Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Physics-Informed Digital Twin Framework for Health Vigilance and Indoor Air Quality Decision Support in Healthcare-Oriented Environments

Ahmed Albashir

ahmed.a.albashir@academy.edu.ly

Libyan Academy for Postgraduate Studies

Research Article

Keywords: indoor air quality, digital twin, health vigilance, physics-informed machine learning, PM2.5 forecasting, decision support, uncertainty quantification

Posted Date: August 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10034990/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Physics-Informed Digital Twin Framework for Health Vigilance and Indoor Air Quality Decision Support in Healthcare-Oriented Environments

Ahmed Ali Albashir*[1,2]

[1]Department of Physics, School of Basic Sciences, Libyan Academy for Postgraduate Studies, Janzur, Libya

[2]Department of Physics, Faculty of Science, University of Tripoli, Tripoli, Libya

Abstract

Indoor air quality (IAQ) monitoring is important for health-vigilance-oriented indoor environments, yet conventional monitoring-only systems may not fully support proactive decision-making. This study proposes a physics-informed IAQ digital twin framework designed for short-horizon $PM_{2.5}$ forecasting and operational decision support under sensor-only constraints. The integrated methodology leverages routine IAQ sensor streams, temporal feature engineering, and an effective mass-balance interpretation layer. The framework evaluates multiple forecasting models alongside baseline approaches using walk-forward validation, uncertainty quantification, and spike-aware reliability assessment. An interactive Streamlit-based dashboard translates these components into a scenario simulation workflow. Empirical evaluations selected LightGBM at a 15-minute horizon as the deployed predictive engine, achieving an RMSE of 11.20 $\mu g/m^3$, MAE of 1.73 $\mu g/m^3$, MedianAE of 0.244 $\mu g/m^3$, and an R^2 of 0.170. While the 95% prediction intervals provided useful uncertainty information, empirical coverage remained below the nominal level, with a $PICP_{95}$ of 0.823. Spike-aware evaluation demonstrated useful moderate-threshold warning behavior, though severe $PM_{2.5}$ spike detection remained limited. The effective physics layer indicated source-dominated dynamics—with approximately 66.6% source contribution and 33.4% removal contribution—exhibiting strong TVOC-driven source behavior. This framework demonstrates how routine IAQ sensor streams can be transformed into an uncertainty-aware, physics-interpretable digital twin workflow for health-vigilance-oriented decision support, especially in settings where direct HVAC, air changes per hour (ACH), building information modeling (BIM), computational fluid dynamics (CFD), room-volume, or ventilation metadata are unavailable. These findings support the value of lightweight, sensor-based IAQ digital twins for proactive environmental decision support, while also identifying uncertainty calibration and severe-spike prediction as key directions for future work.

Keywords: indoor air quality; digital twin; health vigilance; physics-informed machine learning; $PM_{2.5}$ forecasting; decision support; uncertainty quantification.

*Email: ahmed.a.albashir@academy.edu.ly

1 Introduction

Indoor air quality (IAQ) is an important factor in public health and indoor environmental management, given that people spend a large proportion of their time indoors [1, 2]. In healthcare-oriented and sensitive indoor settings, IAQ management has emerged as a focal point for health-vigilance applications because indoor pollutant concentrations fluctuate rapidly due to occupant activity, episodic source emissions, and dynamic environmental parameters [3, 4]. Regularly monitored variables, such as fine particulate matter ($PM_{2.5}$) and carbon dioxide (CO_2), contribute to occupant exposure profiles and remain highly relevant to environmental risk characterization [4, 5]. Consequently, continuous IAQ monitoring provides the essential baseline data infrastructure required to build proactive, automated environmental intelligence frameworks [6, 7].

To capture these dynamic indoor microenvironments, contemporary IAQ management networks increasingly rely on distributed sensor architectures. While analytical reference-grade instruments are often impractical to deploy at scale across multi-zone facilities due to prohibitive capital and maintenance costs, distributed networks of low-cost sensors offer a scalable alternative for continuous spatial monitoring [4, 8]. However, a significant portion of current sensor-based IAQ deployments remains restricted to real-time telemetry, basic visualization, and static alerts [3]. While these basic frameworks establish essential immediate awareness, localized observation alone may not fully support proactive decision-making because it lacks a cohesive operational workflow linking raw data capture with multi-horizon forecasting, predictive uncertainty quantification, mechanistic physical interpretation, and interactive scenario simulation [1, 9].

To transition from passive environmental observation to short-horizon anticipation, researchers have integrated machine-learning and deep-learning algorithms into the IAQ domain. Tree-based ensemble methods and deep sequence models—such as Extreme Gradient Boosting (XGBoost), Random Forest, and Long Short-Term Memory (LSTM) networks—are deployed to capture complex non-linear interactions and temporal dependencies within multivariate environmental data streams [1, 6, 10]. Short-horizon forecasting provides predictive decision-support interfaces with early, actionable notifications regarding near-future concentration changes, enabling preventative intervention before air-quality conditions further deteriorate [5, 11]. Nevertheless, while data-driven configurations often perform well during stable baseline conditions, standard aggregate evaluation metrics, such as RMSE, MAE, and R^2, are structural averages that tend to obscure model failure modes during sudden, high-concentration $PM_{2.5}$ peaks or transient pollution events [3, 12].

Furthermore, although purely statistical forecasting models provide accurate short-term regression, they operate as opaque mathematical "black boxes" that lack the mechanistic transparency required in healthcare-oriented environmental decision-support contexts [13, 14]. High-fidelity physical simulation paradigms, including computational fluid dynamics (CFD) and BIM-supported physical modeling, deliver rich spatial and multi-phase transport insights but depend heavily on comprehensive, site-specific operational metadata—such as complex building geometries, detailed spatial meshes, air change rates (ACH), and HVAC engineering logs [15, 16]. Because these data-rich blueprints are rarely available in legacy or resource-constrained field settings, lightweight physics-informed learning layers present a pragmatic alternative [4, 17]. Specifically, an effective mass-balance layer extracts interpretable, room-level aggregate transport parameters under sensor-only constraints, providing a practical physical abstraction of the indoor system rather than attempting computationally demanding localized airflow field reconstruction

[6, 13].

The unified orchestration of these sensing and modeling modalities is increasingly achieved via digital twin (DT) frameworks, which couple continuous physical monitoring layers with virtual simulation engines and human-in-the-loop decision-support dashboards [7, 8]. Existing digital twins span a wide maturity spectrum, ranging from purely descriptive dashboards to autonomous, control-oriented cyber-physical systems. However, current environmental modeling implementations exhibit two persistent challenges regarding operational reliability. First, systematic uncertainty quantification receives less attention than deterministic point-forecast optimization, despite the practical importance of prediction intervals in decision-support settings and probabilistic calibration metrics like PICP or MPIW [1, 14]. Second, event-level reliability assessment is frequently omitted during evaluation; a model can maintain a low global error profile while systematically missing abrupt, high-concentration $PM_{2.5}$ events [3, 18]. Evaluating threshold-crossing sensitivity and peak transient retention is therefore important for assessing whether the virtual system remains relevant to occupant exposure dynamics and health-vigilance objectives [6, 19].

While low-cost IAQ sensing networks and isolated predictive modeling methods are maturing independently, there remains limited work on integrated, sensor-driven workflows that consolidate these components into an operational decision-support pipeline [4, 20]. The present study addresses this research gap by introducing a physics-informed IAQ digital twin framework that unifies multi-horizon $PM_{2.5}$ forecasting with effective mass-balance parameter identification, uncertainty-aware telemetry modeling, spike-aware classification evaluation, and dashboard-driven scenario simulation under sensor-only deployment conditions. By operating exclusively on routine IAQ sensor data streams without requiring direct HVAC logs, ACH measurements, BIM models, room-volume information, CFD-level spatial reconstruction, or structured ventilation metadata, the framework is designed to support health-vigilance-oriented deployment in sensor-only settings.

The remainder of this paper is structured as follows. Section 2 reviews relevant literature and existing methodologies. Section 3 details the proposed methodology, including data ingestion pipelines, physics-informed layers, and the forecasting architectures. Section 4 presents and discusses the empirical evaluation results. Section 5 summarizes the core contributions and outlines future research pathways.

2 Related Work

2.1 Sensor-based indoor air quality monitoring

The routine monitoring of indoor environments increasingly relies on distributed sensor networks to capture high-frequency environmental variables. While reference-grade instruments are often difficult to deploy at scale across multiple zones due to high capital and maintenance costs, distributed low-cost sensor networks provide a scalable alternative for multi-room indoor air quality (IAQ) monitoring [6, 8, 21, 22]. Typical sensor-based systems track a comprehensive range of contextual indicators, including $PM_{2.5}$, PM_{10}, CO_2, TVOC, ambient temperature, and relative humidity [3, 6, 20]. In standard implementations, the primary objective is real-time observation and static data collection, which provides facility managers with baseline environmental awareness through raw telemetry.

However, a significant portion of current sensor-based IAQ deployments remains focused on real-time visualization and basic alerting mechanisms, frequently lacking integrated predictive forecasting or dynamic, scenario-based decision support [1, 4, 9]. Although recent frameworks extend beyond passive observation

to support automated actuation—such as automated HVAC adjustments and mechanical filtration control [2, 4]—localized monitoring alone is insufficient for proactive decision support. Consequently, raw data collection must be integrated with advanced predictive models to optimize preventative indoor air management.

2.2 Machine-learning and deep-learning models for IAQ forecasting

To extend system functionality beyond passive monitoring, researchers have widely adopted machine-learning (ML) and deep-learning algorithms to forecast IAQ parameters. Tree-based ensemble methods and deep sequence architectures, such as XGBoost, Random Forest, and Long Short-Term Memory (LSTM) networks, are frequently applied to IAQ forecasting due to their capacity to capture complex non-linear relationships within multivariate sensor streams and explicit temporal dependencies [5, 6, 10, 13]. These data-driven models utilize historical lookback windows to learn underlying pollutant patterns, immediate health risks to vulnerable occupants in stable environments.

Nevertheless, standard performance evaluation protocols in IAQ forecasting typically emphasize global, macro-level accuracy metrics such as RMSE, MAE, and R^2. While these algorithms perform well during stable environmental regimes, this aggregate accuracy does not guarantee operational reliability during sudden, highly variable indoor pollution spikes. In this context, aggregate metrics provide limited insight into system reliability during abrupt, transient $PM_{2.5}$ peaks. This limitation highlights an important requirement for forecasting frameworks evaluated not only by global statistical errors but also by specific event-level behavior during transient air degradation episodes [3, 5, 12].

2.3 Physics-based and physics-informed IAQ modeling

While purely data-driven models are effective for short-horizon forecasting, they frequently operate as closed "black boxes" that lack physical interpretability. Traditional physical models, such as computational fluid dynamics (CFD) and grey-box thermal networks, provide deep diagnostic insights by explicitly modeling airflow and particle dispersion dynamics [4, 16]. However, these high-fidelity physical modeling techniques necessitate extensive structural and operational metadata—including complex building geometry, detailed computational meshes, specific air change rates (ACH), and HVAC engineering specifications—which are frequently unavailable in practical field settings [13, 15, 21, 23].

To bridge this gap, recent studies have explored hybrid and physics-informed learning strategies that integrate analytical physical laws into machine learning workflows [6, 14]. The integration of simplified physical principles, such as effective mass-balance equations, into data-driven frameworks provides a mechanistic interpretation layer under sensor-only constraints. This approach serves as a practical intermediate solution, balancing predictive utility and physical interpretability without the computational overhead of full CFD simulations [1, 4, 6, 13]. In this study, the proposed framework does not reconstruct detailed spatial airflow fields. Instead, it leverages an effective mass-balance layer to extract interpretable aggregate dynamics from available sensor streams, providing room-level interpretability as a valid physical abstraction rather than a complete replacement for detailed ventilation or CFD analysis.

