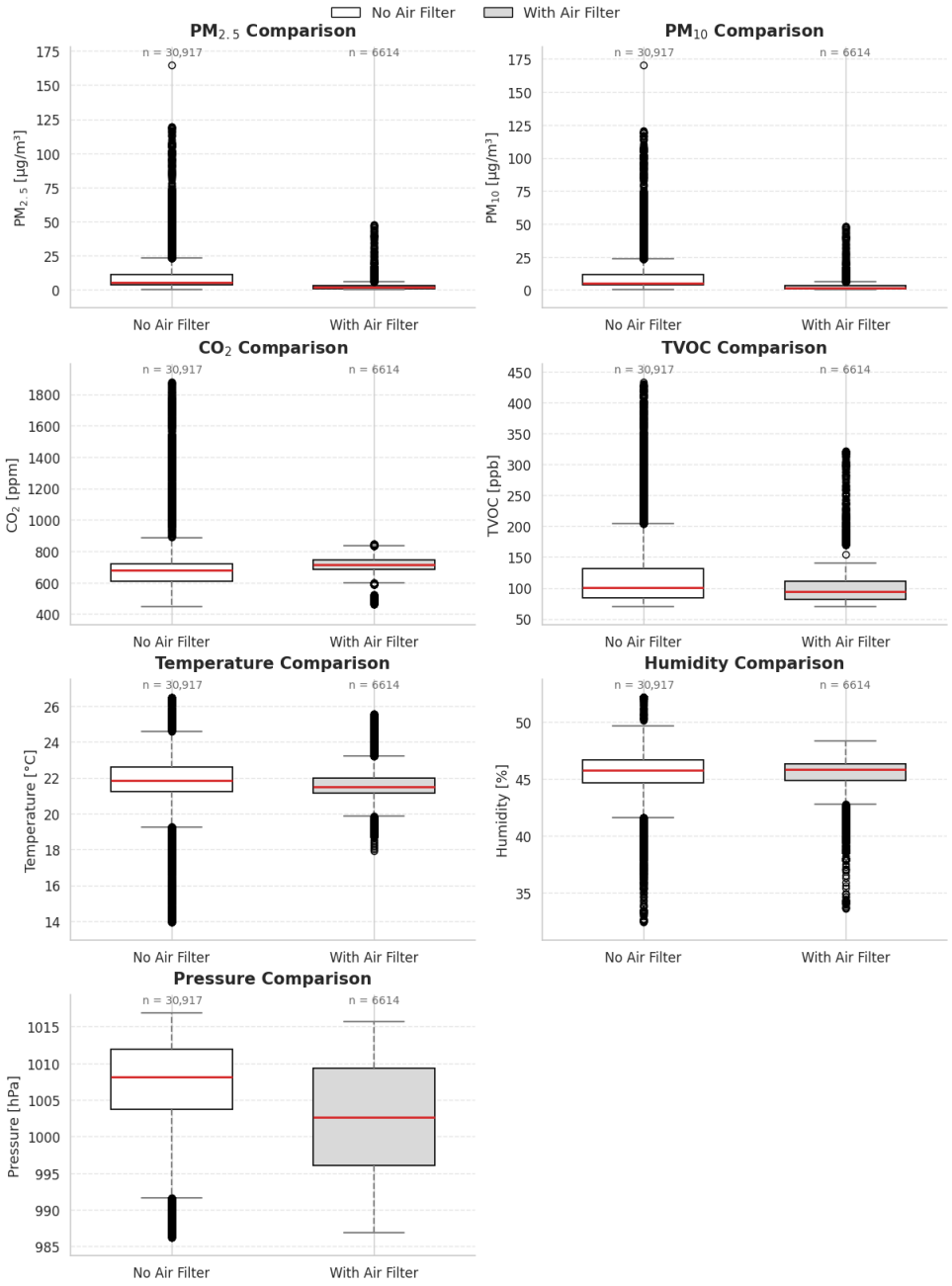

Figure 2. Comparative Analysis of Indoor Environmental Parameters with and without Air Purification (Median Across Six Sensors).

Both $PM_{2.5}$ and PM_{10} concentrations decreased significantly when the air purifier was active. Median values decreased markedly, and the overall distributions were more compact, suggesting effective removal of fine and ultrafine particles by the filtration unit. The reduction in variability also indicates more stable IAQ conditions under filtration. For CO_2, the difference between the two conditions was less pronounced. Although median concentrations were slightly lower with the filter operating, the distributions overlapped considerably. This suggests that CO_2 levels are primarily influenced by occupancy and ventilation rather than filtration efficiency. A similar trend was observed for TVOCs, with the air purifier use resulting in lower median concentrations and a narrower range. This indicates partial removal or adsorption of VOCs by the purification device, likely due to the presence of activated carbon or similar adsorptive media within the system. In contrast, temperature, humidity, and pressure remained relatively stable across both conditions. Minor differences observed are likely due to normal fluctuations in indoor climate control rather than to the air purification process itself. The observed reduction in $PM_{2.5}$ and PM_{10} concentrations during air purifier operation is consistent with previous studies in residential indoor environments, which report substantial PM removal when portable air cleaners are used under natural or mixed-ventilation conditions. Similar reductions in median PM concentrations have been reported, highlighting the effectiveness of portable filtration devices in suppressing elevated indoor particle levels [41,42].

