PIAD-SRNN: Physics-Informed Adaptive Decomposition in State-Space RNN

Ahmad Mohammadshirazi [1] **Pinaki Prasad Guha Neogi** [1] **Rajiv Ramnath** [1]

Abstract

Time series forecasting often demands a trade-off between accuracy and efficiency. While recent Transformer models have improved forecasting capabilities, they come with high computational costs. Linear-based models have shown better accuracy than Transformers but still fall short of ideal performance. We propose PIAD-SRNN—a physics-informed adaptive decomposition state-space RNN—that separates seasonal and trend components and embeds domain equations in a recurrent framework. We evaluate PIAD-SRNN's performance on indoor air quality datasets, focusing on CO_2 concentration prediction across various forecasting horizons, and results demonstrate that it consistently outperforms SoTA models in both long-term and short-term time series forecasting, including transformer-based architectures, in terms of both MSE and MAE. Besides proposing PIAD-SRNN which balances accuracy with efficiency, this paper also provides four curated datasets. Code and data: https://github.com/ahmad-shirazi/DSSRNN

1. Introduction

Time series forecasting demands both expressiveness and efficiency while also robustly handling missing data. On the one hand, self-attention forecasting models (e.g., Transformers) capture long-range dependencies but incur quadratic compute and tuning overhead; conversely, linear-decomposition methods run fast but miss nonlinear dynamics (Zeng et al., 2023). In addition, existing imputation approaches (for missing values in real-time data) either ignore system insights or add significant computational cost.

To address all these gaps in one step, we introduce the Physics-Informed Adaptive Decomposition State-Space Recurrent Neural Network (PIAD-SRNN). This idea was inspired by the success of nature-inspired (Neogi, 2019; Guha Neogi et al., 2020) and physics-oriented (Guha Neogi & Goswami, 2021; Kar & Neogi, 2020) modeling paradigms in ML and optimization problems. PIAD-SRNN embeds physics-based state-space equations within a recurrent architecture that decomposes each time series into trend and seasonal components. By merging mechanistic modeling with data-driven decomposition (inspired by DLinear (Zeng et al., 2023), FEDformer (Zhou et al., 2022)), PIAD-SRNN captures both global patterns and fine-grained temporal variations while retaining linear-time complexity per step. We validate PIAD-SRNN on four Indoor Air Quality (IAQ) datasets we created, focusing on CO_2 concentration forecasting across multiple horizons (up to 720 steps). Compared to leading Transformer and linear baselines, PIAD-SRNN achieves substantial gains in MSE and MAE, while requiring minimal inference time and memory. Our contributions are: **(1) Model.** We propose PIAD-SRNN, a novel architecture that integrates physics-informed state-space modeling with adaptive decomposition for efficient and accurate long-term and short-term time series forecasting and imputation, and is also robust to outliers. **(2) Datasets.** We release four curated IAQ datasets and all code to foster reproducibility and advance research in physics-informed time series learning. **(3) Evaluation.** Through extensive experiments we demonstrate that PIAD-SRNN consistently outperforms SoTA baselines in both predictive accuracy as well as computational efficiency.

2. Related Work

2.1. Time Series Imputation

Time series imputation has evolved from traditional RNN-based methods like GRU-D (Che et al., 2018), BRITS (Cao et al., 2018), and M-RNN (Yoon et al., 2018), which incorporate time decay factors and bidirectional processing but struggle with long-term dependencies and computational intensity. Generative approaches using GANs and VAEs (Luo et al., 2018; Liu et al., 2019; Fortuin et al., 2020) offer sophisticated distribution learning but face training complexity and scalability issues. Recent self-attention-based models like SAITS (Du et al., 2023) optimize imputation through diagonally-masked attention blocks, though computational efficiency remains challenging.

[*]Equal contribution [1]Department of Computer Science and Engineering, Ohio State University, Ohio, US. Correspondence to: Ahmad Mohammadshirazi <mohammadshirazi.2@osu.edu>.

Proceedings of the 41st International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025. Copyright 2025 by the author(s).

2.2. Time Series Forecasting

Classical statistical methods like ARIMA (Siami-Namini et al., 2018) and (Hyndman, 2018) excel at capturing linear relationships and explicit seasonal patterns in univariate time series, but they face limitations with nonlinear dynamics and complex multivariate relationships. Transformer architectures (Vaswani et al., 2017) revolutionized forecasting through self-attention mechanisms, spawning specialized variants: Informer (Zhou et al., 2021) addresses quadratic complexity with ProbSparse attention, Autoformer (Wu et al., 2021) employs Auto-Correlation mechanisms, and iTransformer (Liu et al., 2023) processes variate tokens for multivariate modeling. Surprisingly, DLinear (Zeng et al., 2023) demonstrated that simple linear decomposition can outperform complex Transformers, challenging the necessity of architectural complexity. Recent approaches like SegRNN (Lin et al., 2023) uses segment-wise iterations, & TiDE (Das et al., 2023) combines linear models with MLPs.

PIAD-SRNN model integrates physics-informed state-space modeling with decomposition techniques from DLinear and Autoformer, achieving superior computational efficiency while maintaining the interpretability and robustness required for real-time environmental applications. For a detailed literature review, see Appendix A.

3. Data Preparation

3.1. Dataset Description

The scarcity of IAQ datasets and their limited availability present a significant challenge to research in this domain. For our study, we created datasets from Lawrence Berkeley National Laboratory's Building 59 in California, covering the period from Aug. 19, 2019, to Dec. 31, 2021, creating 20,760 data points from different office settings. It encompasses hourly indoor CO_2 measurements along with other crucial environmental variables—a collection of data infrequently encountered in existing studies. The location of each office numbered 44, 45, 62, and 68 within the building is detailed in a map provided in (Luo et al., 2022), offering contextual insight into the data collection environment. A key aspect of our research is the prediction of indoor CO_2 concentration (CO_{2_in}), using five current environmental variables such as current indoor CO_2 levels, time of day (Hour), day of the week (Num_Week), and both indoor and outdoor temperatures (T_{in} and T_{out}) (Mohammadshirazi et al., 2022; 2023). Appendix B presents data analysis & validations showing complexities of the real-world datasets.

3.2. Handling Missing Observations

Table 1 provides a detailed examination of the data absence across various input variables collected from four distinct office environments. To effectively manage this dataset, we segregated it into training, validation, and testing subsets, with all instances of missing data allocated to the testing subset. The proposed PIAD-SRNN model embeds physics-informed state-space equations within an adaptive seasonal–trend decomposition, enabling it to perform imputation and forecasting in a single recurrent framework. We fit the model by minimizing MSE and MAE on the validation set—first recovering missing CO_2 values, then seamlessly sliding into multi-horizon prediction (and outlier clarification-Appendix G). The injected physical priors yield robustness to substantial data gaps, ensuring reliable IAQ reconstruction and long-term accuracy. Appendix C shows the detection and distribution of outliers in our datasets.