2.4 Digital twins, dashboards, and decision-support systems

Digital twin (DT) architectures represent a dynamic convergence of continuous environmental monitoring, physical modeling, and interactive visualization. Digital twin implementations for IAQ vary significantly, spanning from basic monitoring dashboards to predictive, BIM-integrated, and control-oriented systems [7, 15, 24, 25]. While BIM-integrated and CFD-assisted digital twins are powerful tools for operational optimization, they depend heavily on highly structured metadata and modern building infrastructure.

Conversely, lightweight, sensor-based DT systems represent a more deployable alternative for older or resource-constrained healthcare environments where comprehensive architectural models are missing [3]. However, relatively little work has integrated multi-horizon forecasting, physical interpretation, uncertainty-aware assessment, and interactive "what-if" scenario simulation within a single decision-support workflow without requiring CFD-level spatial reconstruction or detailed structural metadata [1, 7, 15, 21]. This gap limits the utility of dashboards in complex, data-poor deployment contexts.

2.5 Uncertainty quantification and spike-aware reliability in IAQ forecasting

A rigorous assessment of forecast reliability is essential for health-vigilance-oriented applications. While deterministic point predictions remain the primary focus of most IAQ research, uncertainty quantification receives substantially less attention than point-forecast accuracy in contemporary studies [1, 14, 21]. Some recent investigations integrate probabilistic methods, such as Bayesian neural networks and bootstrap conformal prediction, to establish calibrated prediction intervals, utilizing metrics including PICP, MPIW, and CRPS to assess empirical coverage, sharpness, and probabilistic forecast quality [1, 14].

Alongside uncertainty, comprehensive evaluations must prioritize event-level reliability; classification metrics, such as precision, recall, or threshold-crossing behavior, are less commonly reported than global regression averages [1, 6, 13]. Indoor $PM_{2.5}$ concentrations are frequently characterized by discrete emission events rather than smooth background variation [3, 5]. Consequently, spike-aware evaluation is uniquely relevant for health-vigilance-oriented decision support, given that an algorithm can exhibit high average accuracy under ordinary regimes but still fail to detect the abrupt, high-concentration $PM_{2.5}$ peaks that present events that may strongly influence occupant exposure.

2.6 Research gap and positioning of the present study

Despite the proliferation of data-driven IAQ monitoring systems, current research remains largely fragmented, focusing on isolated components such as real-time visualization, point-prediction accuracy, or specialized physical modeling [1, 7, 15]. A significant limitation of existing high-fidelity physical approaches is their strict dependence on extensive structural and operational metadata, which are often unavailable in routine field deployments [4, 13, 21]. Furthermore, most studies prioritize aggregate global error metrics while neglecting systematic uncertainty quantification and event-level reliability during transient pollutant spikes [3, 5, 6]. of sensor-based frameworks that unify multi-horizon forecasting, uncertainty estimation, and scenario simulation in environments lacking direct metadata infrastructure.

The present study addresses this gap by proposing a physics-informed IAQ digital twin framework for health-vigilance-oriented decision support under sensor-only constraints. Instead of treating forecasting as an isolated regression task, the framework connects short-horizon prediction with effective physical interpretation, uncertainty-aware monitoring, spike-aware reliability assessment, and scenario-based

dashboard interaction. The following section details the proposed methodology, including the system architecture, the effective mass-balance layer, the forecasting models and baselines, and the temporal validation protocol used to evaluate the framework.

3 Materials and Methods

3.1 Study design and digital twin architecture

The proposed framework follows a modular digital twin structure designed to process raw environmental telemetry into actionable indoor air quality (IAQ) insights. As illustrated in Figure 1, the architecture consists of five functional layers that guide the data from initial acquisition to final decision support [20, 24]. Each layer executes a specific methodological function to maintain data integrity and physical consistency.

The first layer, the sensor layer, is responsible for collecting routine IAQ sensor streams. It acquires high-frequency readings of particulate matter ($PM_{2.5}$), gas-phase pollutants (CO_2 and TVOC), ambient sound levels, and general environmental conditions (temperature, relative humidity, and atmospheric pressure).

The second layer, the data-processing layer, focuses on cleaning, synchronizing, and resampling the data. Raw sensor timestamps are standardized and resampled to a uniform 5-minute frequency to align the disparate sensor streams. Following basic validation checks, the time-series data is transformed into sequences using a sliding window method. A lookback window of 12 steps (60 minutes) is applied to provide the necessary historical context for subsequent modeling [4].

The third layer, the feature-engineering layer, entails constructing temporal and dynamic sensor-based features. This includes cyclical temporal encoding (sine and cosine transformations) for diurnal and weekly patterns, autoregressive elements (one-step lags), and local statistical metrics (30-minute rolling means and standard deviations). Additionally, it generates dynamic proxies—such as an indoor activity index and a human emission proxy—derived directly from CO_2 variations and ambient sound levels.

The fourth layer, the modeling layer, is the computational core dedicated to calibrating an effective mass-balance layer and training forecasting models. It incorporates a comparative suite of algorithms, ranging from tree-based machine learning models (LightGBM, XGBoost, CatBoost) to sequence-aware deep learning architectures (LSTM, Temporal Fusion Transformer-inspired networks). An important component of this layer is the integration of physical equations, which extracts effective source and removal parameters directly from standard indoor telemetry to provide physically interpretable constraints [6, 17].

The final layer, the decision-support layer, translates these computational outputs into an interactive dashboard for monitoring, forecasting, uncertainty display, and scenario simulation. It generates short-horizon forecasts (15, 30, and 60 minutes) coupled with prediction intervals for uncertainty quantification. To evaluate operational reliability, the system implements a spike-aware evaluation protocol. The architecture was specifically designed to operate under sensor-only constraints, functioning effectively without requiring HVAC schedules, mechanical air change rate (ACH) logs, Building Information Modeling (BIM) frameworks, or room-volume metadata.

To ensure computational transparency and support multi-site reproducibility, the complete open-source software implementation—including the authoritative data pipelines, physical calibration scripts, trained model architectures, and the deployment-ready dashboard—has been compiled and permanently archived.

This research-grade code repository is publicly accessible via the Zenodo repository [26].

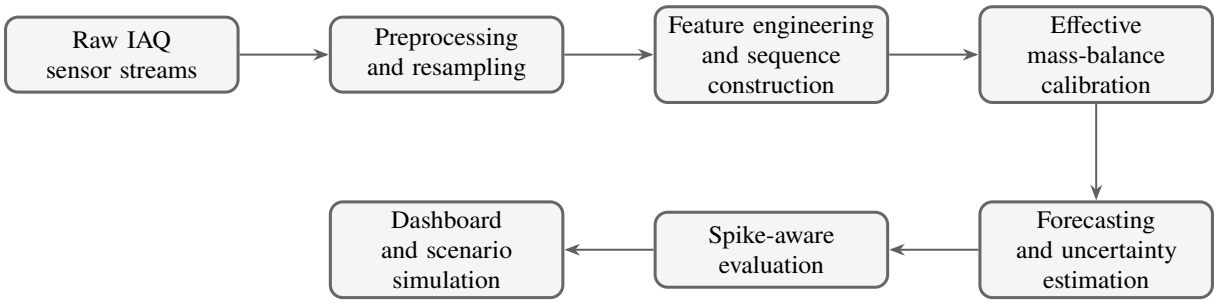


Figure 1: Overall digital twin workflow. The modular architecture converts routine IAQ sensor streams into a physics-informed decision-support workflow through preprocessing, feature engineering, effective mass-balance calibration, forecasting, uncertainty estimation, spike-aware evaluation, and dashboard-based scenario simulation.

3.2 Sensor data and preprocessing

The dataset comprises continuous environmental telemetry collected from two distinct indoor settings: a controlled laboratory and a one-room apartment. The raw sensor data (`laboratory.csv` and `one_room_apartement.csv`) was sourced from the "2023 Indoor Air Quality Dataset Germany" repository [27]. This dual-source setup provides empirical data representing different baseline conditions and typical residential occupancy patterns. The sensor array captures primary air quality and environmental indicators: particulate matter ($PM_{2.5}$), carbon dioxide (CO_2), total volatile organic compounds (TVOC), temperature, relative humidity, atmospheric pressure, and ambient sound levels.

To ensure temporal alignment across disparate sensor streams, a rigorous preprocessing pipeline was implemented. The specific steps were executed as follows:

- **Timestamp Standardization:** Raw timestamps were parsed into a standardized format and chronologically sorted to maintain the strict temporal order required for sliding-window sequence generation.
- **Synchronization and Resampling:** The independent sensor streams were synchronized and resampled to a uniform 5-minute frequency. This interval serves as the fundamental time step for both the physical mass-balance calibration and the predictive forecasting models.
- **Aggregation:** During the 5-minute resampling, continuous variables were aggregated using the mean of the observations within each window.
- **Missing Value Handling:** Missing values introduced by sensor downtime or the synchronization process were imputed using linear interpolation for brief gaps (under 30 minutes). Longer gaps were excluded from sequence generation to avoid fabricating extended temporal dynamics.
- **Unit Verification and Scaling:** All variables underwent unit harmonization. To facilitate gradient-based optimization in the neural network models, continuous features were normalized using Min-Max scaling. Critically, to prevent future information leakage, the scaling parameters were fitted exclusively on the training datasets and subsequently applied to the respective validation and test sets.

Within this workflow, zero and near-zero $PM_{2.5}$ readings are intentionally retained. Rather than applying strict filtering routines to eliminate these low measurements, they are preserved because they reflect the empirical sensor distribution associated with typical low-concentration indoor states. Retaining this data supports the mathematical identification of removal-dominated periods and captures the full decay curves of pollutants, which is required for estimating background concentration levels (C_{out}).

The presence of these near-zero values dictates the selection of evaluation metrics. Mean Absolute Percentage Error (MAPE) is explicitly excluded from the primary evaluation framework. Because percentage errors involve dividing by the observed value, MAPE becomes mathematically unstable and artificially inflates the error when evaluated on datasets containing values close to zero. Instead, the framework assesses model performance using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Median Absolute Error (MedianAE). These absolute and squared-error metrics are more stable for this zero-heavy distribution, providing a more robust measure of deterministic accuracy without the numerical distortions inherent to ratio-based metrics [6, 12].

3.3 Feature engineering under sensor-only constraints

In many practical indoor settings, direct measurements of mechanical ventilation rates (ACH) or building structural metadata are unavailable. To address this information gap, the proposed framework implements a feature engineering strategy that transforms raw sensor telemetry into indicators of unobserved or indirectly observed dynamics. The engineered features are categorized into five distinct groups, as summarized in Table 1, to provide the predictive models with temporal and causal context.

First, the raw IAQ sensor features comprise the observed environmental state at each 5-minute interval, including $PM_{2.5}$, CO_2, TVOC, temperature, humidity, atmospheric pressure, and ambient sound levels. Second, the system captures short-term memory by generating one-step lagged features for the primary pollutants. Third, rolling features, specifically 30-minute rolling means and standard deviations, are computed to capture local trends and environmental volatility. Fourth, cyclic temporal features encode diurnal and weekly patterns using sine and cosine transformations. This conversion ensures mathematical continuity between the end of one day and the start of the next, representing an established best practice in temporal forecasting [10].