3.3. CFD Simulations

To investigate air flow dynamics and particle behavior within the experimental room, CFD simulations were performed using SimScale. The simulations aimed to visualize and quantify airflow patterns and their influence on particle transport under controlled ventilation conditions. Figure 3 presents the velocity magnitude distributions within the room across the X, Y, and Z planes. The top row corresponds to the flow field after 2500 s, while the bottom row shows the evolution after 5000 s. The color contours illustrate spatial variations in air velocity, highlighting regions of recirculation and stagnation. Over time, the airflow distribution becomes more homogeneous, with turbulent mixing zones forming near the inlets and outlets. These turbulent zones enhance air exchange and improve the dispersion of airborne substances.

To further examine PM dynamics, 10 µm particles were introduced into the simulated domain. Figure 4 shows the resulting particle trajectories after 5000 simulation seconds, visualized in the X, Y, and Z planes. Each trajectory corresponds to the motion of an individual particle, colored by its instantaneous velocity magnitude. The results reveal complex, swirling flow structures indicative of turbulence-driven particle motion. Turbulent regions are associated with reduced deposition rates and increased particle resuspension, leading to locally higher particle concentrations.

These findings suggest that ventilation effects, such as those induced by mechanical air exchange systems or open windows, can increase suspended PM concentrations while simultaneously reducing gaseous pollutant concentrations, such as CO_2. However, SimScale currently lacks wall–particle interaction modeling. Consequently, near-wall velocity magnitudes are higher than expected because the software does not account for frictional velocity reduction at solid boundaries. This limitation should be considered when interpreting near-wall particle dynamics. The airflow structures and particle transport patterns identified through CFD simulations are consistent with previous numerical studies of naturally ventilated indoor spaces, which similarly emphasize the influence of room geometry, ventilation openings, and purifier placement on airflow organization and particle dispersion [43,44]. In this study, filtration optimization refers to a qualitative, CFD-informed evaluation of air purifier placement and operation rather than a formal

numerical optimization, with the aim of identifying configurations that promote effective airflow interaction and pollutant removal.

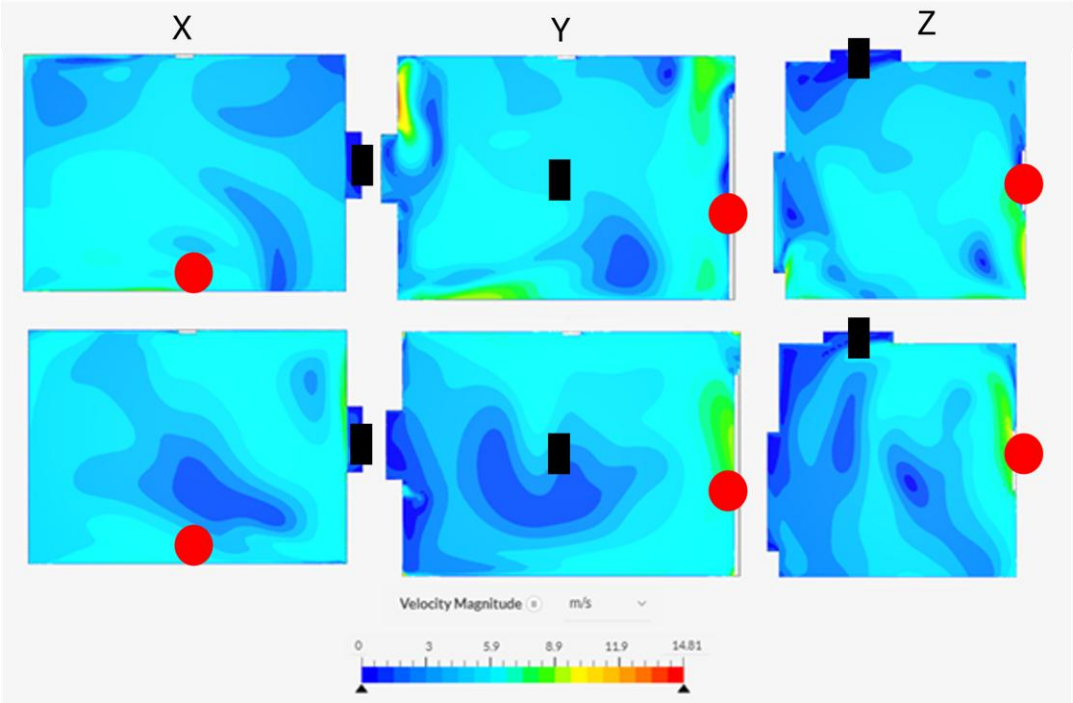

Figure 3. CFD simulations of air velocity. The first row shows the state of the room after 2500 s, and the second row after 5000 s, with the columns representing X, Y, and Z. Black rectangles indicate the window opening representing the primary ventilation boundary condition. For easier orientation, the red dot indicates the room's door.

