XLVI CILAMCE
VITÓRIA - ES - BRAZIL - 2025
XLVI IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGENEERING

Designing forecasting models using composite AutoML for air quality in a coastal urban area

Julia Borisova[1], Daniil Kravchenko[1], Xeniya Bashkova[1], Alexey Lapin[1], Georgii Lopatenko[1], Nikolay O. Nikitin[1], Leonardo Goliatt[2]

[1]*NSS Lab, AI Institute, ITMO University*
Kronverksky Pr. 49, bldg. A, St. Petersburg, 197101, Russia
[2]*Department of Computational and Applied Mechanics, Federal University of Juiz de Fora, 36036-900, MG, Juiz de Fora, Brazil*
leonardo.goliatt@ufjf.br

Abstract.

Air pollution remains a major environmental health risk, especially in urban areas with high industrial and vehicular emissions. Accurate air quality forecasting is essential for public health and aligns with global sustainability efforts. This study explores the application of machine learning and automated machine learning (AutoML) in air quality forecasting. We used data from Vitória, Brazil, a coastal city often exposed to air pollution. Models were built and tuned using the FEDOT AutoML framework to learn both linear and non-linear time patterns. The AutoML models were compared to standard prediction methods and showed similar accuracy, with much less need for manual design. Furthermore, it was observed that the integration of domain-specific knowledge is still critical for correct exploratory data analysis and formulation of machine learning tasks. This work highlights the potential of AutoML to simplify the development of adaptive and scalable models for air quality monitoring that can be used for environmental policy-making, public health planning, and sustainable urban development.

Keywords: air quality forecasting, automated machine learning, time series prediction, FEDOT, urban pollution

1 Introduction

Air pollution remains a major concern for public health and urban sustainability [1]. Elevated concentrations of particulate matter (PM2.5, PM10), nitrogen dioxide (NO_2), and ozone (O_3) are directly linked to respiratory and cardiovascular diseases [2–4]. Accurate short-term forecasts of pollutant levels are essential for timely risk communication, individual exposure reduction, and preventive actions by authorities [5].

Traditional prediction models based on dispersion equations or statistical regressions require extensive inventories and meteorological data [6]. They often lose accuracy when transferred to new locations or when monitoring coverage is sparse or emission patterns change rapidly [7].

Recent years have seen a shift toward data-driven forecasting. Machine-learning algorithms applied to historical and meteorological data—ranging from ensemble methods (Random Forest, Gradient Boosting) to deep architectures (LSTM, attention, graph networks)—capture complex temporal and spatial patterns [8, 9]. Explainable techniques such as SHAP enhance interpretability, while accuracy is typically assessed using R^2, RMSE, MAE, NSE, and KGE [10]. Nevertheless, operational deployment remains challenging due to station heterogeneity, limited data availability [11], sensor outages [12], and manual parameter tuning [13]. Moreover, few models provide analytical transparency required for regulatory use [14].

[10].

Forecasting physical variables from observations alone remains demanding. Recent advances such as SINDY [15] and EPDE [16] demonstrate the potential of discovering governing equations directly from data—an important step toward interpretable physics-informed forecasting.

This study investigates automated procedures for short-term air-quality prediction using data from thirteen monitoring stations in a coastal Brazilian city. Forecasts are produced through both manually configured models and AutoML pipelines assembled by the **FEDOT** framework, enabling direct comparison in accuracy, efficiency, and ease of deployment. The station-based organization supports inputs from neighbouring sites and related pollutant variables, ensuring continuity even when individual measurements are missing.

To ensure transparency, symbolic regression is integrated to derive explicit equations describing pollutant relationships, providing a means for physical validation and anomaly detection. The dataset and procedures reflect the typical limitations of Latin American monitoring networks—irregular data quality, sparse coverage, and constrained computational resources. The proposed workflow aims to combine predictive accuracy with interpretability and to offer a transferable, resource-efficient solution for regions facing similar infrastructural challenges.