Input Variables	CO_{2-in}	T_{in}	T_{out}
Office1	2,823 (13.6%)	3,027 (14.6%)	2,823 (13.6%)
Office2,3,4	2,823 (13.6%)	2,823 (13.6%)	2,823 (13.6%)

Table 1. Input Variables with Missing Entries for Predicting Indoor CO_2 Concentrations across Four Office Environments

4. Physics-Informed Recurrent Modeling

Here, we present two nested architectures. First, **PI-SRNN** (Physics-Informed State-Space RNN) embeds a classical CO_2 mass-balance into an RNN; then **PIAD-SRNN** builds on PI-SRNN by splitting each series into trend and seasonal components for improved multi-horizon forecasting.

4.1. Physics-Informed State-Space RNN (PI-SRNN)

The PI-SRNN architecture represents a novel blend of traditional state-space models with the dynamic capabilities of recurrent neural networks (RNNs). At its core, PI-SRNN aims to model time series data by capturing both the underlying physical processes and the temporal dependencies. This is achieved through a combination of state variables that embody the physical state of the system, and RNN layers that process sequential data, allowing for the integration of prior knowledge and temporal context. The PI-SRNN architecture is designed to efficiently handle the flow of information through time, making it particularly adept at predicting complex dynamics in time series data, such as those encountered in IAQ monitoring. The transformation from physics to ML equations begins with the foundational state-space representation of indoor CO_2 dynamics (Persily, 2022), as shown in Equation 1.

$$\frac{\partial}{\partial t} CO_{2_in} = \frac{\dot{m}}{\rho V}(CO_{2_out} - CO_{2_in}) + \frac{M_t}{\rho} \quad (1)$$

This equation models the temporal change in indoor CO_2 concentration by relating it to variables that capture both the physical characteristics of the environment and the dy-

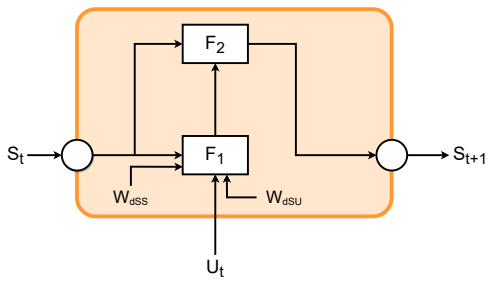


Figure 1. Physics-Informed State-Space RNN (PI-SRNN).

namics of CO_2 exchange. Specifically, the mass flow rate $\dot{m}$ reflects the air exchange due to ventilation, represented by $\frac{\dot{m}}{\rho V} \times CO_{2_out}$, ρ signifies air density, V stands for the volume of the indoor space, and M_t represents the internal generation of CO_2, such as those generated by occupants or appliances within the space. The variable CO_{2_out}, meanwhile, captures the concentration of CO_2 outside the building. Together, these variables form the basis of our state-space approach, setting the stage for its translation into an ML framework.

Equation 2 is where the conversion to an ML framework takes place. The term A represents the state matrix, encapsulating the system dynamics, and B the input matrix, incorporating external influences. These matrices interact with the state vector S_t (Equation 3) and input vector U_t (Equation 4) to form a discretized version of the physics-based model, respectively. This interaction is key to defining the subsequent state vector S_{t+1} and the differential state change dS_{t+1}. These state vectors, embedded within the ML architecture, allow neural networks to approximate the intricate relationships found within the dataset. The continuous state update from S_t to S_{t+1} (Equation 3), facilitated by the differential state dS_{t+1}, is a discrete analog of the physical process. This reformulation allows us to leverage the power of neural networks to approximate the complex relationships within the data, retaining the interpretability and structure provided by the physical model.

$$\underbrace{\frac{\partial}{\partial t}(CO_{2_in})}_{dS_{t+1}} = \underbrace{-\frac{\dot{m}}{\rho V}}_{A} \times \underbrace{CO_{2_in}}_{S_t} + \left\{ \underbrace{\frac{\dot{m}}{\rho V} \times CO_{2_out} + \frac{M_t}{\rho}}_{B \times U_t} \right\} \quad (2)$$

$$S_{t+1} \leftarrow CO_{2_in_{t+1}}, S_t \leftarrow CO_{2_in_t} \quad (3)$$

$$U_t \leftarrow CO_{2_in_t}, Hour_t, T_{in_t}, T_{out_t}, Num_week_t \quad (4)$$

The PI-SRNN model, presented in Figure 1, is meticulously crafted to process inputs within each discrete time step. The model assimilates the current state vector S_t and the input vector U_t to advance its internal state. The transformation of S_t alongside U_t is governed by a nonlinear function F_1, resulting in the interim state change dS_{t+1}, as shown in equation 5. This intermediate state is then sculpted by another nonlinear function F_2, culminating in the updated state S_{t+1} (Equation 6). Weight matrices W_{dSS} and W_{dSU}, alongside bias vectors b_{dSS} and b_{dSU}, underpin this evolution, modulating the influence of past and present information and embedding non-linearity via ReLU activation functions, denoted as F_1 and F_2 (see equation 7).

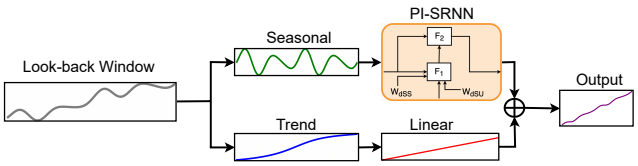


Figure 2. Physics-Informed Adaptive Decomposition State-Space Recurrent Neural Network (PIAD-SRNN).

Based on Figure 1 representation, the PI-SRNN model state space and its changes are updated as follows:

$$dS_{t+1} \leftarrow F_1(S_t \times W_{dSS} + b_{dSS} + U_t \times W_{dSU} + b_{dSU}) \quad (5)$$

$$S_{t+1} \leftarrow F_2(S_t + dS_{t+1}) \quad (6)$$

$$S_{t+1} \leftarrow F_2(S_t + F_1(S_t \times W_{dSS} + b_{dSS} + U_t \times W_{dSU} + b_{dSU}) \quad (7)$$

It should be noted that the equation referenced in 1, while primarily developed for modeling CO_2, can be adapted to encompass a variety of indoor pollutants such as NO_2, O_3, PM_1, $PM_{2.5}$, and PM_{10}. This adaptation necessitates modifying the generation term M_t for each specific pollutant's source strength and incorporating particular deposition or removal processes. Such modifications allow the PI-SRNN model to extend its applicability across a range of indoor pollutants, demonstrating its versatility in different environmental conditions. In the following subsection, we will expound on the intricacies of the PIAD-SRNN and its implementation for enhanced predictive performance.