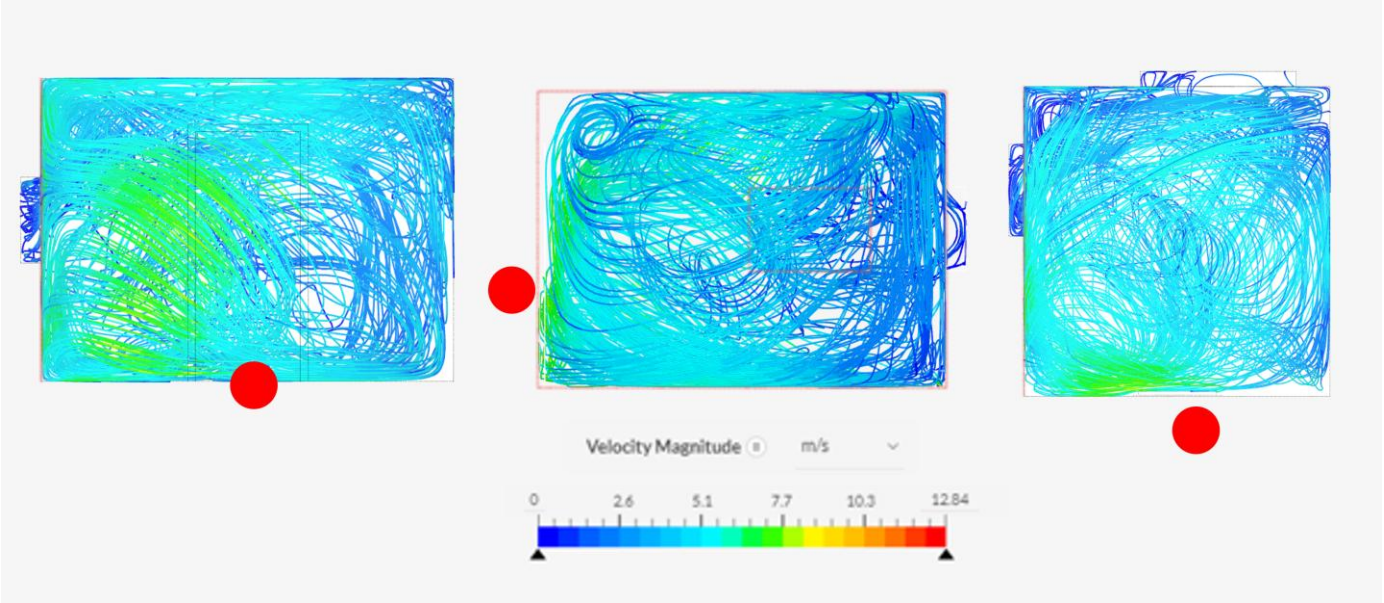

Figure 4. Particle movement during CFD simulations. For easier orientation, the red dot indicates the door to the room.

3.4. Machine Learning Model Results

The performance of all models was evaluated under two conditions of the air purifying unit: on and off. Tables 1 and 2 summarize the results of model predictions on the test dataset for each measured parameter. Model performance was assessed using the R^2,

RMSE, and *MAE*. When the air purifying unit was off (Table 1), the LR and XGBoost models consistently outperformed other approaches across most parameters. Both achieved high R^2 values, typically above 0.9, and the lowest *RMSE* and *MAE*. Deep learning models such as the LSTM and MLP exhibited higher errors and lower R^2 values, particularly for humidity and temperature, where the LSTM model produced large deviations ($R^2 < 0$). The results suggest that during periods of unfiltered air, the system dynamics were more complex, and the simpler regression models better captured the dominant linear relationships.

Table 1. Results of model predictions on the test set when the air purifying unit was off.

Parameter	Metric	LSTM	MLP	LR	RF	XGBoost
CO_2	R^2	−0.28	−0.82	0.99	0.68	0.99
	RMSE	88.26	105.22	0.01	44.35	5.53
	MAE	68.23	77.19	0.02	28.37	3.04
TVOC	R^2	0.52	0.97	0.99	0.74	0.93
	RMSE	60.57	16.05	0.10	44.73	23.14
	MAE	34.16	10.37	0.05	22.75	9.56
$PM_{2.5}$	R^2	0.64	0.98	0.99	0.99	0.99
	RMSE	6.18	1.56	0.03	0.45	0.66
	MAE	3.91	1.12	0.01	0.16	0.16
PM_{10}	R^2	0.61	0.97	0.98	0.99	0.99
	RMSE	6.57	1.7	0.56	0.31	0.85
	MAE	3.65	1.19	0.45	0.13	0.27
Temperature	R^2	0.59	−3.47	0.99	0.97	0.99
	RMSE	0.94	3.09	0.06	0.27	0.08
	MAE	0.71	2.25	0.02	0.15	0.05
Humidity	R^2	0.33	−11.63	0.99	0.98	0.99
	RMSE	1.41	6.14	0.03	0.27	0.05
	MAE	1.15	4.38	0.08	0.19	0.03
Atm Pressure	R^2	−62.16	−3.66	0.99	0.99	0.99
	RMSE	41.52	11.27	0.07	0.05	0.09
	MAE	13.56	7.75	0.3	0.02	0.03

When the air purifying unit was on (Table 2), all models showed improved predictive accuracy, with LR achieving the best overall performance across all measured parameters ($R^2 \approx 0.99$ and $RMSE < 0.1$ for most cases). The activation of the filtering device led to more stable environmental conditions, reflected in lower data variability and smaller prediction errors. Although RF and XGBoost maintained competitive performance, their results were slightly inferior to LR, indicating that simpler models can effectively generalize under steady-state condition

Figure 5 shows a comparison of measured and predicted values on the test dataset using the regression model that achieved the highest predictive performance across all monitored parameters, as evaluated by R^2, *RMSE*, and *MAE*. Mean hourly data are presented. The green-shaded regions represent periods when the air purifying unit was active, while the red-shaded regions correspond to periods when it was off. The predicted values for both conditions (filter on and filter off) closely follow the actual measurements, particularly during the filtering phases. The largest discrepancies were observed in CO_2 and TVOC concentrations during transient periods, likely caused by rapid changes in indoor air exchange and occupant activity.