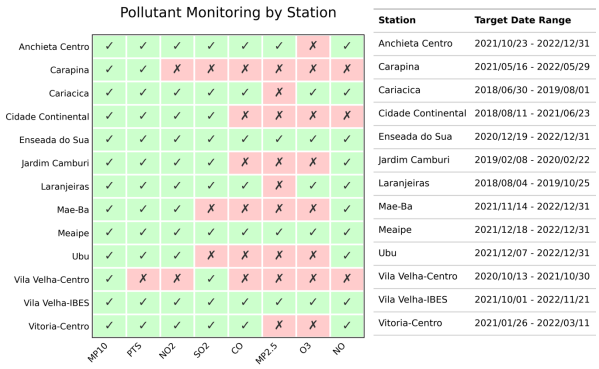
2 Material and Methods

2.1 Problem Statement

The dataset comprises hourly time series from 13 monitoring stations located along the Espírito Santo coastline (Figure 1a). Each station measures a distinct subset of pollutants over varying time spans. Although the records contain gaps, the 1-hour resolution allows for stable temporal modeling. Due to inconsistent coverage, the data cannot be treated as a unified multivariate series; thus, each station is modeled independently. Figure 1b summarizes pollutant availability and illustrates the time ranges considered for training and testing, where the last available year of observations at each station forms the test set.



(a) Spatial position of monitoring stations that are included in the dataset



Station	Target Date Range
Anchieta Centro	2021/10/23 - 2022/12/31
Carapina	2021/05/16 - 2022/05/29
Cariacica	2018/06/30 - 2019/08/01
Cidade Continental	2018/08/11 - 2021/06/23
Enseada do Sua	2020/12/19 - 2022/12/31
Jardim Camburi	2019/02/08 - 2020/02/22
Laranjeiras	2018/08/04 - 2019/10/25
Mae-Ba	2021/11/14 - 2022/12/31
Meaipe	2021/12/18 - 2022/12/31
Ubu	2021/12/07 - 2022/12/31
Vila Velha-Centro	2020/10/13 - 2021/10/30
Vila Velha-IBES	2021/10/01 - 2022/11/21
Vitoria-Centro	2021/01/26 - 2022/03/11

(b) Analysis of pollutant measurements per station and target dates of time series per station

MP_{10} (PM_{10} - Particulate Matter 10) — a type of air pollutant consisting of solid or liquid particles with a diameter of 10 micrometers (μm) or smaller was selected as target variable. Brazil CONAMA resolution provide 24 hours mean concentration gradation for MP_{10} pollutant. To work with informative data and reduce stochastic component time series was processed with sliding window, averaging values for 24 hours. Examples of data before and after processing are presented on Figure 2.

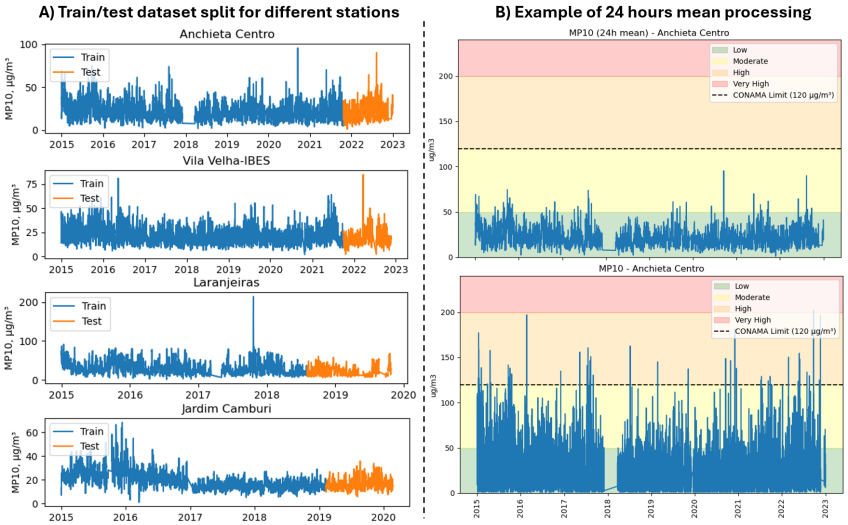


Figure 2. A) Examples of time series split for train and test sets. B) Visualization of time series before (MP_{10}) and after rolling window processing (MP_{10} 24h mean).