4.2. Physics-Informed Adaptive Decomposition (PIAD-SRNN)

Although PI-SRNN effectively blends domain knowledge with sequence learning, complex IAQ patterns benefit from separating slow trends from fast cycles. PIAD-SRNN applies a 24 h moving average to each series, yielding:

$$X_t = T_t + S_t,$$

Methods	Metric	Office 1 (44)				Office 2 (45)				Office 3 (62)				Office 4 (68)			
		T = 96	T = 192	T = 336	T = 720	T = 96	T = 192	T = 336	T = 720	T = 96	T = 192	T = 336	T = 720	T = 96	T = 192	T = 336	T = 720
Informer	MSE	0.531	0.567	0.514	0.625	0.523	0.555	0.549	0.593	0.522	0.535	0.510	0.529	0.576	0.577	0.522	0.520
	MAE	0.508	0.527	0.482	0.553	0.504	0.519	0.509	0.538	0.477	0.471	0.465	0.484	0.503	0.516	0.502	0.492
PatchTST	MSE	0.431	0.484	0.497	0.526	0.410	0.461	0.475	0.496	0.382	0.442	0.456	0.491	0.422	0.484	0.501	0.516
	MAE	0.460	0.485	0.489	0.504	0.452	0.476	0.480	0.489	0.420	0.453	0.460	0.473	0.453	0.484	0.491	0.495
TiDE	MSE	0.491	0.521	0.557	0.612	0.481	0.526	0.573	0.648	0.391	0.404	0.411	0.436	0.507	0.536	0.561	0.613
	MAE	0.471	0.489	0.502	0.535	0.473	0.490	0.510	0.549	0.415	0.422	0.432	0.447	0.488	0.500	0.513	0.541
Linear	MSE	0.452	0.503	0.508	0.536	0.426	0.477	0.483	0.502	0.397	0.450	0.463	0.497	0.436	0.495	0.507	0.519
	MAE	0.475	0.499	0.501	0.509	0.462	0.487	0.489	0.498	0.438	0.466	0.475	0.488	0.469	0.498	0.504	0.502
DLinear	MSE	0.454	0.500	0.504	0.535	0.422	0.474	0.479	0.499	0.392	0.446	0.458	0.494	0.429	0.491	0.503	0.517
	MAE	0.475	0.497	0.498	0.508	0.459	0.485	0.486	0.496	0.434	0.463	0.472	0.485	0.464	0.495	0.501	0.501
SegRNN	MSE	0.491	0.564	0.554	0.576	0.483	0.571	0.560	0.544	0.453	0.549	0.548	0.580	0.509	0.600	0.630	0.613
	MAE	0.488	0.523	0.520	0.530	0.486	0.524	0.518	0.513	0.462	0.506	0.509	0.521	0.504	0.548	0.566	0.558
PI-SRNN	MSE	0.393	0.444	0.450	0.477	0.370	0.418	0.424	0.444	0.342	0.391	0.403	0.438	0.378	0.437	0.448	0.460
	MAE	0.416	0.440	0.442	0.450	0.406	0.428	0.430	0.439	0.382	0.406	0.416	0.429	0.410	0.439	0.444	0.443
PIAD-SRNN	MSE	**0.378**	**0.428**	**0.434**	**0.463**	**0.353**	**0.403**	**0.409**	**0.429**	**0.320**	**0.375**	**0.388**	**0.425**	**0.359**	**0.421**	**0.433**	**0.445**
	MAE	**0.401**	**0.426**	**0.428**	**0.437**	**0.390**	**0.415**	**0.416**	**0.426**	**0.362**	**0.392**	**0.402**	**0.416**	**0.394**	**0.424**	**0.431**	**0.430**

Table 2. Comparative Forecasting Performance: This table presents the MSE and MAE across forecasting horizons (T) of 96, 192, 336, 720 for four office environments, with input sequence length L=96. Bold figures indicate the best overall performance, while underlined figures show the top results for transformer-based, MLP-based, and linear-based models. (More comparisons in Appendix D-Table 5)

where T_t (trend) and S_t (seasonal) feed into two parallel paths (Fig. 2): **(1) Trend branch:** a lightweight linear layer captures smooth, long-term shifts with minimal computation. **(2) Seasonal branch:** a PI-SRNN block models cyclical fluctuations via its physics-informed recurrent updates. The final prediction is simply

$$\widehat{X}_{t+h} = \widehat{T}_{t+h} + \widehat{S}_{t+h},$$

This regularizes learning across time scales and reduces interference between components. Empirically, PIAD-SRNN yields superior MSE/MAE on multi-horizon IAQ benchmarks while maintaining low inference cost.

5. Experiments

In the results section of our paper, we evaluate the proposed models using the robust NVIDIA A100 GPU at the Ohio Supercomputer Center (OSC) [1]. The forthcoming subsections will present a detailed account of the model's imputation and prediction capabilities, each scrutinized through rigorous experimentation. These findings will illustrate PIAD-SRNN's strengths in accurately forecasting and handling missing data within the context of IAQ assessment.

5.1. Imputation Analysis

Table 3 compares single-step (T=1) imputation performance of PIAD-SRNN with SoTA techniques like DLinear (Zeng et al., 2023), and SAITS (Du et al., 2023). PIAD-SRNN consistently achieves the lowest MSE and MAE across all 4 offices, underscoring its exceptional ability to reconstruct missing IAQ measurements with unparalleled precision. By embedding physics-driven state-space dynamics within an adaptive seasonal–trend decomposition, PIAD-SRNN captures complex temporal patterns and robustly handles data sparsity. This evaluation demonstrates that our model's advanced architecture is instrumental in learning and imputing data patterns, outshining SoTA imputation methods.

[1] https://www.osc.edu/

Method	Metric	Office 1	Office 2	Office 3	Office 4
PIAD-SRNN	MSE	**0.357**	**0.180**	**0.217**	**0.373**
	MAE	**0.375**	**0.284**	**0.293**	**0.367**
DLinear	MSE	0.383	0.199	0.233	0.397
	MAE	0.393	0.299	0.304	0.382
SAITS	MSE	1.079	0.956	0.922	0.954
	MAE	0.686	0.701	0.669	0.710

Table 3. One-step imputation (T=1) performance: MSE and MAE for PIAD-SRNN, DLinear, and SAITS across four offices. Bold entries denote best results.

5.2. Prediction Analysis

We benchmark PIAD-SRNN against SoTA baselines following Zeng et al. (Zeng et al., 2023) for multi-horizon forecasting at $T = \{96, 192, 336, 720\}$. Table 2 (and Appendix D) shows that PIAD-SRNN achieves the lowest MSE/MAE across all horizons and office environments, outperforming DLinear—the top linear model—as well as Transformer (listed in Table 5) at shorter horizons and Informer (Zhou et al., 2021) at longer horizons.

We also evaluate the impact of the look-back window size for both long-term (720 steps) and short-term (24 steps) forecasting in Appendix E-Fig. 4. PIAD-SRNN consistently balances accuracy and efficiency, outperforming high-cost models like TiDE (Das et al., 2023) in most settings. These results underscore PIAD-SRNN's adaptability across temporal scales and its superior accuracy–efficiency trade-off.

Additional Analyses To complement our core evaluations, we include the following appendices: Appendices D and E present the additional model performance analysis, and the extended visual analysis of the model performance respectively. Appendix F reports PIAD-SRNN's computational efficiency—runtime, memory footprint, and inference latency—against all baselines. Finally, Appendix G presents extended regression-to-classification results when forecasting is recast as an event detection task.