Table 2. Results of model predictions on the test set when the air purifying unit was on.

Parameter	Metric	LSTM	MLP	LR	RF	XGBoost
CO_2	R^2	0.35	0.82	0.99	0.78	0.94
	RMSE	65.12	28.41	0.01	9.23	4.81
	MAE	47.83	15.64	0.02	5.33	3.02
TVOC	R^2	0.48	0.89	0.99	0.95	0.96
	RMSE	44.1	10.72	0.05	7.85	8.14
	MAE	24.52	6.14	0.03	4.25	5.08
$PM_{2.5}$	R^2	0.51	0.90	0.99	0.97	0.98
	RMSE	4.32	1.02	0.02	0.38	0.42
	MAE	2.95	0.88	0.01	0.14	0.16
PM_{10}	R^2	0.47	0.86	0.99	0.98	0.98
	RMSE	5.31	1.28	0.04	0.25	0.36
	MAE	2.85	0.93	0.02	0.11	0.18
Temperature	R^2	0.42	0.79	0.99	0.96	0.97
	RMSE	0.66	1.52	0.03	0.18	0.09
	MAE	0.49	1.10	0.01	0.12	0.05
Humidity	R^2	0.31	0.75	0.99	0.97	0.97
	RMSE	0.35	0.82	0.99	0.78	0.94
	MAE	65.12	28.41	0.01	9.23	4.81
Atm Pressure	R^2	47.83	15.64	0.02	5.33	3.02
	RMSE	0.48	0.89	0.99	0.95	0.96
	MAE	44.10	10.72	0.05	7.85	8.14

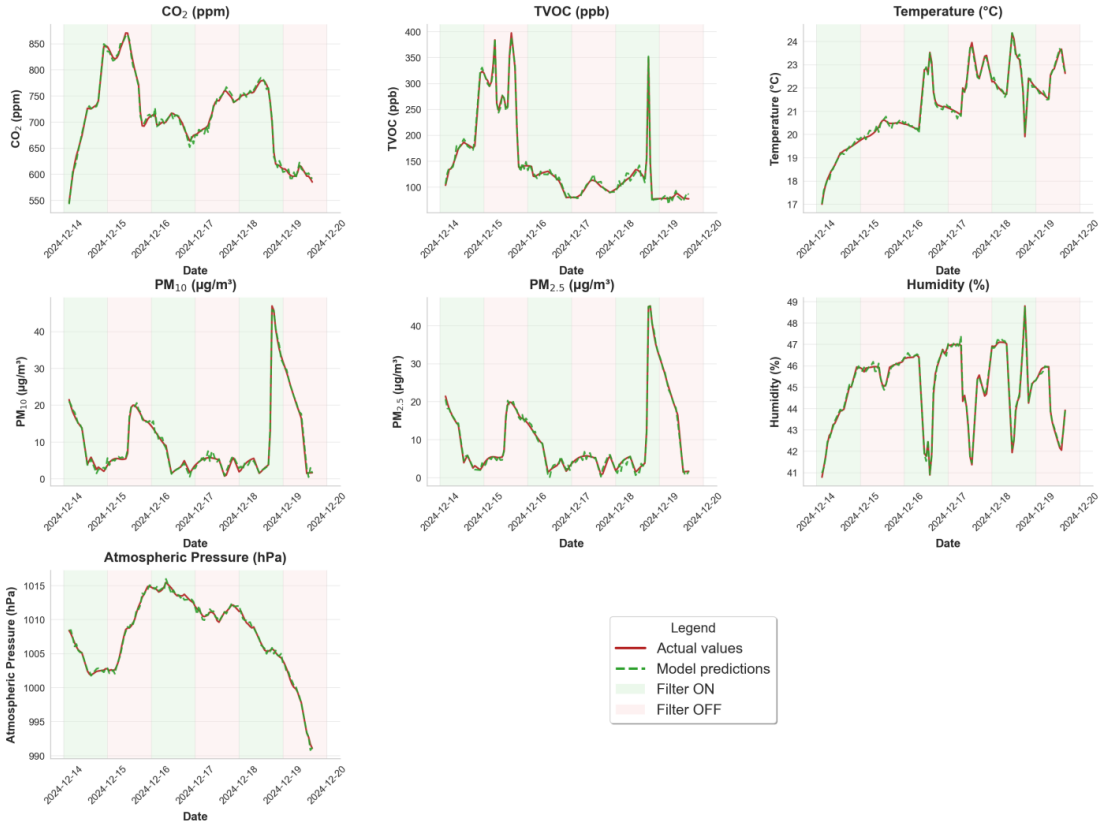

Figure 5. Comparison of measured and predicted parameters for air purifier ON and OFF conditions on the test set.