Dataset forming was produced in terms of regression problem statement: fixed window with two previous

days (48 hours) is used to predict next one hour. Sliding window by time series form table of data for each station separately.

2.2 Expert-designed models

For comparison with automatic and semi-automatic strategies, four expert-designed baseline models were implemented.

The first three are intentionally simple yet form a solid foundation for short-term forecasting, particularly for time series without clear periodic or trend components. The **Chronos** zero-shot baseline, based on a large pre-trained language model, is equally straightforward: it operates without fine-tuning and can be applied directly to new datasets. This makes it both powerful and accessible, offering competitive performance out of the box while retaining the simplicity of classical approaches.

2.3 Classic AutoML tools

Although deep learning and large foundation models receive growing attention, classical AutoML frameworks built on conventional machine-learning algorithms remain strong and computationally efficient baselines for many forecasting tasks. We employ **FEDOT** [17], an open-source AutoML system that constructs multi-step pipelines through evolutionary search.

FEDOT represents each workflow as a directed acyclic graph (DAG) comprising preprocessing and regression nodes, evolved via genetic programming. Compared with task-agnostic solutions such as AutoSklearn or H2O AutoML, it offers advantages particularly relevant for operational air-quality monitoring: transparent, interpretable pipelines; lightweight deployment using only scikit-learn components; and customizable objective functions that align optimization with health-related metrics.

In our experiments, the evolutionary search runs for 30 generations with a population of 20 individuals and 5-fold internal cross-validation. The best pipeline—typically a Lagged feature generator followed by ridge regression—is retrained on the full training set and evaluated on the 8,760-hour test split described in Section 3. This setup yields a strong, interpretable baseline for comparison with more advanced models.

The evolutionary search in FEDOT was configured for 30 generations with a population of 20 individuals and 5-fold internal cross-validation. The search space included nine primitive operations: a `Lagged` feature generator with window sizes of 12, 24, 48 and eight regression models from `scikit-learn` (linear, ridge, lasso, elastic-net, random forest, extra-trees, gradient boosting, and k-nearest neighbours). Structural and parametric mutations were applied with equal probability, using a maximum pipeline depth of three nodes. Each optimization run was limited to 240 s per station on a 4-core CPU, typically converging within 1–2 minutes to a two-layer composition `Lagged`,→,`Ridge`. End-to-end training and inference for all 13 stations required under 30 minutes, confirming that classical AutoML remains a practical solution for rapid model prototyping under limited computational budgets.

2.4 Physics Informed models

To assess the potential of physics-informed forecasting, we employed the **EPDE** framework [16] for automated discovery of differential equations with embedded physical priors. This approach is particularly suitable for environmental applications, where pollutant dynamics follow partially known physical laws.

Applying EPDE to high-resolution MP_{10} time series revealed two distinct dynamical regimes—daytime turbulence and nocturnal stabilization—characterized by time-varying coefficients that reflect advection and diurnal forcing effects. The resulting data-driven equations can be summarized as

$$A(t)\frac{dx}{dt} + B(t) = \frac{dx}{dt}\cos\big(C(t)t + D(t)\big), \qquad t < T_1 \tag{1}$$

$$E(t)x + F(t) = \frac{dx}{dt}, \qquad t \geq T_2 \tag{2}$$

where $T_1 < T_2$ mark transitions between turbulent and stable atmospheric phases. Coefficients $A(t), E(t)$ describe transport balance, $B(t), F(t)$ represent emission–deposition terms, and $C(t), D(t)$ capture diurnal solar forcing.