6. Conclusion and Future Work

This study introduced the Physics-Informed Adaptive Decomposition State-Space Recurrent Neural Network (PIAD-SRNN), which combines physics-based modeling with state-space methodologies for accurate time-series forecasting and data imputation, demonstrating superior computational efficiency and outperforming transformer-based models in both MSE and MAE to deliver a robust solution for environmental data analysis. Future work will extend PIAD-SRNN to broader applications, including energy consumption forecasting and climate-control optimization, thereby validating its generalizability across diverse domain-specific contexts.

References

Bansal, P., Deshpande, P., and Sarawagi, S. Missing value imputation on multidimensional time series. *arXiv preprint arXiv:2103.01600*, 2021.

Cao, W., Wang, D., Li, J., Zhou, H., Li, L., and Li, Y. Brits: Bidirectional recurrent imputation for time series. *Advances in neural information processing systems*, 31, 2018.

Casale, F. P., Dalca, A., Saglietti, L., Listgarten, J., and Fusi, N. Gaussian process prior variational autoencoders. *Advances in neural information processing systems*, 31, 2018.

Che, Z., Purushotham, S., Cho, K., Sontag, D., and Liu, Y. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8(1):6085, 2018.

Das, A., Kong, W., Leach, A., Mathur, S., Sen, R., and Yu, R. Long-term forecasting with tide: Time-series dense encoder. *arXiv preprint arXiv:2304.08424*, 2023.

Du, W., Côté, D., and Liu, Y. Saits: Self-attention-based imputation for time series. *Expert Systems with Applications*, 219:119619, 2023.

Fortuin, V., Baranchuk, D., Rätsch, G., and Mandt, S. Gp-vae: Deep probabilistic time series imputation. In *International conference on artificial intelligence and statistics*, pp. 1651–1661. PMLR, 2020.

Guha Neogi, P. P. and Goswami, S. Force of gravity oriented classification technique in machine learning. In Sharma, N., Chakrabarti, A., Balas, V. E., and Martinovic, J. (eds.), *Data Management, Analytics and Innovation*, pp. 299–310, Singapore, 2021. Springer Singapore. ISBN 978-981-15-5616-6.

Guha Neogi, P. P., Goswami, S., and Mustafi, J. Intelligent water drops-based image steganography. In Mandal, J. K. and Mukhopadhyay, S. (eds.), *Proceedings of the Global AI Congress 2019*, pp. 363–375, Singapore, 2020. Springer Singapore. ISBN 978-981-15-2188-1.

Hyndman, R. *Forecasting: principles and practice*. OTexts, 2018.

Kar, A. and Neogi, P. Triangular coil pattern of local radius of gyration face for heterogeneous face recognition. *Applied Intelligence*, 50(3):698–716, 2020. doi: 10.1007/s10489-019-01545-x.

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Lin, S., Lin, W., Wu, W., Zhao, F., Mo, R., and Zhang, H. Segrnn: Segment recurrent neural network for long-term time series forecasting. *arXiv preprint arXiv:2308.11200*, 2023.

Liu, Y., Yu, R., Zheng, S., Zhan, E., and Yue, Y. Naomi: Non-autoregressive multiresolution sequence imputation. *Advances in neural information processing systems*, 32, 2019.

Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., and Long, M. itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*, 2023.

Luo, D. and Wang, X. Moderntcn: A modern pure convolution structure for general time series analysis. In *The Twelfth International Conference on Learning Representations*, 2024.

Luo, N., Wang, Z., Blum, D., Weyandt, C., Bourassa, N., Piette, M. A., and Hong, T. A three-year dataset supporting research on building energy management and occupancy analytics. *Scientific Data*, 9(1):156, 2022.

Luo, Y., Cai, X., Zhang, Y., Xu, J., et al. Multivariate time series imputation with generative adversarial networks. *Advances in neural information processing systems*, 31, 2018.

Mohammadshirazi, A., Kalkhorani, V. A., Humes, J., Speno, B., Rike, J., Ramnath, R., and Clark, J. D. Predicting airborne pollutant concentrations and events in a commercial building using low-cost pollutant sensors and machine learning: a case study. *Building and Environment*, 213: 108833, 2022.

Mohammadshirazi, A., Nadafian, A., Monsefi, A. K., Rafiei, M. H., and Ramnath, R. Novel physics-based machine-learning models for indoor air quality approximations. *arXiv preprint arXiv:2308.01438*, 2023.

Neogi, P. P. G. Iwd-based mobile sink route determination for wsn under non-uniform data constraints. In *2019 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1–6, 2019. doi: 10.1109/ICCCI.2019.8821993.

Nie, Y., Nguyen, N. H., Sinthong, P., and Kalagnanam, J. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*, 2022.

Persily, A. Development and application of an indoor carbon dioxide metric. *Indoor Air*, 32(7):e13059, 2022.

Shan, S., Li, Y., and Oliva, J. B. Nrtsi: Non-recurrent time series imputation. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2023.

Siami-Namini, S., Tavakoli, N., and Namin, A. S. A comparison of arima and lstm in forecasting time series. In *2018 17th IEEE international conference on machine learning and applications (ICMLA)*, pp. 1394–1401. Ieee, 2018.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Wu, H., Xu, J., Wang, J., and Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems*, 34:22419–22430, 2021.

Yoon, J., Zame, W. R., and van der Schaar, M. Estimating missing data in temporal data streams using multi-directional recurrent neural networks. *IEEE Transactions on Biomedical Engineering*, 66(5):1477–1490, 2018.

Zeng, A., Chen, M., Zhang, L., and Xu, Q. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 11121–11128, 2023.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., and Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 11106–11115, 2021.

Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., and Jin, R. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International Conference on Machine Learning*, pp. 27268–27286. PMLR, 2022.

Appendix

A. Related Work (Detailed)

We survey background work in two areas: imputation of missing time-series data and multi-horizon forecasting.

A.1. Imputation Methods

Accurate imputation in time series is crucial for dataset preparation, as in our case detailed in Table 1. Traditional RNN-based methods, such as GRU-D (Che et al., 2018), have introduced time decay factors to address missing data. Approaches like BRITS (Cao et al., 2018) and M-RNN (Yoon et al., 2018) enhance this with bidirectional RNNs, yet they grapple with the inherent complexities of long-term dependencies and computational intensity. Generative models, including GANs and VAEs, provide alternative strategies (Luo et al., 2018; Liu et al., 2019; Casale et al., 2018; Fortuin et al., 2020), but their training complexity and interpretability issues limit their practical scalability.

The emergence of self-attention-based models has introduced a new paradigm in time series imputation. Models like DeepMVI (Bansal et al., 2021) and NRTSI (Shan et al., 2023) leverage Transformer architectures for handling multidimensional and irregularly sampled time series data. Despite their novel approach, these models face challenges in computational efficiency and generalizability. Du et al. (2023) recently advanced this field with SAITS, a model that optimizes imputation through diagonally-masked self-attention blocks, addressing both temporal dependencies and feature correlations with enhanced efficiency (Du et al., 2023).