Overall, the results demonstrate that the LR model provides robust, stable predictions across both operational regimes. The strong performance of simpler and tree-based models under data-limited conditions observed in this study is consistent with previous IAQ prediction research, which has shown that LR and ensemble methods often outperform deep learning architectures when training data are limited or noisy [45,46]. The presence of the air purifying unit significantly improved the predictive accuracy across all metrics, confirming its stabilizing influence on IAQ parameters. Short-term increases in PM and TVOC concentrations during air filter operation coincide with periods of elevated outdoor pollution and increased natural ventilation, indicating that external pollutant ingress can temporarily exceed the filtering unit's removal efficiency.

3.5. Prophet Model Performance

The Prophet model was tested to assess the potential of classical time-series forecasting for IAQ prediction. However, the model's performance was suboptimal (Figure S7). The limited dataset size and the spatial uniformity of the experimental room led to highly correlated sensor readings, preventing the model from detecting meaningful temporal patterns. As a result, all sensors produced nearly identical forecasts with poor predictive accuracy. Therefore, Prophet was not included in the final DT framework.

3.6. System Architecture, Scalability, and Spatial Optimization

The proposed DT architecture is designed to operate in a cloud-based, distributed manner, enabling scalable deployment across multiple indoor environments. Sensor data acquisition, preprocessing, and model inference can be performed independently for each monitored space, while centralized cloud resources support data storage, model training, and visualization. This modular structure allows the system to scale from a single room to larger buildings or multi-room environments without fundamental changes to the modeling framework. The use of LCSs, lightweight physical models, and computationally efficient ML algorithms further supports scalability by minimizing hardware and computational requirements. In parallel, spatial optimization within the framework is informed by CFD-based analysis of airflow and particle transport patterns. The CFD results demonstrate that air purifier placement influences airflow organization, recirculation zones, and particle suspension, which in turn affects pollutant removal effectiveness. By comparing airflow structures and particle trajectories for different purifier locations, the simulations provide qualitative guidance on favorable placement strategies and support the interpretation of sensor measurements, highlighting the importance of spatial configuration in maximizing air purification performance.

3.7. Graphical User Interface and Health-Related Implications

The developed graphical user interface (GUI) provides an accessible front end to the DT, enabling users to interactively explore IAQ conditions, model predictions, and system states. The GUI supports intuitive visualization of measured and predicted parameters and allows users to compare scenarios, such as air purifier ON and OFF, facilitating practical interpretation without requiring technical expertise. While the current implementation focuses on near–real-time visualization and exploratory analysis rather than closed-loop control, the architecture supports future extensions toward real-time decision support and automated intervention. From a health perspective, the observed reductions in PM and stabilization of gaseous pollutant concentrations during air purifier operation indicate improved IAQ relative to baseline conditions. Although direct health outcomes were not measured, these improvements are relevant given established evidence linking lower PM and VOC exposure to reduced respiratory and cardiovascular health risks. The results, therefore, suggest potential health benefits associated with the proposed system through

improved exposure conditions. A formal health impact assessment or cost–benefit analysis was beyond the scope of this study; however, the combination of LCSs, cloud-based deployment, and computationally efficient models indicates a favorable cost-to-benefit ratio for scalable IAQ monitoring and management applications.

3.8. Limitations and Future Development

While the proposed DT demonstrates the feasibility and benefits of a hybrid physical–data-driven approach for IAQ monitoring, several limitations should be acknowledged. The experimental campaign was conducted in a single residential room under natural-ventilation conditions, which limits direct generalization to larger buildings or to mechanically ventilated environments. In addition, the study relied on manufacturer-calibrated LCS without long-term recalibration, and potential sensor drift was not explicitly compensated for over extended deployment periods. Future work will focus on the continuous development of the framework by integrating external data sources, such as meteorological data, outdoor air quality networks, and occupancy-related information, to improve predictive robustness and contextual awareness. Long-term deployments will enable the implementation of automated drift detection and compensation techniques, including periodic reference calibration, cross-sensor consistency checks, and physics-informed correction schemes. Scalability to other indoor environments, including multi-room apartments, offices, and public buildings, will be explored by extending the distributed cloud-based architecture and adapting the modeling framework to different ventilation systems and spatial configurations. These developments will further enhance the applicability of the proposed DT for real-world IAQ management.

4. Conclusions

This study developed and evaluated a hybrid DT framework for residential IAQ monitoring and analysis, demonstrating its functionality through a 26-day field deployment. The proposed system integrates LCS networks, physics-based models (mass balance and CFD), and ML techniques within a scalable, cloud-based architecture. Results show that, within the proposed framework, simpler ML models can outperform more complex architectures under data-limited conditions, with LR providing stable, accurate predictions across multiple IAQ pollutants as targets. Sensor measurements and statistical analysis indicate that operating the air-purifying unit is associated with lower median PM concentrations than baseline conditions, as shown in the boxplot, confirming its effectiveness in improving IAQ under the studied conditions. The hybrid modeling approach enhances interpretability and robustness by embedding physically meaningful information into data-driven components while maintaining computational efficiency suitable for near–real-time operation. CFD-based airflow and particle-trace simulations provide qualitative insight into pollutant transport and show that air purifier placement affects airflow patterns and particle removal efficiency, underscoring the importance of spatial configuration.