Unlike fixed theoretical models, EPDE adaptively adjusts both the structure and parameters of equations to observed conditions. When evaluated over 72-hour horizons (Figure 3), the derived models reproduced realistic pollutant dynamics with competitive accuracy while maintaining full physical interpretability. Although EPDE

was not directly benchmarked against machine-learning baselines due to data-resolution limitations, its main contribution lies in revealing interpretable governing laws that explain diurnal pollutant behavior. Future work will focus on quantitative comparison and integration of symbolic discovery with AutoML pipelines.

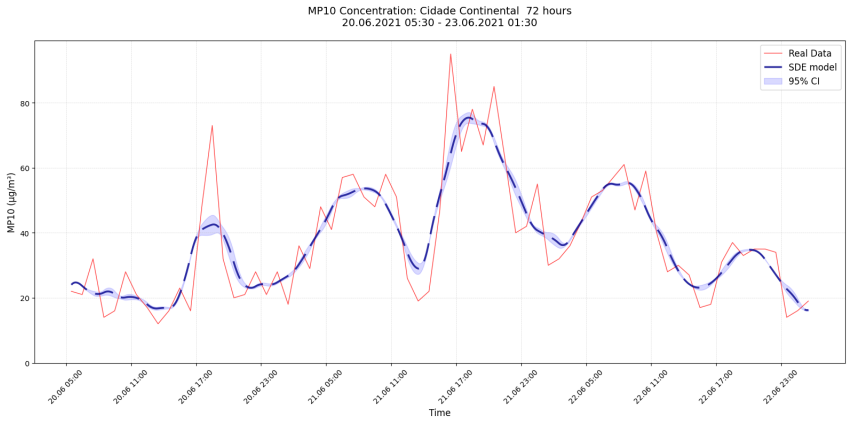


Figure 3. Validation of discovered governing equations predictive capasity for MP_{10} dynamics for 72 hours.

2.5 Zero-code LLM-based AutoML tools

FEDOT.ASSISTANT[1] is a multi-AutoML agent that employs large language models (LLMs) to translate natural-language task descriptions into executable configurations for supported AutoML frameworks. It handles classification, regression, multimodal analysis, and time-series forecasting. Beyond zero-code operation, FEDOT.ASSISTANT integrates into existing codebases, simplifying setup and enabling rapid framework switching, allowing researchers to focus on data analysis rather than implementation details.

Key advantages:

- **Integration with proven AutoML systems.** Unlike purely LLM-driven solutions that explore vast search spaces with limited convergence guarantees, FEDOT.ASSISTANT delegates optimization to established AutoML frameworks, ensuring efficient resource use and stable model quality.
- **Budget efficiency.** Demonstrates strong performance even with local open-source models by leveraging existing optimization pipelines.
- **Framework agnosticism.** Enables fast switching between frameworks and task types without code rewriting.

In our experiments, FEDOT.ASSISTANT used **FEDOT** as the AutoML backend with a 30-minute runtime limit and the `best_quality` preset. The underlying LLM was **Llama3.1:8B**, responsible for interpreting the task description shown in Listing 1.

Listing 1. Task description for air quality prediction (prompt).

```
This task focuses on air quality prediction using data from Brazilian air monitoring
    stations. Develop an accurate model to predict MP10 concentration. The goal is
   to create reliable predictions. Models are evaluated using MAE metric.
```

FEDOT.ASSISTANT automatically identified dataset structure, task type, evaluation metric, and configuration parameters, executing the full AutoML cycle—from pipeline composition to prediction generation—thus accelerating model prototyping and improving accessibility for researchers.

3 Experimental studies

The experiments were executed on: *Hardware* — Intel Core i5 (10th gen, 4 cores) with 16 GB RAM for classical/AutoML runs. *Software* — Python 3.12; scikit-learn 1.6.1; FEDOT 0.7.5; EPDE 1.2.12; NeuralForecast 3.0 (AutoNBEATS); Optuna 4.3; PyTorch 2.1. *Budgets* — FEDOT search ≤ 240 s per station; FEDOT.ASSISTANT 30 min per station (includes pipeline search). End-to-end for 13 stations $<$ 30–40 min on CPU for classical AutoML. Code and data are available in `https://github.com/ITMO-NSS-team/Air-quality-AutoMLs`.