Finally, the framework derives dynamic behavioral proxies. It is important to emphasize that these features do not directly measure ventilation, occupancy, or human emissions. Instead, they act as proxy indicators inferred from available sensor streams. These features provide proxy information about occupancy-related and source-related dynamics when direct HVAC, ACH, or BIM metadata are unavailable.

Specifically, the causal CO_2 derivative (dCO_2/dt) is computed using a backward finite difference method over the 5-minute time step to capture sudden shifts in indoor air composition [4]. Furthermore, the system constructs an Indoor Activity Index and a Human Emission Proxy (HEP). The Indoor Activity Index is generated using Principal Component Analysis (PCA) to extract the primary axis of variance from combined CO_2 and sound-level fluctuations. The HEP is formulated as a composite metric scaling ambient sound by the concurrent rate of CO_2 accumulation. To rigorously prevent data leakage, the PCA components and all scaling parameters for these dynamic proxy features were fitted strictly on the training folds and subsequently applied to the validation and test sets. Once derived, all dynamic proxy features were normalized alongside the continuous environmental variables prior to model ingestion.

Table 1: Feature groups and methodological role

Feature Group	Examples	Methodological Role
Raw IAQ Sensors	$PM_{2.5}$, CO_2, TVOC, Sound	Represents the observed environmental state at time t.
Lagged Features	$PM_{2.5}$ lag, CO_2 lag	Provides short-term memory of recent pollutant states.
Rolling Features	Rolling mean/std (30 min)	Captures local trends and environmental volatility.
Temporal Features	Hour/day cyclic features	Encodes daily and weekly human activity patterns.
Dynamic Proxies	dCO_2/dt, Activity Index, HEP	Provides proxy information under missing HVAC/BIM metadata by inferring unobserved or indirectly observed dynamics.

3.4 Effective mass-balance physics layer

The architecture incorporates an effective mass-balance layer designed to provide a physics-informed representation of $PM_{2.5}$ dynamics, as illustrated in Figure 2. This component represents an effective macro-level abstraction rather than a high-fidelity spatial airflow model. Consequently, it does not directly estimate room volume, absolute air change rates (ACH), direct ventilation rates, true HVAC mechanical performance, or computational airflow fields. Instead, the framework translates raw sensor fluctuations into physically meaningful proxy metrics.

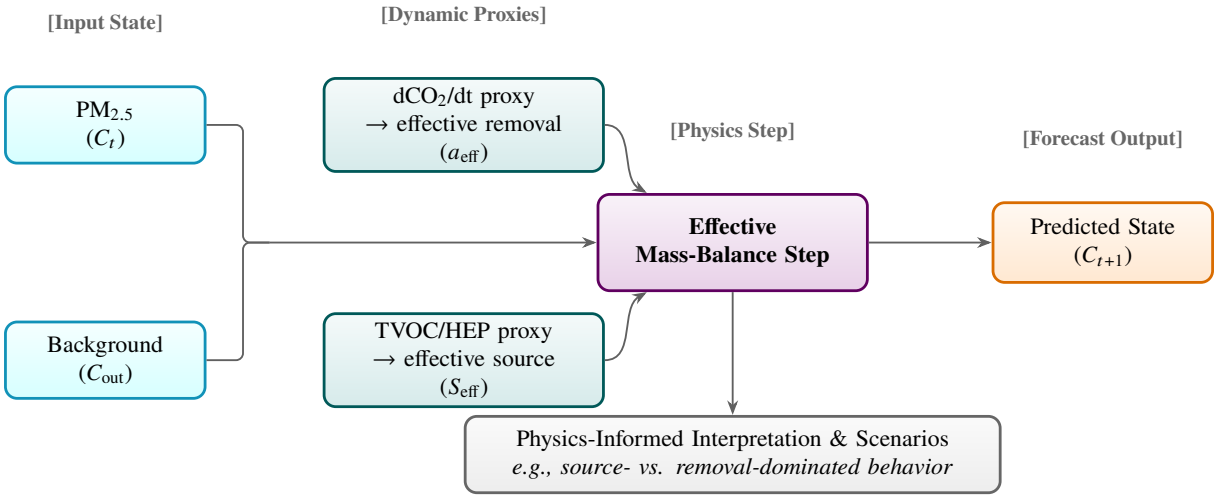


Figure 2: Effective mass-balance interpretation layer. The workflow integrates observed input states and dynamically derived behavioral proxies to parameterize a physics-informed mass-balance step, yielding both predicted future states and actionable operational interpretations without requiring structural building metadata.

The temporal evolution of $PM_{2.5}$ concentrations is modeled using a first-order differential equation, analytically solved over a 5-minute discrete time step (Δt):

$$C_{t+1} = C_t e^{-a_{\text{eff}}\Delta t} + \left(C_{\text{out}} + \frac{S_{\text{eff}}}{a_{\text{eff}} + \epsilon}\right)(1 - e^{-a_{\text{eff}}\Delta t}) \tag{1}$$

In this equation, C_t represents the current particulate concentration. The term C_{out} functions as an effective background concentration; because direct outdoor measurements were unavailable, this variable was estimated from the lowest stable baseline states of the indoor sensor stream. The terms a_{eff} and S_{eff} represent the effective removal coefficient and the effective source term, respectively, both of which are inferred dynamically from the available sensor telemetry. A small constant ϵ is added to ensure numerical stability.

In this methodology, a_{eff} and S_{eff} are treated as effective parameters rather than direct physical constraints derived from architectural metadata. These parameters were calibrated from the available sensor streams, while bootstrap resampling served as the primary method for parameter uncertainty estimation. Although Markov Chain Monte Carlo (MCMC) diagnostics were explored during development, they were not retained because they proved uninformative under highly non-stationary conditions. In the Physics-LSTM variant, the mass-balance equation provides a regularizing constraint to encourage physically consistent behavior during model training. In the deployed digital twin, this physical layer supports practical diagnosis—such as classifying the environment into source-dominated or removal-dominated operational regimes—and enables scenario-based decision support.

3.5 Forecasting models and baselines

To evaluate the predictive performance of the digital twin, the methodology implements a comparative modeling suite, as summarized in Table 2. This selection is designed to systematically test the capacity of the framework to handle non-linear dynamics and temporal dependencies across a short prediction horizon [5].

The suite includes gradient-boosted decision trees (GBDT), specifically LightGBM, XGBoost, and CatBoost. These algorithms serve as baseline tabular forecasting models capable of capturing complex, non-linear feature interactions without explicit sequence modeling [10]. To evaluate temporal representation learning, the system also implements deep sequence models utilizing a 60-minute lookback window (12 steps). These architectures include a standard Vanilla LSTM and a Temporal Fusion Transformer (TFT)-inspired network, which provide a comparative assessment. Additionally, a Physics-LSTM is evaluated to determine whether integrating the effective mass-balance equations directly into the network architecture improves predictive performance by enforcing physically plausible behavior through a physics-informed constraint.

Finally, the performance of these predictive models is benchmarked against strict operational baselines to quantify the value added beyond simple persistence or smoothing. The Persistence baseline assumes the future concentration equals the last observed value ($C_{t+1} = C_t$). The Rolling Mean baseline predicts future states using a 30-minute moving average. Furthermore, a Seasonal-TimeOnly baseline is employed as a control mechanism. This control model ignores pollutant telemetry and relies exclusively on cyclic temporal features, thereby verifying whether the primary models actively learn environmental dynamics or merely replicate daily and weekly periodic patterns.

3.6 Training, validation, and model selection protocol

To evaluate model performance on unseen future environmental states, the framework implements a strict temporal validation protocol. The system utilizes a walk-forward evaluation strategy, also known as nested time-series cross-validation, as illustrated in Figure 3. This approach preserves the chronological sequence

Table 2: Forecasting models and benchmark role

Model Class	Models	Role in Evaluation
Tree-based ML	LightGBM, XGBoost, CatBoost	Serve as strong tabular forecasting models for capturing non-linear feature interactions.
Deep Sequence Models	Vanilla LSTM, Physics-LSTM, TFT-inspired	Test temporal representation learning; Physics-LSTM tests the impact of an effective mass-balance constraint.
Operational Baselines	Persistence, Rolling Mean (30 min)	Quantify the value added beyond naive forecasting or simple signal smoothing.
Time-only Baseline	Seasonal-TimeOnly	Checks whether models are only learning daily or weekly periodic patterns.

of the sensor streams by partitioning the dataset into successive windows, where each test window strictly succeeds its corresponding training window. Consequently, this chronological partitioning simulates real-time operational conditions and prevents future information leakage [5].

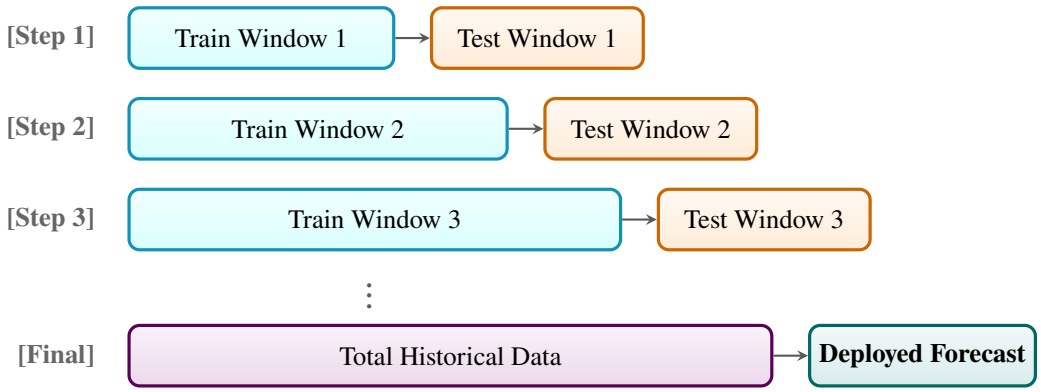


Figure 3: Walk-forward validation scheme. The dataset is chronologically partitioned into successive windows to prevent information leakage. This ensures that the protocol evaluates performance strictly on unseen future states, providing a robust measure of operational reliability.

Specifically, the validation scheme is executed using five outer walk-forward folds to evaluate out-of-sample performance. Within each outer training window, a nested, three-fold chronological split is utilized for hyperparameter optimization. The Optuna framework systematically searches the hyperparameter space for each architecture, using the validation Root Mean Square Error (RMSE) as the target optimization metric. To mitigate overfitting, early stopping criteria are applied across all architectures, halting the optimization process if the validation RMSE does not improve for a predefined number of consecutive iterations or epochs. Furthermore, to enforce strict data isolation, all normalization scalers, Principal Component Analysis (PCA) transformations, and data-derived preprocessing parameters are fitted exclusively on the active training windows before translation to the respective validation and test sets.

The framework generates forecasts across three prediction horizons: 15, 30, and 60 minutes. The final optimal configuration across these horizons is selected by balancing deterministic accuracy (RMSE and MAE) with probabilistic uncertainty calibration and real-time operational relevance. Based on these criteria, the 15-minute LightGBM configuration was selected for deployment because it minimized

the 15-minute prediction RMSE while maintaining low computational latency for real-time dashboard visualization.

3.7 Uncertainty quantification

To account for the inherent noise in low-cost sensor data and the approximations within the effective mass-balance framework, the methodology integrates a probabilistic estimation layer. Instead of providing only a deterministic point estimate, the system generates an uncertainty-aware forecast that quantifies the confidence level associated with each $PM_{2.5}$ prediction. Because the modeling suite encompasses fundamentally different architectures, uncertainty is estimated using two distinct approaches tailored to neural networks and tree-based models, respectively.