Overall, the proposed system addresses key limitations of existing IAQ solutions by combining dense sensing, physics-informed interpretation, and data-driven prediction within a unified DT framework. While the study is limited to a single residential room under natural ventilation, the results demonstrate the approach's potential as a decision-support tool for IAQ management. Future work will focus on multi-room scalability, long-term sensor drift compensation, integration with mechanical ventilation systems, incorporation of external data sources, and extended validation in larger and more diverse indoor environments.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/atmos17020184/s1: Figure S1. Flowchart of the study phases; Figure S2. Satellite and street view of the study location and photographs of the test site; Figure S3. Calculation of λ and λ_d; Figure S4. Mass Balance results for one sensor; Figure S5. GUI with tab CFD Simulation of Air Fluctuation; Figure S6. GUI with tab CFD Simulation of Participles in a Room; Figure S7. Results of the Prophet model on the test set; Table S1: Sensor placement in the room; Table S2: Sensor Accuracy of Wings devices; Table S3. Sensor Accuracy of Labservice Analytica devices; Table S4. List of all measured parameters, Table S5. Filtering unit placement in the room.

Author Contributions: Conceptualization, V.P., N.R., M.L. and D.S.; Data curation, V.P.; Formal analysis, V.P.; Methodology, V.P., N.R. and S.K.; Resources, M.L., D.S., F.M., T.H. and S.K.; Supervision, D.S., M.L., F.M. and T.H.; Visualization, V.P.; Writing—original draft, N.R. and V.P.; Writing—review and editing, T.H., S.K., F.M. and M.L. All authors have read and agreed to the published version of the manuscript.

Funding: V.P. received an academic license for SimScale (https://www.simscale.com/academic-program/ (accessed on 22 January 2026)). N.R. is supported by Next Generation EU under the EnvironPollutHealth project, Program Contract of 8 December 2023, Class: 643-02/23-01/00016, Reg. no. 533-03-23-0006. M.L. and N.R. are supported by the EU-Commission Grant Nr. 101217310—NextAIRE. M.L., F.M. and D.S. are supported by the EU-Commission Grant Nr. 101057497—EDIAQI.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Acknowledgments: The authors express their gratitude to Bruno Kraljinović (from company Ascalia d.o.o., Ul Trate 16, 40000, Čakovec) and creation of GUI. The authors also acknowledge SimScale for granting an academic license used in this study, granted to Valentino Petrić for his doctoral work.

Conflicts of Interest: Author Dejan Strbad was employed by the company Ascalia d.o.o. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IAQ	Indoor Air Quality
IAM	Indoor Aerosol Model
API	Application Programming Interface
DT	Digital Twin
CFD	Computational Fluid Dynamics
PM	Particulate Matter
PM_1	Particulate Matter with aerodynamic diameter $\leq$ 1 µm
$PM_{2.5}$	Particulate Matter with aerodynamic diameter $\leq$ 2.5 µm
PM_{10}	Particulate Matter with aerodynamic diameter $\leq$ 10 µm
VOC	Volatile Organic Compounds
TVOC	Total Volatile Organic Compounds
CO_2	Carbon Dioxide
CO	Carbon Monoxide
NO_2	Nitrogen Dioxide
NO	Nitric Oxide
O_3	Ozone
SO_2	Sulfur Dioxide
ML	Machine Learning
DL	Deep Learning

LCS	Low-Cost Sensors
LR	Linear Regression
RF	Random Forest
RH	Relative Humidity
SST	Shear Stress Transport
SI	Supplementary Information
LSTM	Long Short-Term Memory
MLP	Multi-Layer Perceptron
GUI	Graphical User Interface
HVAC	Heating, Ventilation, and Air Conditioning
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error