[1] `https://github.com/AaLexUser/Fedot-assistant`

In addition, the reproducibility of results was improved by publishing all code and data in open repositories. To streamline this process, we applied the open-source OSA tool[2].

The study design is summarized below. For each monitoring station, the last 365 days (8,760 h) of data are reserved for testing, and the remaining records are used for training. The task is formulated as a one-hour-ahead forecast using a 48-hour input window. Model accuracy is evaluated by the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), computed per station and averaged across all thirteen stations.

- **Train/test split**: last 365 days (8,760 h) for testing; remainder for training.
- **Forecast horizon**: $h = 1$ h.
- **Context window**: $w = 48$ h.
- **Metrics**: MAE and RMSE (per-station and averaged).

This setup defines a short-term air-quality forecasting task applicable to operational monitoring systems.

3.1 Models under comparison

The evaluation includes both classical forecasting baselines and modern data-driven approaches. For the baseline methods, we consider simple yet effective strategies: the *Naïve* model, which uses the last observed value as the forecast ($\hat{y}_t = y_{t-1}$); the *Moving Average (MA-48)*, which averages the previous 48 observations; and *Linear Regression*, fitted via ordinary least squares using the 48 most recent lagged values as predictors.

A summary f the classical baselines are shown below:

1. *Naïve*: $\hat{y}_t = y_{t-1}$.
2. *Moving average (MA-48)*: $\hat{y}_t = \frac{1}{48}\sum_{i=1}^{48} y_{t-i}$.
3. *Linear regression*: ordinary least squares trained on the 48 past lags.

As more advanced alternatives, we evaluate several AutoML and deep learning solutions. This includes **FEDOT AutoML**, which evolves pipelines of the form "Lagged $\rightarrow$ Ridge" using genetic programming for hyperparameter optimization; **FEDOT.ASSISTANT**, a hybrid system that combines evolutionary search with pipeline suggestions generated by a large language model; **Chronos-T5-small**, a pre-trained time-series foundation model developed by Amazon Science, applied in a zero-shot setting without fine-tuning; and **AutoNBEATS**, a wrapper from the NeuralForecast library that uses Optuna for automated architecture search. Due to its high computational cost, AutoNBEATS is evaluated only on the four major urban monitoring stations. Together, these models represent a diverse set of paradigms, from traditional statistical methods to cutting-edge, scalable forecasting frameworks.

The AutoML and deep learning solutions are summarized as follows:

1. **FEDOT AutoML** — pipeline "Lagged $\rightarrow$ Ridge" with GP-based hyperparameter tuning.
2. **FEDOT.ASSISTANT** — hybrid evolutionary optimisation guided by a large language model.
3. **Chronos-T5-small** — pre-trained time-series foundation model evaluated in a zero-shot mode.
4. **AutoNBEATS** — NeuralForecast wrapper with Optuna search, evaluated on four urban stations (Ubu, Vila Velha-Centro, Vila Velha-IBES, and Vitória-Centro) due to long training time.

3.2 Aggregate results and Representative forecasts

Figure 4 presents the aggregated test results across the 13 monitoring stations. Among all approaches, **FEDOT.ASSISTANT**, **FEDOT**, and **Chronos** consistently outperform the classical baselines, achieving on average about 25,A paired Wilcoxon signed-rank test indicates that the performance differences among these three modern methods are not statistically significant at the 0.05 level. **AutoNBEATS** performs slightly worse in this setup, likely due to the short one-hour forecast horizon ($h = 1$) and a limited hyperparameter optimization budget.

Detailed MAE results for each station are provided in Table 1. Figure 5 illustrates representative one-step forecasts for the Vila Velha–Centro station, produced by the best-performing models.

3.3 Discussion

The results highlight key insights into the robustness, efficiency, and interpretability of different forecasting approaches for air-quality modeling.