For neural network architectures (Vanilla LSTM, Physics-LSTM, and TFT-inspired models), the primary method for generating predictive distributions is Monte Carlo (MC) Dropout. During the inference phase, the dropout layers remain active to perform multiple stochastic forward passes ($N = 10$) for each time step. The mean of these passes serves as the point forecast ($\hat{y}$), while the standard deviation (σ_y) quantifies the empirical model uncertainty. Under a standard normal distribution assumption at a 95% nominal confidence level, the prediction intervals for the neural models are mathematically defined as follows:

$$\text{Lower Bound} = \max(\hat{y} - z \cdot \sigma_y, 0) \tag{2}$$

$$\text{Upper Bound} = \hat{y} + z \cdot \sigma_y \tag{3}$$

where $z \approx 1.96$ represents the important value for 95% confidence. The explicit maximum function applied to the lower bound ensures that predicted pollutant concentrations never drop below zero, maintaining physical feasibility.

In contrast, tree-based models—including the deployed LightGBM configuration—yield deterministic point forecasts rather than native predictive distributions in standard implementations. Therefore, prediction intervals for these models are generated using empirical residual quantiles. Historical out-of-fold validation residuals are aggregated to compute the 2.5th and 97.5th percentiles of the error distribution. These empirical margins are subsequently applied to the test-set point forecasts to establish the 95% bounds, with the lower bound similarly constrained at zero. Consequently, because deterministic operational baselines (such as Persistence and Rolling Mean) do not yield calibrated predictive distributions, interval metrics and the Continuous Ranked Probability Score (CRPS) are not reported for those control models.

The quality and practical utility of these generated intervals are assessed using three primary metrics. $PICP_{95}$ and $MPIW_{95}$ evaluate empirical interval coverage and practical interval width, respectively. Specifically, $PICP_{95}$ measures the proportion of actual ground-truth observations falling within the predicted bounds, while $MPIW_{95}$ quantifies the average spread of the intervals. Additionally, CRPS evaluates both the accuracy and the calibration of the entire predictive distribution, effectively penalizing both excessively wide intervals and poor point-forecast accuracy.

Providing probabilistic confidence intervals aligns with emerging trends in building environmental characterization [14]. However, due to the complex and non-stationary nature of indoor air dynamics, empirical coverage can occasionally deviate from the nominal 95% level. In this context, prediction intervals are used as uncertainty-aware diagnostic indicators rather than as perfectly calibrated statistical

guarantees.

3.8 Spike-aware evaluation

Standard evaluation metrics, such as RMSE and MAE, provide a useful global summary of model accuracy but can mask a system's failure to capture sudden, high-concentration indoor air quality events. To address this limitation, the framework incorporates a spike-aware evaluation strategy. While many contemporary methodologies focus primarily on global predictive accuracy and often filter out transient peaks to smooth stochastic variability [6], this methodology explicitly evaluates model reliability during sudden pollutant spikes.

The evaluation categorizes $PM_{2.5}$ concentrations using two operational event thresholds: a moderate threshold (15 $\mu g/m^3$) to detect early air quality degradation, and a severe spike threshold (35 $\mu g/m^3$) representing elevated short-term conditions. At these designated levels, the continuous deterministic forecast is re-evaluated as a binary detection task. Performance is measured using two standard classification metrics. Recall measures the fraction of true physical events successfully detected by the model, which quantifies sensitivity. In contrast, Precision measures the fraction of model-predicted events that correspond to actual physical events, serving as a verification mechanism against excessive false alarms.

This event-level analysis serves as a secondary reliability test and diagnostic tool rather than the primary model-training objective. Consequently, this multi-tiered approach distinguishes ordinary forecasting performance from event-level reliability during abrupt indoor pollution episodes.

3.9 Evaluation metrics

To provide a comprehensive assessment of the digital twin, the methodology employs a multi-dimensional evaluation suite, as summarized in Table 3. This framework assesses deterministic accuracy, probabilistic uncertainty, and operational sensitivity to elevated pollutant events.

For deterministic point-forecast accuracy, the framework utilizes Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). These metrics represent established standards in indoor air quality (IAQ) forecasting [4, 17, 28]. Additionally, Median Absolute Error (MedianAE) is included to assess model performance during stable operational regimes, given its resilience to extreme, transient outliers [12]. Notably, Mean Absolute Percentage Error (MAPE) is explicitly excluded from the headline metrics. Because zero and near-zero $PM_{2.5}$ values are intentionally retained within the dataset, MAPE becomes mathematically unstable and artificially inflates the reported error.

For probabilistic diagnostics, uncertainty is evaluated using Prediction Interval Coverage Probability ($PICP_{95}$) and Mean Prediction Interval Width ($MPIW_{95}$). Rather than assuming absolute statistical guarantees, these metrics verify whether the generated intervals provide valid empirical coverage and practical sharpness [1, 29]. To provide a comprehensive probabilistic evaluation, the Continuous Ranked Probability Score (CRPS) is also implemented, which simultaneously accounts for point-forecast accuracy and the calibration of the overall predictive distribution [14]. Probabilistic metrics are reported exclusively when a predictive distribution or calibrated interval is available.

Finally, to measure operational reliability, the framework incorporates a spike-aware evaluation module. While standard continuous regression metrics quantify global performance [5, 10], this methodology evaluates the forecasting task as a binary detection problem at specific operational thresholds (15 and

35 $\mu g/m^3$). By calculating Recall and Precision at these operational boundaries, the framework quantifies model sensitivity and prediction reliability during sudden pollution events [8, 9].

Table 3: Evaluation metrics and interpretation

Metric	**Purpose**	**Methodological Interpretation**
RMSE	Sensitivity to large errors	Penalizes large errors and is sensitive to spikes.
MAE	Average absolute error	Measures average absolute error.
MedianAE	Ordinary regime error	Describes typical error and is less sensitive to rare large residuals.
R^2	Explained variance	Summarizes explained variance but may be low in spike-dominated datasets.
$PICP_{95}$	Interval coverage	Measures empirical interval coverage.
$MPIW_{95}$	Interval width	Measures average interval width.
CRPS	Probabilistic quality	Evaluates probabilistic forecast quality when a predictive distribution is available.
Recall@15/35	Detection sensitivity	Measures event detection sensitivity.
Precision@15/35	Detection reliability	Measures event prediction reliability.

3.10 Dashboard and scenario simulation

The final layer of the architecture comprises a research-grade Streamlit dashboard prototype. This interface transforms the underlying computational models into an interactive decision-support system, translating predictive outputs into real-time IAQ monitoring, forecasting, uncertainty-aware interpretation, effective physics diagnostics, and scenario simulation. The development of interactive dashboards for monitoring IAQ trends represents an expanding paradigm in smart building management [20, 30]; however, this tool is explicitly designed as a research and decision-support prototype rather than a regulatory or clinical system.

The dashboard provides continuous IAQ monitoring alongside a 15-minute $PM_{2.5}$ forecast, complete with an interval display to communicate empirical model uncertainty [14]. To maintain operational transparency, the interface features a dynamic model leaderboard that ranks the deployed algorithm against the baseline models [25]. Additionally, a physics diagnostic module characterizes the immediate indoor environment via a source/removal interpretation, classifying the current operational state as either "source-dominated" or "removal-dominated." While alternative approaches utilize source terms to model occupancy dynamics [31], explicitly exposing these effective mass-balance ratios as a real-time diagnostic indicator provides enhanced interpretability relative to standard black-box visualizations.

The core operational feature of the dashboard is its scenario simulation engine. This module leverages the analytical mass-balance formulation to project how $PM_{2.5}$ concentrations would evolve under hypothetical interventions. Users can interactively adjust effective parameters, including effective removal (a_{eff}), effective source (S_{eff}), and background concentration (C_{out}). This flexibility enables the evaluation of specific conditions, as detailed in Table 4, including ventilation deterioration, air purification, indoor emission surges, external ingress, and combined mitigation strategies. While advanced digital twins often rely on spatially resolved computational fluid dynamics (CFD) [4] or city-scale partial differential

equation (PDE) solvers [6, 13], this framework prioritizes the interactive manipulation of inferred physical parameters. This transition from a static predictive model to a simulation-driven interface provides scenario-based, physics-informed IAQ decision support. Consequently, this design facilitates low-cost, sensor-based IAQ vigilance in healthcare-oriented and resource-constrained environments where direct HVAC metadata, Building Information Modeling (BIM) records, or physical ventilation measurements are unavailable.

Table 4: Scenario simulation parameters and decision-support purpose

Scenario	**Changed Component**	**Decision-Support Purpose**
Ventilation Deterioration	Effective removal (a_{eff}) decreases	Stress-tests the impact of ventilation failure or window closure on pollutant accumulation.
Air Purification	Effective removal (a_{eff}) increases	Simulates the introduction of mechanical filtration or increased ACH to accelerate decay.
Human/Source Surge	Effective source (S_{eff}) increases	Simulates indoor emission events, such as cooking, cleaning, or localized occupancy spikes.
External Ingress	Background level (C_{out}) increases	Predicts indoor concentration changes driven by rising outdoor pollution or infiltration.
Combined Mitigation	Effective source decreases and effective removal increases	Evaluates a combined source-control and removal-improvement scenario.

4 Results and Discussion

4.1 Data readiness and feature construction

The scientific motivation for the proposed digital twin framework relies on establishing a non-intrusive, sensor-based indoor air quality (IAQ) monitoring architecture. Consequently, the dataset was transformed into a sensor-based modeling structure designed to support short-horizon operational decision-making. Table 5 summarizes the monitored sites, sensor streams, forecasting horizons, and engineered features utilized in this workflow. The framework incorporates 19 features, a 12-step lookback window representing 60 minutes of historical context, and three primary prediction horizons at 15, 30, and 60 minutes. This feature set integrates raw sensor variables, lagged variables, rolling statistics, temporal features, and causal proxies, including causal CO_2 change, an indoor activity index, and a human emission proxy. As a result, the proposed workflow can be implemented without direct HVAC schedules, air changes per hour (ACH) logs, building information modeling (BIM), or room-volume metadata in the present sensor-only setting.

Regarding the underlying data distribution, the dataset preserves the natural variations of indoor particulate matter across heterogeneous contexts. The laboratory site presents a high fraction of zero $PM_{2.5}$ readings, whereas the apartment site exhibits a lower fraction. These zero and near-zero values were retained as part of the observed sensor distribution. This structural sparsity dictates the evaluation strategy; specifically, ordinary Mean Absolute Percentage Error (MAPE) is omitted as a headline metric

because percentage errors become unstable and artificially inflated when target values approach zero. In this context, the constructed data matrix provides a suitable foundation for short-horizon IAQ forecasting models.

By utilizing routine sensor data, this feature engineering strategy demonstrates that a functional digital twin can be established under practical constraints. Furthermore, retaining the zero and near-zero $PM_{2.5}$ readings ensures that the framework accurately represents true exposure conditions, which supports health vigilance by preserving authentic indoor air quality baselines.

4.2 Forecasting performance

Table 6 presents the empirical results across the 15, 30, and 60-minute prediction horizons. Because the 15-minute horizon represents the primary target for real-time proactive adjustments, the comparative analysis focuses heavily on this interval. LightGBM achieved the lowest 15-min RMSE, while the other tree-based models remained closely competitive. These models consistently demonstrated stable performance across the evaluated metrics. Consequently, LightGBM at the 15-minute horizon was selected as the deployed predictive model for the digital twin architecture.