References

1. WHO Guidelines for Indoor Air Quality: Selected Pollutants. Available online: https://www.who.int/publications/i/item/9789289002134 (accessed on 10 December 2025).
2. Mannan, M.; Al-Ghamdi, S.G. Indoor Air Quality in Buildings: A Comprehensive Review on the Factors Influencing Air Pollution in Residential and Commercial Structure. *Int. J. Environ. Res. Public Health* **2021**, *18*, 3276. [CrossRef] [PubMed]
3. Saraga, D.E.; Querol, X.; Duarte, R.M.B.O.; Aquilina, N.J.; Canha, N.; Alvarez, E.G.; Jovasevic-Stojanovic, M.; Bekö, G.; Byčenkienė, S.; Kovacevic, R.; et al. Source Apportionment for Indoor Air Pollution: Current Challenges and Future Directions. *Sci. Total Environ.* **2023**, *900*, 165744. [CrossRef] [PubMed]
4. Martins, C.; Teófilo, V.; Clemente, M.; Corda, M.; Fermoso, J.; Aguado, A.; Rodriguez, S.; Moshammer, H.; Kristian, A.; Ferri, M.; et al. Sources, Levels, and Determinants of Indoor Air Pollutants in Europe: A Systematic Review. *Sci. Total Environ.* **2025**, *964*, 178574. [CrossRef] [PubMed]
5. Hussein, T.; Glytsos, T.; Ondráček, J.; Dohányosová, P.; Ždímal, V.; Hämeri, K.; Lazaridis, M.; Smolík, J.; Kulmala, M. Particle Size Characterization and Emission Rates during Indoor Activities in a House. *Atmos. Environ.* **2006**, *40*, 4285–4307. [CrossRef]
6. Račić, N.; Terzić, I.; Karlović, N.; Bošnjaković, A.; Terzić, T.; Jakovljević, I.; Pehnec, G.; Horvat, T.; Gajski, G.; Gerić, M.; et al. Volatile Organic Compounds (VOCs) and Polycyclic Aromatic Hydrocarbons (PAHs) in Indoor Environments: A Review and Analysis of Measured Concentrations in Europe. *Indoor Air* **2025**, *2025*, 5945455. [CrossRef]
7. Satish, U.; Mendell, M.J.; Shekhar, K.; Hotchi, T.; Sullivan, D.; Streufert, S.; Fisk, W.J. Is CO_2 an Indoor Pollutant? Direct Effects of Low-to-Moderate CO_2 Concentrations on Human Decision-Making Performance. *Environ. Health Perspect.* **2012**, *120*, 1671–1677. [CrossRef]
8. Anderson, J.O.; Thundiyil, J.G.; Stolbach, A. Clearing the Air: A Review of the Effects of Particulate Matter Air Pollution on Human Health. *J. Med. Toxicol.* **2012**, *8*, 166–175. [CrossRef]
9. Grahame, T.J.; Klemm, R.; Schlesinger, R.B. Public Health and Components of Particulate Matter: The Changing Assessment of Black Carbon. *J. Air Waste Manag. Assoc.* **2014**, *64*, 620–660. [CrossRef]
10. Pope, C.A.; Dockery, D.W. Health Effects of Fine Particulate Air Pollution: Lines That Connect. *J. Air Waste Manag. Assoc.* **2006**, *56*, 709–742. [CrossRef]
11. Jansen, K.L.; Larson, T.V.; Koenig, J.Q.; Mar, T.F.; Fields, C.; Stewart, J.; Lippmann, M. Associations between Health Effects and Particulate Matter and Black Carbon in Subjects with Respiratory Disease. *Environ. Health Perspect.* **2005**, *113*, 1741–1746. [CrossRef]
12. Badura, M.; Batog, P.; Drzeniecka-Osiadacz, A.; Modzel, P. Evaluation of Low-Cost Sensors for Ambient $PM_{2.5}$ Monitoring. *J. Sens.* **2018**, *2018*, 5096540. [CrossRef]
13. Castell, N.; Dauge, F.R.; Schneider, P.; Vogt, M.; Lerner, U.; Fishbain, B.; Broday, D.; Bartonova, A. Can Commercial Low-Cost Sensor Platforms Contribute to Air Quality Monitoring and Exposure Estimates? *Environ. Int.* **2017**, *99*, 293–302. [CrossRef] [PubMed]
14. Hayward, I.; Martin, N.A.; Ferracci, V.; Kazemimanesh, M.; Kumar, P. Low-Cost Air Quality Sensors: Biases, Corrections and Challenges in Their Comparability. *Atmosphere* **2024**, *15*, 1523. [CrossRef]
15. Kang, Y.; Aye, L.; Ngo, T.D.; Zhou, J. Performance Evaluation of Low-Cost Air Quality Sensors: A Review. *Sci. Total Environ.* **2022**, *818*, 151769. [CrossRef] [PubMed]
16. Zaidan, M.A.; Motlagh, N.H.; Fung, P.L.; Khalaf, A.S.; Matsumi, Y.; Ding, A.; Tarkoma, S.; Petaja, T.; Kulmala, M.; Hussein, T. Intelligent Air Pollution Sensors Calibration for Extreme Events and Drifts Monitoring. *IEEE Trans. Ind. Inf.* **2023**, *19*, 1366–1379. [CrossRef]