- **Model robustness and generalization.** The top-performing models—**FEDOT.ASSISTANT**, **FEDOT**, and **Chronos-T5-small**—maintained MAE below 0.75 μg/m^3 for most stations, demonstrating strong cross-location generalization. Their stability across urban and rural environments confirms suitability for operational air-quality monitoring.

[2] https://github.com/aimclub/OSA

Mean test MAE and RMSE across all stations.
AutoNBEATS is averaged over four evaluated stations.

Model	MAE	RMSE
Moving average	8.192	10.696
Naïve	0.812	0.812
Linear regression	0.592	0.912
FEDOT AutoML	0.45	1.308
FEDOT.ASSISTANT	0.408	1.245
Chronos-T5-small	0.444	1.131
AutoNBEATS	4.01	5.648

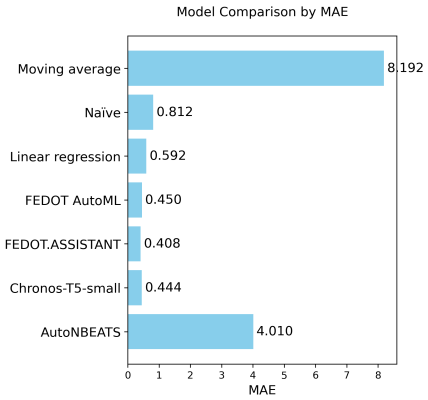


Figure 4. Average test-set MAE and RMSE across all monitoring stations (AutoNBEATS results averaged over four evaluated sites). Right: comparison of MAE values for different model classes.

Table 1. Mean Absolute Error (MAE, µg/m^3) for MP_{10} hourly prediction across monitoring stations. The best result per station is highlighted in bold.

Station	Linear	Naive	Chronos	FEDOT	FEDOT Assistant
Anchieta Centro	0.55	0.82	**0.42**	0.43	0.43
Carapina	0.62	0.87	**0.43**	**0.43**	**0.43**
Cariacica	0.72	1.07	0.52	0.53	**0.42**
Cidade Continental	0.85	1.07	0.63	0.59	**0.52**
Enseada do Suá	0.56	0.83	**0.34**	0.38	0.39
Jardim Camburi	0.44	0.54	0.32	0.32	**0.26**
Laranjeiras	0.69	0.93	0.50	0.51	**0.43**
Mãe-Bá	0.55	0.70	0.46	0.50	**0.44**
Meaípe	0.59	0.84	0.46	0.48	**0.40**
Ubu	0.59	0.82	0.49	0.48	**0.42**
Vila Velha Centro	0.51	0.70	0.39	0.39	**0.34**
Vila Velha IBES	0.58	0.73	0.46	**0.44**	0.46
Vitória Centro	0.46	0.63	0.37	0.36	**0.35**

- **Complexity–accuracy trade-off.** Linear regression served as a simple baseline but underperformed by about 25
- **Operational relevance.** All advanced models produced reliable short-term forecasts suitable for municipal warning systems. Their low training cost (30 min per station) and easy deployment make AutoML and pre-trained models practical for cities with limited technical resources, whereas complex deep-learning frameworks such as AutoNBEATS offered no clear advantage.
- **Physical interpretability.** Symbolic regression (EPDE) revealed diurnal MP_{10} regimes—turbulent evening and stable night phases—providing physically meaningful insights. Future hybrid approaches could merge symbolic and data-driven methods for improved accuracy and explainability.

Limitations include the current focus on univariate MP_{10} forecasting. Future work will extend the framework to multi-pollutant and meteorological variables and assess robustness under abrupt emission changes.