In contrast, simple baselines such as RollingMean-30min and Persistence-LastValue may show low Median Absolute Error (MedianAE) values because many observations fall in stable or low-concentration regimes. However, these baselines are less suitable as deployed predictive engines because they have weaker aggregate reliability and limited uncertainty representation compared to the machine learning implementations.

Figure 4 illustrates the performance characteristics of the selected 15-minute LightGBM model. The observed-versus-predicted correlation shows good ordinary low-to-moderate tracking, while high $PM_{2.5}$ spikes are underestimated relative to the ideal line $y = x$. The residual distribution profile centers around zero, while the prediction interval bounds reflect the capacity for uncertainty-aware forecasting. Finally, the internal leaderboard supports the optimal ranking of the 15-minute model. These findings suggest that the system is operationally useful for short-horizon tracking, although it is not a universal extreme-spike predictor.

The stable performance of tree-based algorithms highlights their capability to map the non-linear relationships between diverse sensor streams. While the model underestimates extreme $PM_{2.5}$ spikes, this behavior aligns with the practical goals of short-horizon decision support. Rather than providing absolute peak prediction, the predictive engine serves as an early-warning-oriented decision support tool, linking available sensor data to proactive management steps.

4.3 Uncertainty and spike-aware reliability

Evaluating prediction uncertainty and response reliability under transient conditions is necessary to determine the operational limits of the framework. As shown in Table 6, the uncertainty metrics for the selected 15-minute LightGBM model include $PICP_{95}$, $MPIW_{95}$, and CRPS. The prediction intervals provide useful uncertainty information, although empirical coverage remains below the nominal 95% level. This variance indicates the difficulty of bounding rapid indoor air-quality transients through standard interval generation methods.

Figure 5 details the error distribution across distinct operational regimes. Residual magnitudes scale proportionally with observed $PM_{2.5}$ concentrations, showing that the median absolute error remains low

Table 5: Dataset and feature-engineering summary. The proposed framework is built from routine, non-intrusive IAQ sensor streams, enabling a physics-informed digital twin workflow without requiring direct HVAC schedules, ACH logs, BIM models, or room-volume metadata.

Component	**Value**	**Relevance to the IAQ digital twin**
Number of monitored sites	2	Provides two distinct indoor monitoring contexts for evaluating the proposed workflow.
Site identifiers	`laboratory`; `one_room_apartement`	Captures heterogeneous indoor sensing conditions and pollutant profiles.
Sampling interval	5.0 min	Supports short-horizon tracking of indoor air-quality dynamics.
Lookback window	12 steps, equivalent to 60 min	Provides recent temporal context for pollutant accumulation, decay, and short-term variation.
Forecasting horizons	15 min; 30 min; 60 min	Supports multi-horizon decision support for proactive IAQ management.
Target variable	`pm2_5`	Represents the primary particulate indicator used for health-vigilance-oriented forecasting.
Number of model features	19	Combines raw sensor streams, lagged variables, rolling statistics, and temporal descriptors.
Main sensor streams	`pm2_5`, `co2`, `tvoc`, `sound`, `temperature`, `humidity`, `pressure`	Defines the observable sensor-only state used by the digital twin without requiring building-management-system metadata.
Engineered temporal features	`hour_sin`, `hour_cos`, `dayofweek_sin`, `dayofweek_cos`, `rolling_mean_30min`, `rolling_std_30min`	Encodes short-term trends and recurring temporal patterns relevant to IAQ variation.
Engineered causal/dynamic features	`dCO2dt_causal`, `Indoor_Activity_Index`, `human_emission_proxy`, `pm25_lag1`, `co2_lag1`, `tvoc_lag1`	Provides proxy information about recent occupancy-related and source-related dynamics inferred from available sensors.
Laboratory $PM_{2.5}$ zero fraction	47.89%	Indicates sparse or event-driven particulate behavior in the laboratory site.
Apartment $PM_{2.5}$ zero fraction	4.06%	Indicates more persistent non-zero particulate readings in the residential site.
$PM_{2.5}$ 99.5th percentile by site	Laboratory: 26.04; apartment: 134.58	Highlights site-dependent upper-tail particulate behavior relevant to spike-aware evaluation.
TVOC 99.5th percentile by site	Laboratory: 1149.47; apartment: 22632.75	Indicates substantial site-dependent chemical-emission variability captured by the sensor streams.
Data-quality note	Zero and near-zero $PM_{2.5}$ values retained	Preserves the observed sensor distribution and avoids artificially removing low-concentration states.

Note: Zero and near-zero $PM_{2.5}$ values were retained as part of the observed sensor distribution rather than treated as missing data. This motivates avoiding ordinary MAPE as a headline metric, since percentage errors can be unstable when the denominator approaches zero.

Table 6: Main forecasting benchmark across prediction horizons. The table reports aggregate walk-forward evaluation results for selected forecasting models and operational baselines. LightGBM at the 15-minute horizon was selected as the deployed predictive engine.

Horizon	**Model**	**RMSE** (μg/m^3)	**MAE** (μg/m^3)	**MedianAE** (μg/m^3)	R^2	**PICP**$_{95}$	**MPIW**$_{95}$ (μg/m^3)	**CRPS**
15 min	LightGBM†	**11.20**	1.73	0.244	0.170	0.823	4.10	1.83
	XGBoost	11.21	1.89	0.184	0.169	0.806	5.60	2.05
	CatBoost	11.21	1.51	0.230	0.168	0.817	5.08	1.70
	Physics-LSTM	11.27	2.35	0.901	0.159	0.796	5.36	2.15
	RollingMean-30min	11.62	1.10	0.085	0.106	–	–	–
	Persistence-LastValue	12.92	0.83	0.077	-0.105	–	–	–
30 min	LightGBM	11.48	1.95	0.319	0.128	0.825	5.40	2.07
	XGBoost	**11.43**	1.90	0.227	0.135	0.839	5.85	2.03
	CatBoost	11.48	1.64	0.310	0.128	0.824	5.64	1.82
	Physics-LSTM	11.54	2.23	0.709	0.119	0.826	6.33	2.22
	RollingMean-30min	12.23	1.28	0.100	0.010	–	–	–
	Persistence-LastValue	14.18	1.08	0.085	-0.330	–	–	–
60 min	LightGBM	11.79	2.20	0.450	0.081	0.818	6.30	2.26
	XGBoost	**11.69**	2.04	0.363	0.097	0.833	6.61	2.14
	CatBoost	11.74	1.88	0.360	0.089	0.819	6.24	2.04
	Physics-LSTM	12.70	2.69	0.971	-0.067	0.837	7.10	2.61
	RollingMean-30min	12.93	1.55	0.123	-0.106	–	–	–
	Persistence-LastValue	15.10	1.42	0.119	-0.508	–	–	–

† Deployed predictive engine.

Notes: Best RMSE values within each horizon are shown in bold. RMSE is sensitive to rare abrupt $PM_{2.5}$ spikes, whereas MedianAE better reflects typical low-to-moderate concentration tracking. MAPE is not used as a headline metric because of the high fraction of zero or near-zero $PM_{2.5}$ observations. Probabilistic metrics are reported only for models with calibrated predictive intervals or predictive distributions; they are not reported for deterministic baselines without calibrated uncertainty estimates. Baseline models may show low MedianAE under low-concentration regimes but provide weaker aggregate reliability.

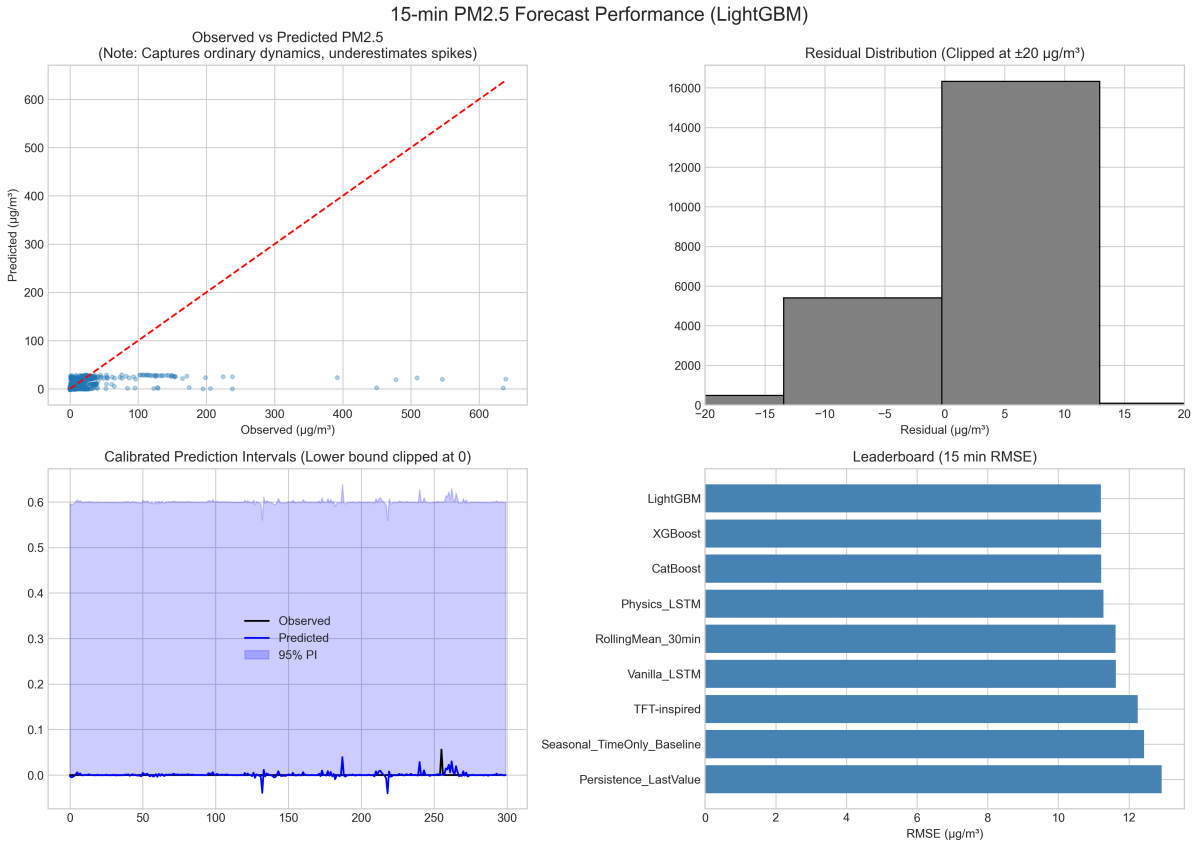


Figure 4: Fifteen-minute $PM_{2.5}$ forecasting performance of the deployed LightGBM model. The observed-versus-predicted panel compares measured and predicted $PM_{2.5}$ concentrations, with the red dashed line representing the ideal prediction line $y = x$. The concentration of points near the low-value region indicates that the model captures ordinary low-to-moderate IAQ dynamics, while the strong deviation from the ideal line at high observed concentrations shows systematic underestimation of rare $PM_{2.5}$ spikes. The residual-distribution panel summarizes the central error behavior after clipping residuals at $\pm 20\ \mu g/m^3$ for visualization. The calibrated prediction-interval panel illustrates uncertainty-aware short-horizon forecasting, and the leaderboard supports that LightGBM achieved the lowest 15-min RMSE among the evaluated models. The prediction-interval panel is shown on the normalized model scale for visual clarity.

within ordinary low-to-moderate regimes but increases sharply during high-concentration spike events. Concurrently, the prediction interval widths vary dynamically with the forecast behavior. Table 7 indicates that spike detection performance is stronger at the moderate threshold (15 μg/m^3) than at the severe threshold (35 μg/m^3). The results support moderate-threshold warning, while severe spike detection remains limited. These constraints reinforce the utility of the framework as a research-grade short-horizon decision-support tool rather than an absolute event detector.