Our approach diverges significantly from these existing methods. As noted in the introduction, we integrate domain knowledge with ML. This enhances the accuracy of our model while also addressing the limitations of existing approaches. Additionally, leveraging neural networks within this framework allows for efficient processing of larger datasets, balancing robustness with computational agility.

A.2. Forecasting Techniques

Time series forecasting has deep roots in statistical methods. The Autoregressive Integrated Moving Average (ARIMA) model (Siami-Namini et al., 2018) has been a cornerstone of time series analysis since the 1970s, decomposing time series into autoregressive and moving average components while handling non-stationarity through differencing. Error, Trend, and Seasonality (ETS) models (Hyndman, 2018) provide another classical approach, explicitly modeling these three components with various combinations of additive and multiplicative effects. While these methods excel at capturing linear relationships and explicit seasonal patterns in univariate time series, they face limitations with nonlinear dynamics and complex multivariate relationships.

Transformer models by Vaswani et al. (2017) (Vaswani et al., 2017), have revolutionized time series forecasting through their innovative use of self-attention mechanisms, enabling the model to focus selectively on relevant parts of the input. This model's parallelizable architecture significantly enhances its efficiency in handling large datasets. There are also newer models based on transformers such as PatchTST (Nie et al., 2022), Informer (Zhou et al., 2021), and Autoformer (Wu et al., 2021).

Furthermore, Informer model by Zhou et al. (2021) (Zhou et al., 2021), which addresses the quadratic time complexity and memory usage inherent in traditional Transformer models, introduces the ProbSparse self-attention mechanism, a distilling operation, and a generative decoder to manage long-sequence time-series forecasting effectively. This model demonstrates a marked improvement in prediction capabilities for long-sequence forecasting tasks by mitigating the limitations of the encoder-decoder architecture.

The FEDformer model, introduced by Zhou et al. (2022) (Zhou et al., 2022), represents another advancement in time series forecasting by integrating Transformer architecture with seasonal-trend decomposition. This hybrid approach enables the FEDformer to capture both the global trends and the intricate details within the time series data. By leveraging the sparsity in the frequency domain, FEDformer achieves enhanced efficiency and effectiveness in long-term forecasting, outpacing conventional Transformer models in both performance and computational efficiency.

In contrast, the Autoformer model by Wu et al. (2021) (Wu et al., 2021) incorporates an Auto-Correlation mechanism and a novel series decomposition strategy within its architecture. This approach allows the Autoformer to dissect and analyze complex time series data, focusing on seasonal patterns while progressively incorporating trend components. This model's unique design facilitates a deeper understanding of temporal dependencies, particularly in subseries levels, thereby offering a robust solution for long-term forecasting challenges.

iTransformer (Liu et al., 2023) introduces a perspective by inverting the traditional transformer architecture. Rather than processing temporal tokens, it operates on variate tokens, enabling more effective modeling of multivariate correlations. This approach maintains the advantages of transformer architectures while addressing their limitations in time series contexts, particularly for long sequences and complex multivariate relationships.

In addition, the introduction of a linear decomposition model, Dlinear, by Zeng et al. (2023) (Zeng et al., 2023)

challenges the prevailing dominance of Transformer-based models in long-term time series forecasting. By simplifying the model to focus on linear relationships within the data, Dlinear astonishingly surpasses more complex models in performance. This revelation underscores the potential of minimalistic approaches in extracting temporal relationships effectively, thereby questioning the necessity of complex architectures for certain forecasting tasks.

Moreover, the Modern Temporal Convolutional Network (ModernTCN) by Luo et al. (2024) (Luo & Wang, 2024), leverages convolutional operations to model temporal dependencies. ModernTCN employs dilated convolutions, residual connections, and layer normalization, resulting in a highly effective time series forecasting model with reduced computational overhead. However, ModernTCN may suffer from limited representation capabilities due to its lightweight backbone, potentially leading to inferior performance in certain tasks.

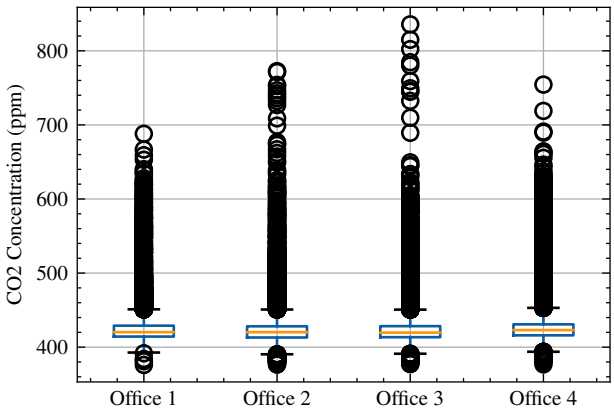
Additionally, SegRNN by Lin et al. (2023) (Lin et al., 2023) revisits RNN-based methods by introducing segment-wise iterations and parallel multi-step forecasting. These techniques reduce the number of recurrent steps and enhance forecasting accuracy and inference speed, making SegRNN a highly efficient tool for long-term time series forecasting. Despite these improvements, SegRNN can be constrained by the inherent limitations of RNNs, such as vanishing gradients and difficulty in capturing long-term dependencies.

Lastly, TiDE by Das et al. (2023) (Das et al., 2023) combines the strengths of linear models and multi-layer perceptrons (MLPs). TiDE's dense encoder architecture captures both linear and nonlinear dependencies in time series data, offering robust performance without the excessive computational demands of more complex models like transformers. Nonetheless, TiDE can struggle with modeling non-linear dependencies and may face challenges when dealing with complex covariate structures.

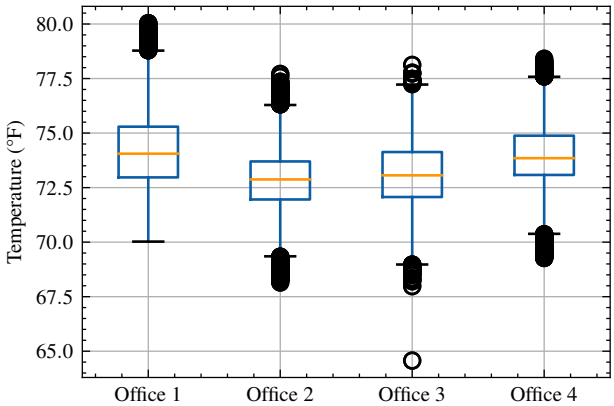
Our PIAD-SRNN model builds upon decomposition techniques used in Dlinear, Autoformer, and Fedformer, while significantly enhancing computational efficiency. By optimizing its architecture, PIAD-SRNN achieves faster training and inference times than leading transformer-based models, making it particularly suitable for real-time applications.