4 Conclusion

This study compared manual baselines, AutoML pipelines and symbolic regression for short-term PM_{10} forecasting across thirteen monitoring stations in Espírito Santo (Brazil). FEDOT and FEDOT.ASSISTANT as

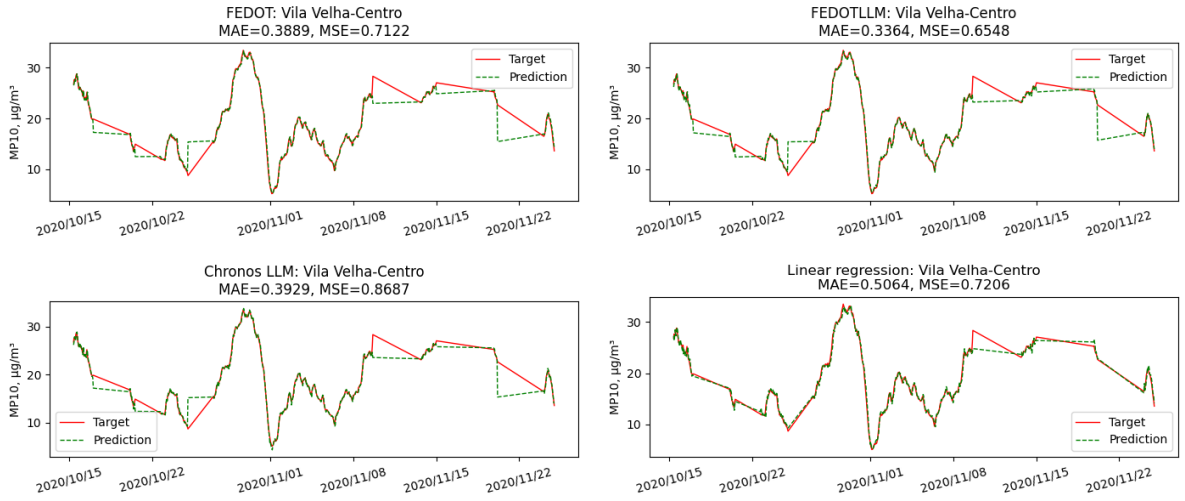


Figure 5. One-step MP_{10} forecasts for the Vila Velha-Centro station: FEDOT AutoML (the first), FEDOT.ASSISTANT (the second), Chronos-T5-small (the third), Linear regression (the fourth).

AutoML approaches achieved accuracy comparable to expert-designed models while reducing manual setup, and produced interpretable pipelines suitable for operational use. Pre-trained models - Chronos-T5-small provided strong zero-shot forecasts without fine-tuning, showing that foundation models can be practical when data or time are limited. EPDE as physics-informed insight yielded explicit equations that captured diurnal pollutant regimes, adding interpretability beyond standard ML metrics. All advanced approaches delivered MAE < 1 µg/m^3, sufficient for public-health alerts and deployment in heterogeneous urban environments what justifies practical impact of instruments research.

Future work will extend the pipeline to multi-pollutant settings, incorporate meteorological variables and evaluate robustness under abrupt emission changes.

Acknowledgements. The research was carried out within the state assignment of Ministry of Science and Higher Education of the Russian Federation (project No. № FSER-2024-0004). L.G. acknowledges the support of the funding agencies CNPq (grants 307688/2022-4, 409433/2022-5 and 304646/2025-3), Fapemig (grants APQ-02513-22, APQ-04458-23 and BPD-00083-22), and Finep (grant SOS Equipamentos 2021 AV02 0062/22).

References

[1] F. J. Kelly and J. C. Fussell. Air pollution and public health: emerging hazards and improved understanding of risk. *Environmental geochemistry and health*, vol. 37, n. 4, pp. 631–649, 2015.

[2] J. N. Sommar, U. A. Hvidtfeldt, C. Geels, L. M. Frohn, J. Brandt, J. H. Christensen, O. Raaschou-Nielsen, and B. Forsberg. Long-term residential exposure to particulate matter and its components, nitrogen dioxide and ozone—a northern sweden cohort study on mortality. *International Journal of Environmental Research and Public Health*, vol. 18, n. 16, pp. 8476, 2021.