Integrating uncertainty metrics directly into the digital twin elevates it from a standard predictive tool to a risk-assessment framework. The observation that empirical interval coverage falls below the nominal 95% level reflects the abrupt and localized nature of indoor emission events. Nevertheless, the system may provide useful warning signals for moderate threshold crossings (15 μg/m^3), directly supporting the core objective of health vigilance.

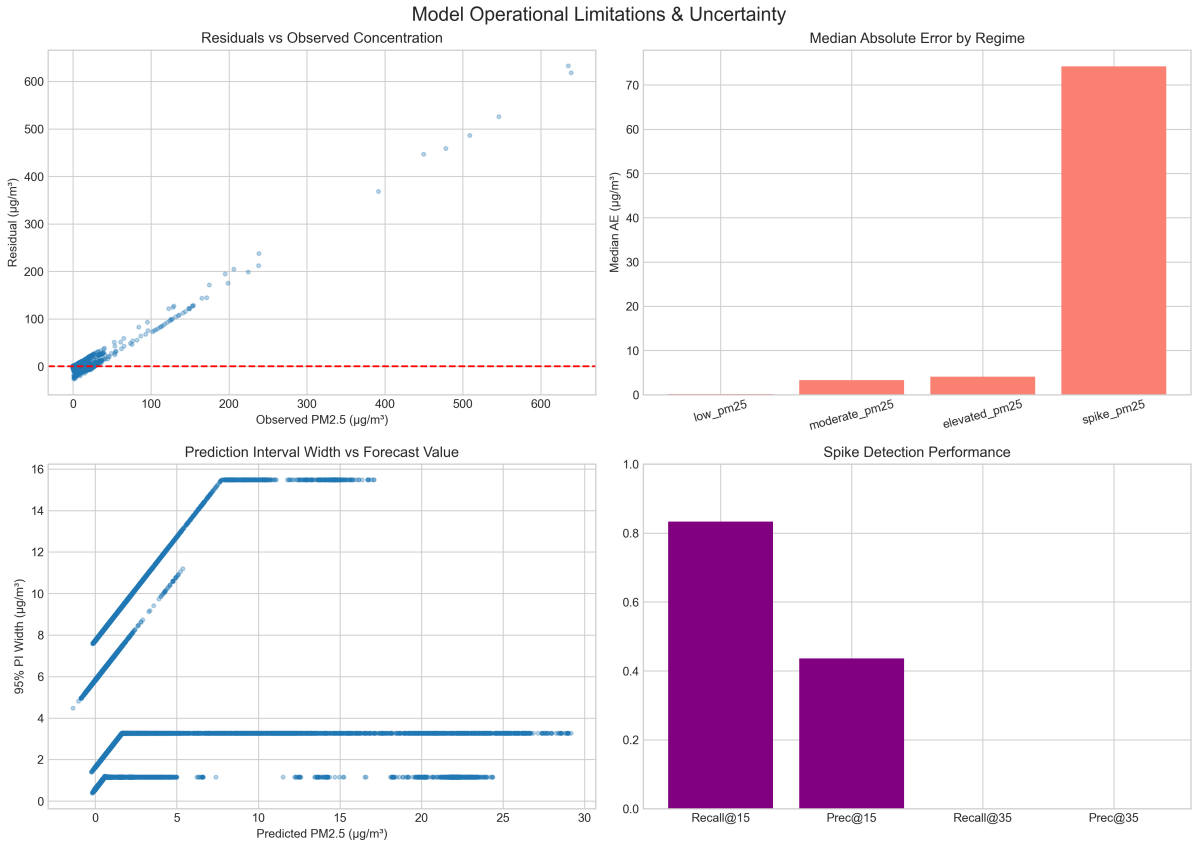


Figure 5: Spike-aware reliability assessment of the digital twin forecasts. The figure summarizes regime-dependent prediction behavior, including residual growth at high observed $PM_{2.5}$ concentrations, median absolute error across concentration regimes, prediction-interval width behavior, and spike-detection performance at moderate and stronger event thresholds. The model provides stable tracking in ordinary low-to-moderate regimes, while rare abrupt $PM_{2.5}$ peaks remain more difficult to capture, motivating future spike-specific extensions.

4.4 Effective physics calibration

The digital twin framework extends standard data-driven modeling by incorporating an effective mass-balance layer to ensure physical interpretability. Table 8 and Figure 6 present the calibrated physical

Table 7: Spike-aware evaluation summary. The table reports event-level reliability metrics for selected models and distinguishes ordinary low-to-moderate IAQ tracking from rare abrupt $PM_{2.5}$ spike events.

Horizon	Model	Recall@15	Precision@15	Recall@35	Precision@35
15 min	LightGBM	0.833	0.436	0.000	0.000
	XGBoost	0.763	0.438	0.000	0.000
	CatBoost	0.522	0.820	0.000	0.000
	RollingMean-30min	0.813	0.771	0.467	0.385
30 min	LightGBM	0.782	0.369	0.000	0.000
	XGBoost	0.668	0.408	0.000	0.000
	CatBoost	0.398	0.794	0.000	0.000
	RollingMean-30min	0.768	0.729	0.367	0.303
60 min	LightGBM	0.694	0.336	0.000	0.000
	XGBoost	0.413	0.596	0.000	0.000
	CatBoost	0.242	0.855	0.000	0.000
	RollingMean-30min	0.712	0.675	0.200	0.165

Notes: Recall@15 and Precision@15 use a moderate $PM_{2.5}$ threshold of 15 $\mu g/m^3$; Recall@35 and Precision@35 use 35 $\mu g/m^3$. Learned models retain useful moderate-event detection but rarely capture severe spikes; RollingMean-30min mainly reflects event persistence.

diagnostics derived from the sensor measurements. The physical analysis indicates that the calibrated dynamics are source-dominated and strongly TVOC-driven, with approximately 33.4% removal contribution, 66.6% source contribution, and TVOC source dominance close to 100%.

Figure 6 summarizes these effective physical insights. The calibrated coefficients are interpreted as effective mass-balance parameters inferred from available sensor streams. They support physical interpretation and scenario simulation, but they are not direct measurements of ACH, ventilation rate, HVAC performance, room volume, or airflow fields. This distinction confirms that the framework yields physically interpretable approximations under sensor-only constraints.

The extraction of effective physical parameters from standard sensor streams offers a practical advantage for indoor environments. By identifying that the local air dynamics are predominantly source-driven and highly correlated with TVOC variations, the digital twin provides contextual information. Consequently, the dashboard can support the evaluation of source-control and removal-related scenarios, bridging the gap between purely statistical forecasting and physically meaningful decision support.

4.5 Dashboard-based decision support

The final operational component of the system consists of a Streamlit-based dashboard designed to translate statistical modeling outputs and physical constraints into actionable metrics. Figure 7 illustrates the interface, which serves as a research-grade operational demonstration rather than a clinical-grade or regulatory deployment. The application is divided into three distinct panels to maintain situational awareness.

Panel A contains the executive overview and forecasting module. This interface tracks real-time variations in $PM_{2.5}$, CO_2, TVOC, temperature, and humidity, while visualizing the 15-minute forecast along with its 95% prediction interval. Panel B contains the statistical and physical diagnostics layer. It reports the validation performance of the deployed 15-minute LightGBM model via standard metrics, including RMSE, MAE, MedianAE, R^2, PICP, MPIW, and CRPS. Additionally, it displays the uncertainty

Table 8: Effective physics calibration summary. The table summarizes the effective mass-balance interpretation layer used by the proposed IAQ digital twin under sensor-only constraints.

Diagnostic / Parameter	**Value**	**Interpretation**
C_{out}	0.01527	Robust background $PM_{2.5}$ concentration used as the effective outdoor/background reference level.
C_{out} estimation method	`positive_5th_percentile`	Estimated from the lower positive tail of the observed $PM_{2.5}$ distribution rather than from direct outdoor monitoring.
`base_removal_per_min`	1.22×10^{-21}	Baseline removal term; effectively near zero in the calibrated sensor-only representation.
`aer_dco2_weight`	0.00471	Dynamic removal-related coefficient inferred from causal CO_2 change information.
`source_tvoc_weight`	9.46×10^{-4}	Effective source coefficient associated with TVOC-driven emission dynamics.
Condition number	332.77	Indicates that the calibration problem is numerically acceptable under the adopted identifiability threshold.
Identifiability status	`true`	The physical calibration is treated as numerically identifiable for the proposed effective model.
Removal contribution ratio	33.363681%	Approximately one-third of the effective calibrated signal is associated with removal-related terms.
Source contribution ratio	66.636300%	The calibrated dynamics are source-dominated under the available sensor streams.
TVOC source dominance ratio	99.999972%	The effective source dynamics are strongly dominated by the TVOC-related source term.
MCMC status	`failed_or_non_informative`	Bayesian MCMC diagnostics were not considered reliable; the reported acceptance rate was outside the useful range.
Primary parameter uncertainty method	`bootstrap`	Bootstrap intervals are used as the primary uncertainty source for calibrated parameters.
Physical interpretation statement	Effective mass-balance layer	The calibrated coefficients provide an effective physical interpretation for decision support, not direct measurements of ventilation rate, ACH, HVAC performance, room volume, or airflow fields.

Note: Coefficients are effective parameters inferred from IAQ sensor streams, not direct ventilation, ACH, HVAC, room-volume, or airflow measurements.

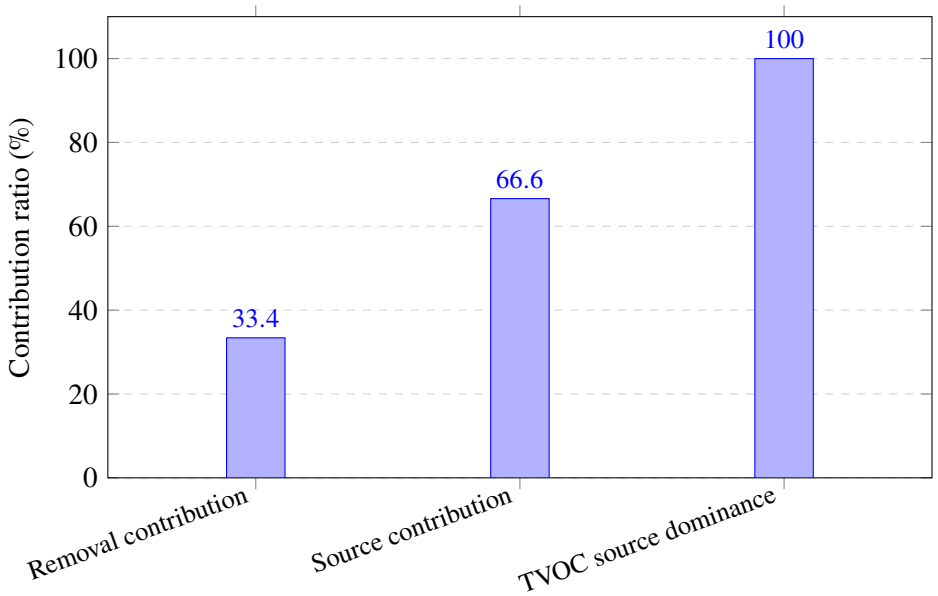
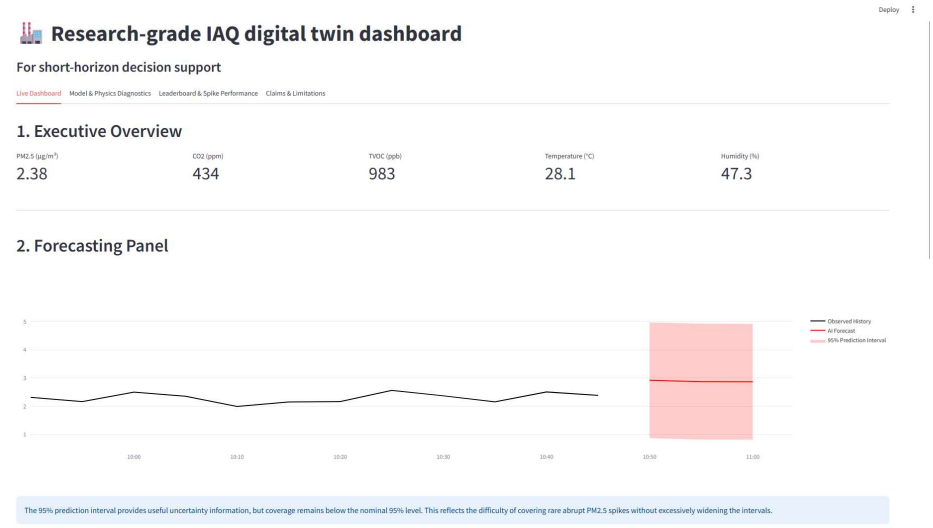