17. Attaran, M.; Celik, B.G. Digital Twin: Benefits, Use Cases, Challenges, and Opportunities. *Decis. Anal. J.* **2023**, *6*, 100165. [CrossRef]
18. Qi, Q.; Tao, F. Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison. *IEEE Access* **2018**, *6*, 3585–3593. [CrossRef]
19. Julien, N.; Martin, E. How to Characterize a Digital Twin: A Usage-Driven Classification. *IFAC-Pap.* **2021**, *54*, 894–899. [CrossRef]
20. Khajavi, S.H.; Motlagh, N.H.; Jaribion, A.; Werner, L.C.; Holmstrom, J. Digital Twin: Vision, Benefits, Boundaries, and Creation for Buildings. *IEEE Access* **2019**, *7*, 147406–147419. [CrossRef]
21. Digital Twin: Enabling Technologies, Challenges and Open Research. Available online: https://ieeexplore.ieee.org/document/9103025 (accessed on 27 January 2026).
22. Boje, C.; Guerriero, A.; Kubicki, S.; Rezgui, Y. Towards a Semantic Construction Digital Twin: Directions for Future Research. *Autom. Constr.* **2020**, *114*, 103179. [CrossRef]
23. Zhang, A.; Yu, Z.; Zhang, X.; Han, L. Advances in the Application of Digital Twins in Smart Cities. In *Proceedings of the 2024 Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Education Digitalization and Computer Science, Shen-zhen, China, 26–28 July 2024*; ACM: New York, NY, USA, 2024; pp. 249–254.
24. Hoffer, J.G.; Ofner, A.B.; Rohrhofer, F.M.; Lovrić, M.; Kern, R.; Lindstaedt, S.; Geiger, B.C. Theory-Inspired Machine Learning—towards a Synergy between Knowledge and Data. *Weld World* **2022**, *66*, 1291–1304. [CrossRef]
25. Razi, M.; Narayan, A.; Kirby, R.M.; Bedrov, D. Fast Predictive Models Based on Multi-Fidelity Sampling of Properties in Molecular Dynamics Simulations. *Comput. Mater. Sci.* **2018**, *152*, 125–133. [CrossRef]
26. Lovrić, M.; Meister, R.; Steck, T.; Fadljević, L.; Gerdenitsch, J.; Schuster, S.; Schiefermüller, L.; Lindstaedt, S.; Kern, R. Parasitic Resistance as a Predictor of Faulty Anodes in Electro Galvanizing: A Comparison of Machine Learning, Physical and Hybrid Models. *Adv. Model. Simul. Eng. Sci.* **2020**, *7*, 46. [CrossRef]
27. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations. *J. Comput. Phys.* **2019**, *378*, 686–707. [CrossRef]
28. Gudelj, I.; Lovrić, M.; Nyarko, E.K. Modelling the Daily Concentration of Airborne Particles Using 1D Convolutional Neural Networks. *Eng. Proc.* **2024**, *68*, 16. [CrossRef]
29. Kampezidou, S.I.; Ray, A.T.; Duncan, S.; Balchanos, M.G.; Mavris, D.N. Real-Time Occupancy Detection with Physics-Informed Pattern-Recognition Machines Based on Limited CO_2 and Temperature Sensors. *Energy Build.* **2021**, *242*, 110863. [CrossRef]
30. Nielsen, P.V. Fifty Years of CFD for Room Air Distribution. *Build. Environ.* **2015**, *91*, 78–90. [CrossRef]
31. Lovrić, M.; Petrić, V.; Strbad, D.; Terzić, T.; Frka, S.; Kušan, A.C.; Fermoso, J.; Düsing, S.; Alas, H.D.; Ladavac, M.D.; et al. Indoor and Ambient Air Pollution Dataset Using a Multi-Instrument Approach and Total Event Monitoring. *Sci. Data* **2025**, *12*, 1584. [CrossRef]
32. Hussein, T.; Kulmala, M. Indoor Aerosol Modeling: Basic Principles and Practical Applications. *Water Air Soil Pollut. Focus.* **2008**, *8*, 23–34. [CrossRef]
33. Hussein, T. Indoor-to-Outdoor Relationship of Aerosol Particles inside a Naturally Ventilated Apartment—A Comparison between Single-Parameter Analysis and Indoor Aerosol Model Simulation. *Sci. Total Environ.* **2017**, *596–597*, 321–330. [CrossRef]
34. De Jode, M. Long Term Monitoring of CO_2 Levels and Ventilation Rates in a Naturally Ventilated Residential Apartment. *Indoor Environ.* **2024**, *1*, 100030. [CrossRef]
35. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
36. Miao, S.; Gangolells, M.; Tejedor, B. Data-Driven Model for Predicting Indoor Air Quality and Thermal Comfort Levels in Naturally Ventilated Educational Buildings Using Easily Accessible Data for Schools. *J. Build. Eng.* **2023**, *80*, 108001. [CrossRef]
37. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv* **2019**, arXiv:1912.01703.
38. Taylor, S.J.; Letham, B. Forecasting at Scale. *Am. Stat.* **2018**, *72*, 37–45. [CrossRef]
39. Hightower, K.; Burns, B.; Beda, J. *Kubernetes, Up and Running: Dive into the Future of Infrastructure*, 1st ed.; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2017; ISBN 978-1-4919-3567-5.
40. Burns, B.; Grant, B.; Oppenheimer, D.; Brewer, E.; Wilkes, J. Borg, Omega, and Kubernetes. *Commun. ACM* **2016**, *59*, 50–57. [CrossRef]
41. Fisk, W.J.; Chan, W.R. Health Benefits and Costs of Filtration Interventions That Reduce Indoor Exposure to PM 2.5 during Wildfires. *Indoor Air* **2017**, *27*, 191–204. [CrossRef]
42. Aparicio-Ruiz, P.; Barbadilla-Martín, E.; Guadix, J.; Muñuzuri, J. A Field Study on Adaptive Thermal Comfort in Spanish Primary Classrooms during Summer Season. *Build. Environ.* **2021**, *203*, 108089. [CrossRef]
43. Zuo, W.; Chen, Q. Real-Time or Faster-than-Real-Time Simulation of Airflow in buildings. *Indoor Air* **2009**, *19*, 33–44. [CrossRef]
44. Nielsen, P.V. Computational Fluid Dynamics and Room Air Movement. *Indoor Air* **2004**, *14*, 134–143. [CrossRef]

45. Sun, Y.; Yang, T.; Gui, H.; Li, X.; Wang, W.; Duan, J.; Mao, S.; Yin, H.; Zhou, B.; Lang, J.; et al. Atmospheric Environment Monitoring Technology and Equipment in China: A Review and Outlook. *J. Environ. Sci.* **2023**, *123*, 41–53. [CrossRef]
46. Wei, G.; Yu, X.; Fang, L.; Wang, Q.; Tanaka, T.; Amano, K.; Yang, X. A Review and Comparison of the Indoor Air Quality Requirements in Selected Building Standards and Certifications. *Build. Environ.* **2022**, *226*, 109709. [CrossRef]