B. Data Analysis

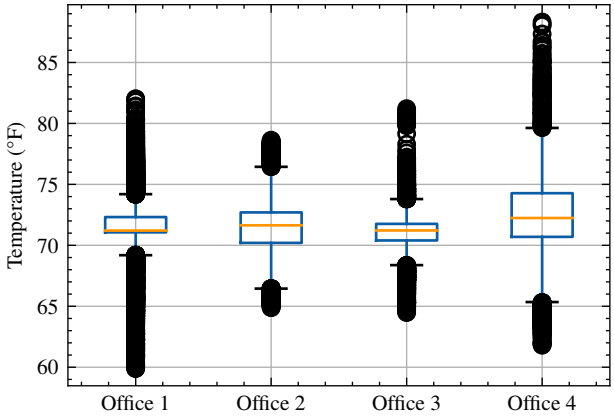
In Figure 3, we present a comprehensive depiction of the environmental conditions across four different office settings through the analysis of indoor temperature, outdoor temperature, and indoor CO_2 concentration. The subfigure 3(a) focuses on the indoor CO_2 concentration levels. Across all offices, the median CO_2 concentration is consistent, indicating a general adherence to acceptable IAQ standards.



(a) Indoor CO_2 Concentration (CO_{2_in})



(b) Indoor Temperature (T_{in})



(c) Outdoor Temperature (T_{out})

Figure 3. Comparative environmental conditions across four office spaces.

However, the presence of numerous high outliers in each office suggests episodic events where the CO_2 levels exceed typical ranges. These instances could potentially be linked to variable occupancy levels, inconsistent operation of ventilation systems, or activities within the offices that temporarily increase CO_2 production.

The indoor temperature, illustrated in the subfigure 3(b), reveals a notable variance between offices. Office 1 exhibits a broad temperature range with several high outliers, indicative of sporadic deviations from the norm, which could be attributed to intermittent HVAC system performance or external factors affecting the indoor climate. In contrast, Offices 2 and 3 display a narrower interquartile range, suggesting a more consistent application of temperature control measures. Office 4, while also showing a relatively stable temperature range, has more outliers on the lower end, pointing to occasional dips in the indoor temperature.

The subfigure 3(c) captures the outdoor temperature conditions for each office location. Offices 1, 3, and 4 show a wider range of temperature values, which may reflect either a larger variance in outdoor climatic conditions or a more extended period of data capture. Office 2 demonstrates tighter temperature controls with minimal outliers, suggesting either less climatic variability or more effective mitigation of outdoor temperature fluctuations. Overall, the multitude of data eliminates the statistical biases, and the variety of the data supports the training of a robust model.

Analysis of the dataset reveals the presence of missing values, as detailed in Table 1, corresponding to the period from January 2020 to April 2020. The missing data is imputed, and appropriate sections of the dataset are selected for training, testing, and validation to maintain the integrity of our analysis. The upcoming section elaborates on this method and its significance in our research.

C. Predicting outlier events

	Threshold	Number of events Training/(Testing)	Percentage Training/(Testing)
Office 1	451.12	1112/(268)	7.70%/(6.46%)
Office 2	450.84	1019/(260)	7.06%/(6.26%)
Office 3	450.75	1175/(292)	8.14%/(7.03%)
Office 4	453.05	928/(271)	6.43%/(6.53%)

Table 4. Comparative analysis across training and testing datasets between the number of high-concentration CO_2 events and the corresponding percentage of data points for each office alongside their third quartile threshold

In order to demonstrate the robustness of our model in predicting outlier events within the CO_2 data, we employ a binary classification approach based on the box-and-whisker plot method. Specifically, CO_2 measurements above the upper whisker—calculated as $Q_3 + 1.5 \times IQR$ where Q_3 is the third quartile and IQR is the interquartile range—are classified as outliers and encoded as '1', while non-outliers are assigned '0' (See the threshold value for each office in Table 4. This binary transformation allows us to specifically focus on the prediction of outlier occurrences. We then compare the performance of our model in predicting these binary outcomes against SoTA models, showcasing our model's capability to identify significant deviations in air quality. The number of outlier events detected for each office is systematically cataloged in Table 4. The comparative results and analysis are presented in detail in the Result section of the paper.

D. Additional Model Performance Analysis

To provide a comprehensive evaluation of our PIAD-SRNN model, we conducted additional comparative analyses with both traditional and modern time series forecasting approaches. Table 5 presents an extended comparison of performance metrics across different models and prediction horizons for all four office environments. Among the traditional statistical models, ARIMA and ETS demonstrate consistent but relatively higher error rates across all prediction horizons. For instance, in Office 1, ARIMA shows MSE values ranging from 0.881 to 1.398, while ETS exhibits similar performance patterns with MSE values between 0.910 and 1.435. These results highlight the limitations of traditional statistical approaches in capturing complex temporal dependencies in indoor air quality data. The neural network-based models (VanillaRNN, GRU, and LSTM) show improved performance compared to statistical methods. Notably, GRU and LSTM achieve competitive results, particularly for shorter prediction horizons (T=96), with MSE values consistently below 0.5 across all offices. This improvement can be attributed to their ability to learn and retain temporal patterns in the data. FEDformer and Transformer architectures demonstrate varying degrees of success, with Transformer showing particularly strong performance in shorter prediction horizons but experiencing some degradation in accuracy for longer forecasting periods. The Autoformer model maintains relatively stable performance across different prediction lengths, though not achieving the optimal results of our proposed PIAD-SRNN model. The Nlinear model, while computationally efficient, shows moderate performance levels that remain relatively consistent across different prediction horizons, suggesting its utility in scenarios where computational resources are constrained.

These supplementary results further validate the effectiveness of our PIAD-SRNN approach, as presented in the manuscript's Table 2, by providing additional context and comparison points across a broader range of methodologies and architectural approaches.

E. Extended Visual Analysis of Model Performance

Figure 4 provides a comprehensive visualization of model performance across different temporal scales and office envi-