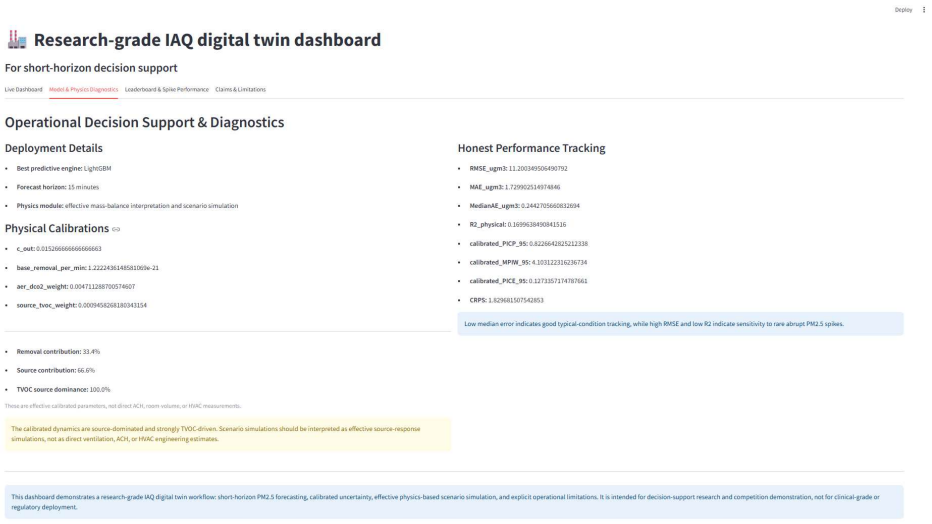
Figure 6: Effective physics diagnostics of the calibrated mass-balance layer. The calibrated dynamics are source-dominated, with the source contribution accounting for approximately 66.6% of the effective signal, compared with approximately 33.4% for removal-related terms. The TVOC source dominance ratio approaches 100%, indicating that the inferred effective source dynamics are strongly TVOC-driven under the available sensor-only calibration.

indicators and effective physics diagnostics, such as the background reference (C_{out}), source-to-removal proportions, and a TVOC dominance alert indicator. Lastly, Panel C facilitates proactive intervention testing through a combined mitigation scenario. This module simulates concurrent modifications to both effective source and effective removal dynamics.

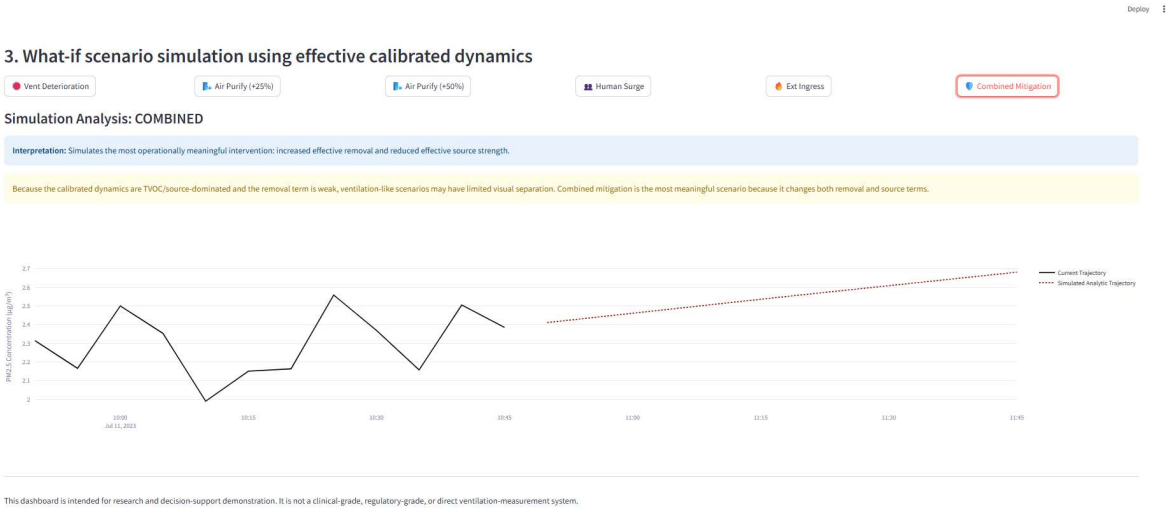
The deployment of this interactive dashboard supports the central hypothesis of the research: routine indoor sensor streams can be effectively transformed into an operational digital twin workflow. By combining predictive forecasting, uncertainty bounds, and physical scenario simulations in a single interface, the framework enables users to explore a proactive, information-driven IAQ management approach. This makes the framework suitable for low-cost, sensor-based IAQ vigilance in healthcare-oriented and resource-constrained environments.



(a) Executive overview and $PM_{2.5}$ forecast with uncertainty.



(b) Model performance, uncertainty, and physics diagnostics.



(c) Combined-mitigation scenario simulation.

Figure 7: Streamlit dashboard for the proposed IAQ digital twin. These panels illustrate real-time environmental monitoring, $PM_{2.5}$ multi-horizon forecasting, diagnostic evaluation layers, and interactive scenario simulations.

5 Conclusion

This study proposed a physics-informed IAQ digital twin framework designed to provide health-vigilance-oriented decision support under strict sensor-only constraints. While traditional monitoring platforms offer valuable real-time observation, they frequently lack the predictive and diagnostic components required to support proactive responses to evolving indoor environmental conditions.

The primary methodological contribution of this work lies in the development of an integrated workflow rather than the isolated optimization of a single forecasting model. The framework synthesizes routine IAQ sensor streams, sensor-based feature engineering, and short-horizon $PM_{2.5}$ forecasting with an effective mass-balance physical interpretation layer. It further incorporates uncertainty-aware monitoring, spike-aware reliability assessment, and scenario-based dashboard decision support. This cohesive integration bridges the gap between passive telemetry collection and actionable environmental intelligence.

Empirical evaluations determined that LightGBM operating at a 15-minute horizon provided the most balanced performance and was selected as the deployed predictive engine. The model achieved an RMSE of 11.20 $\mu g/m^3$, an MAE of 1.73 $\mu g/m^3$, a MedianAE of 0.244 $\mu g/m^3$, and an R^2 of 0.170. Tree-based ensemble models remained closely competitive across various forecasting horizons. While simple baselines occasionally exhibited low MedianAE values during stable, low-concentration periods, they provided weaker aggregate reliability and lacked the capacity to generate calibrated predictive distributions for uncertainty-aware evaluation. The generation of 95% prediction intervals offered valuable operational context; however, empirical coverage remained below the nominal target, yielding a $PICP_{95}$ of 0.823. Furthermore, spike-aware analysis indicated that while the framework demonstrated stronger detection capabilities for moderate-threshold events, anticipating severe, abrupt $PM_{2.5}$ spikes remains an ongoing challenge.

Beyond point forecasting, the effective mass-balance layer provided interpretable aggregate dynamics under sensor-only constraints. The calibrated physical behavior was found to be source-dominated and strongly TVOC-driven. It is important to emphasize that these parameters represent effective coefficients inferred dynamically from sensor streams; they do not constitute direct physical measurements of ACH, ventilation rates, HVAC performance, room volume, or spatial airflow fields. To make these insights actionable, the framework employs a dashboard interface that translates complex model outputs into practical functions, including real-time monitoring, forecast visualization, uncertainty display, physics diagnostics, and combined-mitigation scenario simulation.

Several limitations should be noted when interpreting the results of this study. The framework was developed and evaluated using sensor-only data derived from two specific indoor environments. Direct reference measurements for HVAC operation, ACH, room volume, BIM metadata, CFD simulations, and outdoor $PM_{2.5}$ concentrations were not available. Consequently, the effective physics layer must be interpreted as a simplified aggregate representation rather than a high-fidelity physical reconstruction. Additionally, the empirical prediction interval coverage remained below the nominal 95% level, and severe $PM_{2.5}$ spikes proved difficult to predict reliably. Therefore, wider external validation across a more diverse range of buildings and healthcare-oriented spaces is required to assess the broader generalizability of the framework.

Future research will focus on addressing these limitations through larger, multi-site validation efforts. Key directions include the development of improved spike-specific modeling architectures, enhanced

techniques for probabilistic uncertainty calibration, and the integration of outdoor $PM_{2.5}$ reference data and ventilation metadata where available. Furthermore, subsequent phases of this work will involve applied testing directly within healthcare-oriented environments and structured user studies with facility managers to evaluate the operational utility of the dashboard interface. Future work will also explore the integration of the proposed methodology with broader low-cost IoT deployments to improve scalable building management. Overall, the proposed framework provides a practical step toward lightweight, sensor-based IAQ digital twins that support proactive, health-vigilance-oriented decision-making in resource-constrained indoor environments.

Declaration of AI assistance

During the preparation of this work, the author used Gemini 3.1 Pro to assist with Python scripting, mathematical derivations, language editing, and LaTeX formatting. ChatGPT is used for structural review, and NotebookLM to organize and connect references. The author carefully reviewed, tested, and validated all AI-assisted content and assumes full responsibility for the final content, integrity, and factual accuracy of the manuscript.

Funding Statement

The author declares that no specific funding was received for this work.

References

[1] Dominik Latoń, Jakub Grela, Andrzej Ożadowicz, and Lukasz Wisniewski. Artificial Intelligence and Machine Learning Approaches for Indoor Air Quality Prediction: A Comprehensive Review of Methods and Applications. *Energies*, 18(19):5194, September 2025. ISSN 1996-1073. doi: 10.3390/en18195194. URL https://www.mdpi.com/1996-1073/18/19/5194.

[2] Chiu Kit Lo. The Application of Digital Twin for Indoor Air Quality Management: A Case Study in Hong Kong. *JSBE*, 2(1), January 2025. ISSN 2759-7318, 2760-327X. doi: 10.70731/8sk2aa87. URL https://jandoopress.com/journal/jsbe/article/view/90.