[3] A.-A. Rus, S.-A. Pescariu, A.-S. Zus, D. Gaiţă, and C. Mornoş. Impact of short-term exposure to nitrogen dioxide (no2) and ozone (o3) on hospital admissions for non-st-segment elevation acute coronary syndrome. *Toxics*, vol. 12, n. 2, pp. 123, 2024.

[4] B. N. Aljafen, N. Shaikh, J. M. AlKhalifah, and S. A. Meo. Effect of environmental pollutants particulate matter (pm2. 5, pm10), nitrogen dioxide (no2), sulfur dioxide (so2), carbon monoxide (no) and ground level ozone (o3) on epilepsy. *BMC neurology*, vol. 25, n. 1, pp. 133, 2025.

[5] P. Thangavel, D. Park, and Y.-C. Lee. Recent insights into particulate matter (pm2. 5)-mediated toxicity in humans: an overview. *International journal of environmental research and public health*, vol. 19, n. 12, pp. 7511, 2022.

[6] S. C. Izah, L. Sylva, M. C. Ogwu, A. Shahsavani, S. Bazzazpour, and M. Rahmatinia. Modeling and statistical approaches for air pollution analysis. In *Air Pollutants in the Context of One Health: Fundamentals, Sources, and Impacts*, pp. 425–456. Springer, 2024.

[7] A. Adhikari and C. M. Hussain. From detection to solution: A review of machine learning in pm2. 5 sensing and sustainable green mitigation approaches (2021–2025). *Processes*, vol. 13, n. 7, pp. 2207, 2025.
[8] X. Liu, D. Lu, A. Zhang, Q. Liu, and G. Jiang. Data-driven machine learning in environmental pollution: gains and problems. *Environmental science & technology*, vol. 56, n. 4, pp. 2124–2133, 2022.
[9] I. Essamlali, H. Nhaila, and M. El Khaili. Supervised machine learning approaches for predicting key pollutants and for the sustainable enhancement of urban air quality: A systematic review. *Sustainability*, vol. 16, n. 3, pp. 976, 2024.
[10] M. Méndez, M. G. Merayo, and M. Núñez. Machine learning algorithms to forecast air quality: a survey. *Artificial Intelligence Review*, vol. 56, n. 9, pp. 10031–10066, 2023.
[11] M. Asgari, W. Yang, and M. Farnaghi. Spatiotemporal data partitioning for distributed random forest algorithm: Air quality prediction using imbalanced big spatiotemporal data on spark distributed framework. *Environmental Technology & Innovation*, vol. 27, pp. 102776, 2022.
[12] H. Y. Teh, A. W. Kempa-Liehr, and K. I.-K. Wang. Sensor data quality: A systematic review. *Journal of Big Data*, vol. 7, n. 1, pp. 11, 2020.
[13] T.-P. Nguyen. Aiot-based indoor air quality prediction for building using enhanced metaheuristic algorithm and hybrid deep learning. *Journal of Building Engineering*, vol. 105, pp. 112448, 2025.
[14] A. Masood and K. Ahmad. A review on emerging artificial intelligence (ai) techniques for air pollution forecasting: Fundamentals, application and performance. *Journal of Cleaner Production*, vol. 322, pp. 129072, 2021.
[15] de B. Silva, K. Champion, M. Quade, J.-C. Loiseau, J. Kutz, and S. Brunton. Pysindy: A python package for the sparse identification of nonlinear dynamical systems from data. *Journal of Open Source Software*, vol. 5, n. 49, pp. 2104, 2020.
[16] M. Maslyaev, A. Hvatov, and A. V. Kalyuzhnaya. Partial differential equations discovery with epde framework: application for real and synthetic data. *Journal of Computational Science*, vol. , pp. 101345, 2021.
[17] N. O. Nikitin, P. Vychuzhanin, M. Sarafanov, I. S. Polonskaia, I. Revin, I. V. Barabanova, G. Maximov, A. V. Kalyuzhnaya, and A. Boukhanovsky. Automated evolutionary approach for the design of composite machine learning pipelines. *Future Generation Computer Systems*, vol. , 2021.