Methods	Metric	Office 1 (44)				Office 2 (45)				Office 3 (62)				Office 4 (68)			
		T = 96	T = 192	T = 336	T = 720	T = 96	T = 192	T = 336	T = 720	T = 96	T = 192	T = 336	T = 720	T = 96	T = 192	T = 336	T = 720
ARIMA	MSE	0.881	0.887	1.398	1.359	1.208	1.022	1.320	1.452	1.064	0.939	1.194	1.191	1.076	0.949	1.442	1.561
	MAE	0.640	0.729	0.862	0.838	0.845	0.808	0.858	0.872	0.735	0.764	0.801	0.773	0.751	0.766	0.930	0.911
ETS	MSE	0.910	0.904	1.435	1.387	1.181	1.007	1.301	1.436	1.008	0.910	1.146	1.149	0.960	0.893	1.313	1.433
	MAE	0.655	0.736	0.877	0.847	0.831	0.801	0.850	0.866	0.706	0.753	0.780	0.762	0.700	0.748	0.879	0.874
FEDformer	MSE	1.062	1.208	1.166	1.186	1.008	1.054	0.974	1.143	1.070	1.076	0.944	0.891	1.030	1.137	1.123	0.878
	MAE	0.740	0.792	0.770	0.786	0.723	0.740	0.709	0.768	0.745	0.749	0.709	0.691	0.750	0.793	0.788	0.687
Transformer	MSE	0.528	0.599	0.651	0.643	0.421	0.555	0.572	0.553	0.471	0.636	0.500	0.534	0.528	0.529	0.713	0.581
	MAE	0.497	0.536	0.566	0.565	0.461	0.539	0.533	0.525	0.448	0.522	0.465	0.477	0.497	0.490	0.566	0.525
Autoformer	MSE	0.619	0.667	0.767	0.704	0.586	0.613	0.603	0.599	0.537	0.607	0.934	0.620	0.648	0.729	1.193	0.645
	MAE	0.598	0.583	0.648	0.598	0.556	0.557	0.561	0.561	0.527	0.555	0.698	0.555	0.591	0.629	0.809	0.586
iTransformer	MSE	0.562	0.520	0.585	0.606	0.433	0.483	0.509	0.619	0.562	0.520	0.585	0.606	0.555	0.574	0.659	0.713
	MAE	0.463	0.460	0.486	0.488	0.459	0.484	0.439	0.484	0.463	0.463	0.486	0.488	0.472	0.482	0.496	0.511
ModernTCN	MSE	0.491	0.586	0.586	0.605	0.460	0.549	0.549	0.551	0.428	0.522	0.531	0.567	0.473	0.586	0.615	0.599
	MAE	0.482	0.525	0.526	0.529	0.474	0.516	0.513	0.512	0.445	0.490	0.495	0.506	0.487	0.541	0.556	0.543
Nlinear	MSE	0.513	0.615	0.611	0.621	0.476	0.569	0.569	0.569	0.448	0.547	0.557	0.585	0.494	0.614	0.642	0.616
	MAE	0.500	0.545	0.542	0.542	0.485	0.527	0.523	0.522	0.462	0.506	0.510	0.518	0.501	0.556	0.571	0.554
VanillaRNN	MSE	0.452	0.508	0.498	0.544	0.411	0.463	0.499	0.526	0.399	0.439	0.477	0.477	0.432	0.513	0.485	0.524
	MAE	0.470	0.512	0.493	0.505	0.450	0.465	0.495	0.508	0.428	0.457	0.452	0.471	0.462	0.499	0.481	0.501
GRU	MSE	0.451	0.490	0.528	0.558	0.512	0.542	0.518	0.586	0.459	0.453	0.481	0.504	0.467	0.523	0.549	0.526
	MAE	0.459	0.474	0.483	0.503	0.498	0.514	0.502	0.546	0.446	0.456	0.472	0.476	0.484	0.508	0.520	0.503
LSTM	MSE	0.456	0.507	0.538	0.548	0.422	0.494	0.538	0.538	0.420	0.427	0.447	0.479	0.454	0.556	0.584	0.516
	MAE	0.456	0.480	0.499	0.497	0.453	0.500	0.526	0.529	0.438	0.443	0.457	0.472	0.477	0.520	0.536	0.494

Table 5. Additional Model Comparisons: Performance metrics for supplementary models across different prediction horizons. These results complement the primary findings presented in Table 2.

ronments. The analysis is structured to compare short-term (24 steps) and long-term (720 steps) forecasting capabilities, offering insights into how various models handle different prediction horizons. In the short-term forecasting scenario (24 steps), shown in the top row of Figure 4, we observe that PIAD-SRNN maintains consistently lower MSE values across all four offices compared to other models. This is particularly evident in Office 1 and Office 3, where the performance gap between PIAD-SRNN and other models becomes more pronounced as the look-back window size increases. The Transformer and Autoformer models show competitive performance but exhibit more volatility in their error rates, especially with larger look-back windows. The long-term forecasting results (720 steps), displayed in the bottom row, reveal the robustness of PIAD-SRNN in maintaining prediction accuracy over extended time horizons. While most models show degraded performance with increased prediction length, PIAD-SRNN demonstrates remarkable stability. This is particularly notable in Office 2 and Office 4, where other models show significant performance deterioration with larger look-back windows. The graphs also highlight an interesting pattern regarding the impact of look-back window size on model performance. For most models, including TIDE and FEDformer, increasing the look-back window beyond certain thresholds (typically around 300-400 steps) does not yield substantial improvements in prediction accuracy. In contrast, PIAD-SRNN shows more consistent improvement with increased historical data, suggesting better utilization of temporal information.

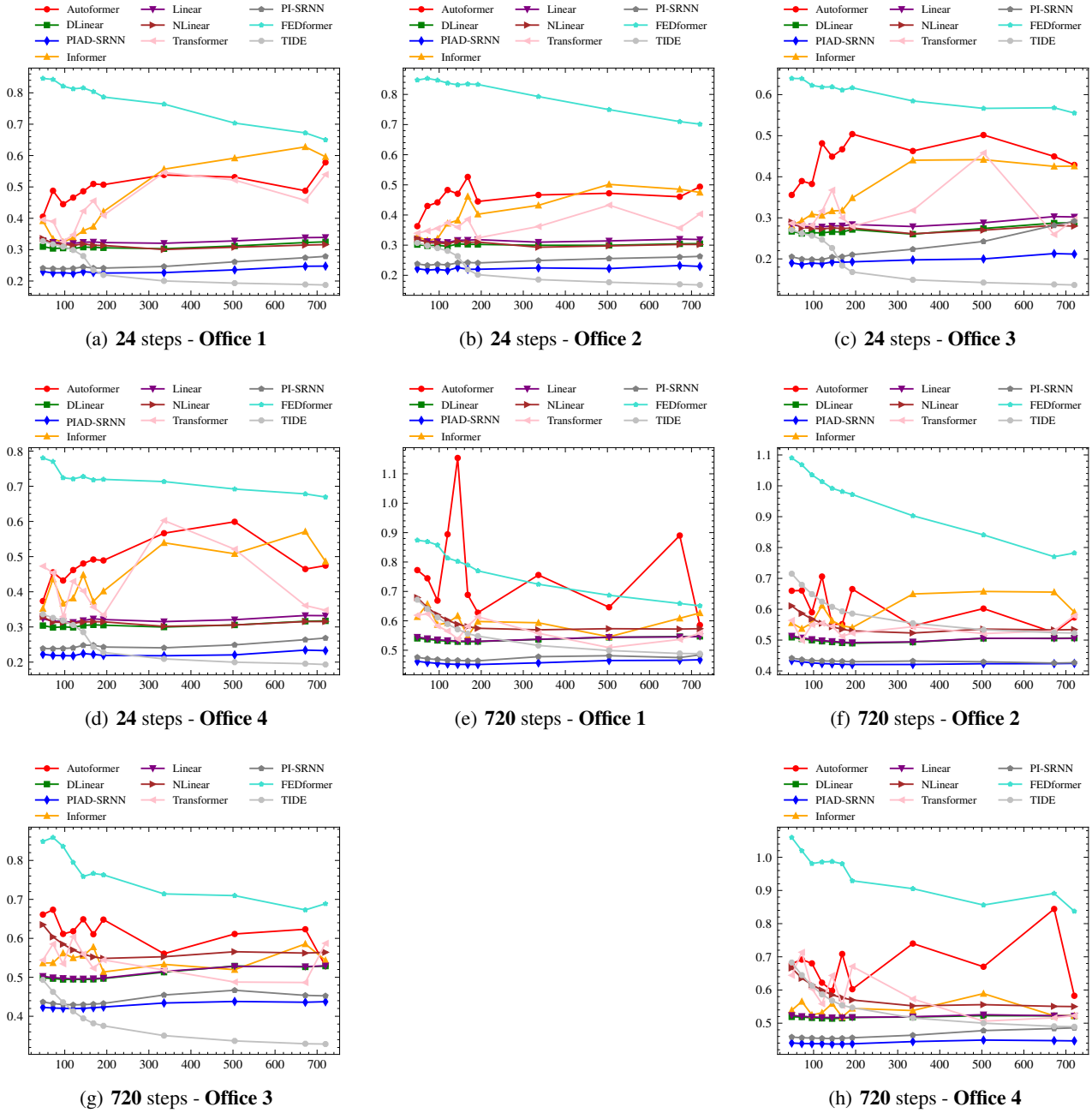
These visualizations support our quantitative findings presented in Table 2 of the manuscript, providing additional evidence for PIAD-SRNN's superior performance in both short-term and long-term forecasting scenarios. The consistent performance across different offices and temporal scales demonstrates the robustness and reliability of our proposed approach.