[3] Angel Marinov, Firgan Feradov, Tamer Abu-Alam, and Boyan Shabanski. Digital-Twin-Assisted Adaptive Sensor Scheduling for Energy Optimization in Battery-Powered Indoor Air Quality (IAQ) IoT Nodes. *Electronics*, 15(11):2395, June 2026. ISSN 2079-9292. doi: 10.3390/electronics15112395. URL https://www.mdpi.com/2079-9292/15/11/2395.

[4] Valentino Petrić, Dejan Strbad, Nikolina Račić, Tareq Hussein, Simonas Kecorius, Francesco Mureddu, and Mario Lovrić. Hybrid Digital Twin Framework for Real-Time Indoor Air Quality Monitoring and Filtration Optimization. *Atmosphere*, 17(2):184, February 2026. ISSN 2073-4433. doi: 10.3390/atmos17020184. URL https://www.mdpi.com/2073-4433/17/2/184.

[5] Cosmina-Mihaela Rosca and Adrian Stancu. Machine Learning Applications with Sensors for Indoor Air Quality Research. *Sensors*, 26(9):2909, May 2026. ISSN 1424-8220. doi: 10.3390/s26092909. URL https://www.mdpi.com/1424-8220/26/9/2909.

[6] Adekunle Dosumu, Seyed Hamed Godasiaei, Obuks A. Ejohwomu, and Saeed Reza Mohandes. Generative physics-informed machine learning for modeling indoor air quality and its impact on student health and performance. *Journal of Environmental Management*, 395:127676, December 2025. ISSN 03014797. doi: 10.1016/j.jenvman.2025.127676. URL https://linkinghub.elsevier.com/retrieve/pii/S0301479725036527.

[7] Ibrahim Yitmen, Amjad Almusaed, Muaz Hussein, and Asaad Almssad. AI-Driven Digital Twins for Enhancing Indoor Environmental Quality and Energy Efficiency in Smart Building Systems. *Buildings*, 15(7):1030, March 2025. ISSN 2075-5309. doi: 10.3390/buildings15071030. URL https://www.mdpi.com/2075-5309/15/7/1030.

[8] Xi Hu and Rayan H. Assaad. A BIM-enabled digital twin framework for real-time indoor environment monitoring and visualization by integrating autonomous robotics, LiDAR-based 3D mobile mapping, IoT sensing, and indoor positioning technologies. *Journal of Building Engineering*, 86:108901, June 2024. ISSN 23527102. doi: 10.1016/j.jobe.2024.108901. URL https://linkinghub.elsevier.com/retrieve/pii/S2352710224004698.

[9] Sena Assaf, Mohamed Assaf, Xinming Li, Ahmed Bouferguene, and Mohamed Al-Hussein. Title of the Specific Paper Here. In *Proceedings of the International Symposium on Automation and Robotics in Construction (ISARC)*, volume 42, pages 885–892, Waterloo, 2025. IAARC Publications.

[10] Ashay Singh, Mohaiminul Islam, and Nga Dinh. Forecasting Indoor Air Quality Using Machine Learning Models. In *2024 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–6, Las Vegas, NV, USA, January 2024. IEEE. ISBN 979-8-3503-2413-6. doi: 10.1109/ICCE59016.2024.10444387. URL https://ieeexplore.ieee.org/document/10444387/.

[11] Fareena Naz, Muhammad Fahim, Adnan Ahmad Cheema, Nguyen Trung Viet, Tuan-Vu Cao, Ruth Hunter, and Trung Q. Duong. Two-Stage Feature Engineering to Predict Air Pollutants in Urban Areas. *IEEE Access*, 12:114073–114085, 2024. ISSN 2169-3536. doi: 10.1109/ACCESS.2024.3443810. URL https://ieeexplore.ieee.org/document/10636142/.

[12] Manzoor Ansari and Mansaf Alam. An Intelligent IoT-Cloud-Based Air Pollution Forecasting Model Using Univariate Time-Series Analysis. *Arab J Sci Eng*, 49(3):3135–3162, March 2024. ISSN 2193-567X, 2191-4281. doi: 10.1007/s13369-023-07876-9. URL https://link.springer.com/10.1007/s13369-023-07876-9.

[13] Khaled M. Alhawiti. A Physics-Informed Multimodal Deep Learning Framework for City-Scale Air-Quality and Health-Risk Prediction. *Systems*, 14(3):320, March 2026. ISSN 2079-8954. doi: 10.3390/systems14030320. URL https://www.mdpi.com/2079-8954/14/3/320.

[14] Mojgan Sami and Francisco Sierra. Physics-informed machine learning for ASHRAE-compliant heat transfer coefficient estimation from short-term IoT monitoring. *Energy and Buildings*, 354:116984, March 2026. ISSN 03787788. doi: 10.1016/j.enbuild.2026.116984. URL https://linkinghub.elsevier.com/retrieve/pii/S0378778826000447.

[15] Fahad Iqbal and Shayan Mirzabeigi. Digital Twin-Enabled Building Information Modeling–Internet of Things (BIM-IoT) Framework for Optimizing Indoor Thermal Comfort Using Machine Learning.

Buildings, 15(10):1584, May 2025. ISSN 2075-5309. doi: 10.3390/buildings15101584. URL https://www.mdpi.com/2075-5309/15/10/1584.

[16] Marcus Vogt, Jan Schlichter, Franziska Aschersleben, Tim Abraham, Lars Wolf, and Christoph Herrmann. Integration of cyber-physical HVAC systems in Incremental Manufacturing to improve Energy Efficiency and Air Quality. *Procedia CIRP*, 104:482–487, 2021. ISSN 22128271. doi: 10.1016/j.procir.2021.11.081. URL https://linkinghub.elsevier.com/retrieve/pii/S2212827121009793.

[17] Jindong Tian, Yuxuan Liang, Ronghui Xu, Peng Chen, Chenjuan Guo, Aoying Zhou, Lujia Pan, Zhongwen Rao, and Bin Yang. Air quality prediction with physics-guided dual neural odes in open systems, 2025. URL https://arxiv.org/abs/2410.19892.

[18] A. AghaKouchak, B. Pan, O. Mazdiyasni, M. Sadegh, S. Jiwa, W. Zhang, C. A. Love, S. Madadgar, S. M. Papalexiou, S. J. Davis, K. Hsu, and S. Sorooshian. Status and prospects for drought forecasting: opportunities in artificial intelligence and hybrid physical–statistical forecasting. *Phil. Trans. R. Soc. A.*, 380(2238):20210288, December 2022. ISSN 1364-503X, 1471-2962. doi: 10.1098/rsta.2021.0288. URL https://royalsocietypublishing.org/doi/10.1098/rsta.2021.0288.

[19] Yuanyuan Wei, Julian Jang-Jaccard, Wen Xu, Fariza Sabrina, Seyit Camtepe, and Mikael Boulic. LSTM-Autoencoder based Anomaly Detection for Indoor Air Quality Time Series Data, April 2022. URL http://arxiv.org/abs/2204.06701. arXiv:2204.06701 [cs.LG].

[20] Giuseppe Piras, Sofia Agostinelli, and Francesco Muzi. Smart Buildings and Digital Twin to Monitoring the Efficiency and Wellness of Working Environments: A Case Study on IoT Integration and Data-Driven Management. *Applied Sciences*, 15(9):4939, April 2025. ISSN 2076-3417. doi: 10.3390/app15094939. URL https://www.mdpi.com/2076-3417/15/9/4939.

[21] Yitong Li, Holly Josephs, Yaocheng Wu, Gediminas Mainelis, Clinton Andrews, and Jie Gong. An air quality digital twin for real-time outdoor air quality monitoring and prediction. *Building and Environment*, 290:114127, February 2026. ISSN 03601323. doi: 10.1016/j.buildenv.2025.114127. URL https://linkinghub.elsevier.com/retrieve/pii/S0360132325015938.

[22] Kabir Bahadur Shah, Dylan Kim, Sai Deepak Pinakana, Mkhitar Hobosyan, Armando Montes, and Amit U. Raysoni. Evaluating Indoor Air Quality in Residential Environments: A Study of PM2.5 and CO2 Dynamics Using Low-Cost Sensors. *Environments*, 11(11):237, October 2024. ISSN 2076-3298. doi: 10.3390/environments11110237. URL https://www.mdpi.com/2076-3298/11/11/237.

[23] Sattar Hedayat, Tina Ziarati, and Matteo Manganelli. A Physics-Informed Reinforcement Learning Framework for HVAC Optimization: Thermodynamically-Constrained Deep Deterministic Policy Gradients with Simulation-Based Validation. *Energies*, 18(23):6310, November 2025. ISSN 1996-1073. doi: 10.3390/en18236310. URL https://www.mdpi.com/1996-1073/18/23/6310.

[24] Eleonora Congiu, Giuseppe Desogus, Emanuela Quaquero, Giulia Rubiu, and Francesca Poggi. Towards a Digital Twin for Buildings IAQ and Thermal Comfort Monitoring. *Applied Sciences*, 15(19):10444, September 2025. ISSN 2076-3417. doi: 10.3390/app151910444. URL https://www.mdpi.com/2076-3417/15/19/10444.

[25] Tae Wook Kang and Yunjeong Mo. A comprehensive digital twin framework for building environment monitoring with emphasis on real-time data connectivity and predictability. *Developments in the Built Environment*, 17:100309, March 2024. ISSN 26661659. doi: 10.1016/j.dibe.2023.100309. URL https://linkinghub.elsevier.com/retrieve/pii/S2666165923001916.

[26] Ahmed Albashir. A Physics-Informed Digital Twin Framework for Health Vigilance and Indoor Air Quality Decision Support in Healthcare-Oriented Environments. https://doi.org/10.5281/zenodo.20688071, June 2026. Software and Code Repository for Research Replication.

[27] Ali Uoa. 2023 indoor air quality dataset germany. https://www.kaggle.com/code/aliuoa/2023-indoor-air-quality-dataset-germany/notebook, 2023. Accessed: June 14, 2026.

[28] Weiwei Ren, Zhongzheng Zhu, Yingzheng Wang, Jianbin Su, Ruijie Zeng, Donghai Zheng, and Xin Li. Comparison of Machine Learning Models in Simulating Glacier Mass Balance: Insights from Maritime and Continental Glaciers in High Mountain Asia. *Remote Sensing*, 16(6):956, March 2024. ISSN 2072-4292. doi: 10.3390/rs16060956. URL https://www.mdpi.com/2072-4292/16/6/956.

[29] Jingyue Pang, Datong Liu, Yu Peng, and Xiyuan Peng. Optimize the Coverage Probability of Prediction Interval for Anomaly Detection of Sensor-Based Monitoring Series. *Sensors*, 18(4): 967, March 2018. ISSN 1424-8220. doi: 10.3390/s18040967. URL https://www.mdpi.com/1424-8220/18/4/967.

[30] Agustín Zaballos, Alan Briones, Alba Massa, Pol Centelles, and Víctor Caballero. A Smart Campus' Digital Twin for Sustainable Comfort Monitoring. *Sustainability*, 12(21):9196, November 2020. ISSN 2071-1050. doi: 10.3390/su12219196. URL https://www.mdpi.com/2071-1050/12/21/9196.

[31] Ming Jin, Nikolaos Bekiaris-Liberis, Kevin Weekly, Costas J. Spanos, and Alexandre M. Bayen. Occupancy Detection via Environmental Sensing. *IEEE Trans. Automat. Sci. Eng.*, 15(2):443–455, April 2018. ISSN 1545-5955, 1558-3783. doi: 10.1109/TASE.2016.2619720. URL https://ieeexplore.ieee.org/document/7742900/.