Method	MACs	Parameter	Time	Memory
PIAD-SRNN	0.11G	408.21K	0.58ms	437MiB
DLinear	0.04G	139.7K	0.15ms	196MiB
TiDE	1.09G	2.69M	32.87ms	1011MiB
PatchTST	4.79G	18.21M	14.03ms	2033MiB
Transformer	4.03G	13.61M	10.31ms	1740MiB
Informer	3.93G	14.39M	18.96ms	1105MiB
FEDformer	4.41G	20.68M	15.58ms	1183MiB

Table 6. Comparative analysis of computational efficiency between PIAD-SRNN, DLinear, and various transformer-based models, evaluated on the Office #1 air quality dataset with input sequence length $L = 96$ and prediction sequence length $T = 720$. Metrics include multiply-accumulate operations (MACs), model parameters, average inference time over five runs, and memory usage.

F. Computational Efficiency

In our comparative analysis, delineated in Table 6, we scrutinize the computational efficiency of our proposed PIAD-SRNN model compared to DLinear and a suite of Transformer-based architectures. This evaluation is predicated on the Office #1 air quality dataset, adopting an input sequence length $L = 96$ and a prediction sequence length $T = 720$. The criteria for this comparison encom-



(a) **24** steps - **Office 1**

(b) **24** steps - **Office 2**

(c) **24** steps - **Office 3**

(d) **24** steps - **Office 4**

(e) **720** steps - **Office 1**

(f) **720** steps - **Office 2**

(g) **720** steps - **Office 3**

(h) **720** steps - **Office 4**

Figure 4. Comparative MSE Analysis for Forecasting Models: Illustrated here are the MSE metrics for models with varying look-back window sizes for both long-term (e.g., 720 steps) and short-term (e.g., 24 steps) forecasting. The Y-axis measures MSE, while the X-axis represents the look-back window size, highlighting performance across four offices.

pass multiply-accumulate operations (MACs), the aggregate number of model parameters, the average inference time over a quintet of runs, and the requisite memory allocation during computational processes.

DLinear distinguished itself as the most computationally efficient model, demanding the least computational resources with a mere 0.04G MACs and minimal memory usage of 196MiB, alongside a remarkable average inference speed of 0.15ms. This exceptional efficiency renders it particularly advantageous for resource-constrained settings. Close on its heels, PIAD-SRNN stands as the second most efficient model, consuming slightly more resources at 0.11G MACs and 437MiB of memory but still delivering a swift inference time of 0.58ms, validating its efficacy.

G. Classification

Performance Metrics	Model	Office 1	Office 2	Office 3	Office 4
True Positive	PIAD-SRNN	192	209	234	218
	DLinear	181	189	223	203
	PatchTST	179	178	219	207
False Positive	PIAD-SRNN	84	80	46	67
	DLinear	137	111	67	95
	PatchTST	98	99	57	98
True Negative	PIAD-SRNN	3,799	3,811	3,813	3,807
	DLinear	3,757	3,780	3,792	3,785
	PatchTST	3,785	3,792	3,802	3,782
False Negative	PIAD-SRNN	76	51	46	73
	DLinear	87	71	69	68
	PatchTST	89	82	73	64
Accuracy	PIAD-SRNN	96.15%	96.84%	97.52%	96.96%
	DLinear	94.87%	95.62%	96.72%	96.07%
	PatchTST	95.50%	95.64%	96.87%	96.10%
Precision	PIAD-SRNN	69.57%	72.32%	84.63%	74.91%
	DLinear	58.96%	63.00%	76.90%	68.12%
	PatchTST	64.62%	64.26%	79.35%	67.87%
Recall	PIAD-SRNN	71.64%	80.38%	80.48%	80.44%
	DLinear	67.54%	72.69%	76.37%	74.91%
	PatchTST	66.79%	68.46%	75.00%	76.38%
F1 Score	PIAD-SRNN	70.59%	76.14%	82.02%	77.58%
	DLinear	62.96%	67.50%	76.63%	71.35%
	PatchTST	65.69%	66.29%	77.11%	71.88%

Table 7. Comparative analysis of model performance metrics for CO_2 concentration event classification over T = 96.

In an effort to further comprehend the regression capabilities of our PIAD-SRNN model, we transformed the regression task into a classification problem, as meticulously explained in Section C. The results of this transformation are depicted in Table 7, where we delineate the thresholds set for event detection, the count of events, and their respective percentages for each office environment.

The classification approach allows us to dissect the model's performance in distinguishing between high-concentration and normal event occurrences. While the DLinear model exhibits commendable recall for high-concentration events—likely a result of its decomposition capability, breaking down time series into seasonal and trend components—the Autoformer model shows a preference for normal events. This tendency is attributed to its Transformer-based architecture, which is proficient in modeling long-range dependencies within the data.

Notwithstanding the strengths displayed by DLinear and Autoformer in their respective domains, the PIAD-SRNN model demonstrates a clear overarching superiority. It transcends the typical decomposition approach by integrating state space physics concepts, offering a profound understanding of the dynamics of the time series data. PIAD-SRNN consistently maintains higher True Positive rates alongside lower False Positive rates, indicating an adeptness at correctly identifying events while minimizing false alarms—a critical aspect for real-world applications. The True Negative rate is also the highest for PIAD-SRNN, reflecting its strong discriminative power in differentiating non-events. Precision and Recall are metrics where PIAD-SRNN distinctly outshines its counterparts. The model's precision is indicative of the reliability and exactitude of its predictions, while the higher recall rates—especially for high-concentration events—demonstrate an increased sensitivity and a lower propensity to overlook significant events.