Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

# Beyond model-specific biases: An explainable multifaceted approach for robust $PM_{10}$ source apportionment

ARTICLE INFO

*Keywords:*
$PM_{10}$ source apportionment
Machine learning
Feature importance
Multifaceted approach
Model interpretability

ABSTRACT

Liu et al. (2025) present an innovative approach to $PM_{10}$ source apportionment in urban environments by integrating Positive Matrix Factorization with machine learning (ML) models including XGBoost, Random Forest (RF), and Support Vector Machine (SVM). Their use of the Lung Performance Optimization (LPO) algorithm for XGBoost and 10-fold cross-validation improved model robustness, with the LPO-XGBoost variant achieving the highest predictive accuracy ($r^2 = 0.88$). SHAP values were employed to interpret feature importance, but concerns arise regarding the reliability of these rankings due to model-specific biases. Tree-based models may overemphasize features selected early in the decision process, while SVM models can obscure original feature relationships through kernel transformations. Although Liu et al. interpret variability in feature importance across models as analytical depth, this may reflect methodological inconsistencies rather than strength. SHAP values, being model-dependent, can inherit and amplify biases, complicating interpretation. In environmental research, where data are often noisy and high-dimensional, such instability can undermine the reliability of insights. Future studies should consider incorporating unsupervised learning techniques and non-parametric statistical methods to improve interpretability and robustness. Specifically, methods such as Feature Agglomeration (FA), Highly Variable Gene Selection (HVGS), Spearman's rho, and Kendall's tau can better capture complex and nonlinear associations, particularly in the context of health risk assessments. By integrating these approaches, researchers can enhance the stability of feature selection, reduce the influence of model-specific biases, and improve the transparency of analytical outcomes. A more systematic and cautious approach to model evaluation will ultimately strengthen reproducibility and support more informed environmental decision-making.

*Letter to the Editor:*

Liu et al. (2025) investigated the prediction and source apportionment of $PM_{10}$ concentrations in urban environments by integrating Positive Matrix Factorization (PMF) with several machine learning (ML) models, including Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost), raising several critical points that warrant further discussion. Their approach incorporated the Lung Performance Optimization (LPO) algorithm to fine-tune the XGBoost model's hyperparameters and utilized 10-fold cross-validation to enhance model robustness. The LPO-XGBoost variant achieved the highest predictive accuracy ($r^2 = 0.88$), outperforming RF and SVM. Feature importance was analyzed using SHapley Additive exPlanations (SHAP), offering a detailed interpretation of the factors influencing $PM_{10}$ distribution. However, the variation in feature importance across models suggests potential methodological biases, highlighting the need for cautious interpretation.

While Liu et al. made a valuable contribution to $PM_{10}$ assessment, their interpretation of feature importance derived from XGBoost, SVM, and RF models using SHAP introduces ambiguity. Although they evaluated predictive performance using metrics such as $r^2$, normalized RMSE, and normalized MAE, it is essential to distinguish between prediction accuracy and the reliability of feature importance rankings. According to existing literature, which includes over 300 peer-reviewed studies, high predictive accuracy does not necessarily ensure valid or consistent feature importance rankings (Fisher et al., 2019; Lenhof et al., 2024; Lipton, 2018; Musolf et al., 2022; Wood et al., 2024). Additional discussion and references can be found in the supplementary material.

XGBoost, like other tree-based models such as RF, tends to introduce bias in feature importance estimation due to its hierarchical structure, which often overrepresents features selected in early splits (Adler and Painsky, 2022; Alaimo Di Loro et al., 2023; Ugirumurera et al., 2024; Salles et al., 2021; Touw et al., 2013). This may distort the perceived relevance of environmental variables. Similarly, SVM models are prone to bias stemming from their use of kernel transformations, which can obscure the interpretability of original features (Faragalli et al., 2025).

Furthermore, while Liu et al. suggest that discrepancies in feature importance rankings among XGBoost, SVM, and RF models enhance the analytical depth of their study, this interpretation may benefit from further clarification. In Section 3.2, Performance comparison among models, they explain that each model was trained separately for individual pollution sources, and that differences in source-species relationships and PMF-derived concentration contributions contributed to variations in model performance. Although this rationale is understandable, interpreting such inconsistencies as a unique strength may overlook the implications of methodological divergence. The observed variability in feature importance across models could reflect underlying inconsistencies in identifying key predictors, which may affect the interpretability and reliability of the results. Given that each model

https://doi.org/10.1016/j.envres.2025.122656
Received 6 August 2025; Accepted 19 August 2025
Available online 22 August 2025

operates under distinct assumptions and algorithmic biases, a more systematic evaluation of feature stability could strengthen the robustness of the findings.

Additionally, although SHAP values are designed to clarify feature importance, these values can inherit biases from the underlying ML models and may, in certain cases, even amplify those biases (Bilodeau et al., 2024; Huang and Marques-Silva, 2024; Kumar et al., 2021). This dependency is evident in the formulation explain = SHAP(model), which directly ties the explanation to the model's output. As a result, SHAP's interpretability is inherently influenced by the model's internal mechanisms and assumptions, potentially leading to misleading conclusions. While predictive accuracy remains a central objective in ML, it is important to recognize that high accuracy does not necessarily imply trustworthy or stable feature importance estimates.

Validating feature importance in ML models is inherently difficult due to the absence of ground truth. Different algorithms, such as XGBoost and SVM, introduce model-specific biases that result in inconsistent rankings. This challenge is particularly evident in Liu et al.'s study, where high-dimensional and collinear features complicate interpretation and increase the risk of overfitting. In environmental research, where data are often noisy and complex, such instability can undermine the reliability of model-derived insights. Moreover, SHAP values, while widely used for interpretability, are directly dependent on model outputs and may inherit or even amplify existing biases. When combined with tree-based models like XGBoost, which tend to overemphasize features used in early splits, this can exacerbate interpretability issues. Therefore, claims regarding the identification of predictive features using such pipelines should be approached with caution, and future studies would benefit from more robust validation strategies.

To ensure accurate interpretations in health risk assessment, a robust analytical framework is essential. This approach should incorporate methodologies better suited for capturing complex associations within health data, such as unsupervised learning techniques including Feature Agglomeration (FA) and Highly Variable Gene Selection (HVGS) (Zhang et al., 2020; Xie et al., 2025). Additionally, non-parametric statistical methods like Spearman's rho or Kendall's tau would be highly beneficial (Okoye and Hosseini, 2024; Yu and Hutson, 2024). These methods can detect various types of relationships, offering enhanced interpretability crucial for translating findings into actionable clinical insights. Ultimately, this multi-faceted approach is indispensable for generating accurate, reproducible, and clinically relevant insights that can truly advance health risk assessment.

In conclusion, Liu et al. proposed a promising framework for $PM_{10}$ source apportionment using ML and SHAP-based interpretation. However, their study presents several methodological challenges, particularly regarding the stability of feature importance, model-specific biases, and interpretability. Environmental datasets are often noisy and complex, which makes it essential to adopt analytical strategies that can reliably distinguish meaningful patterns from random variation. Future research should consider incorporating unsupervised learning techniques and non-parametric statistical methods to improve robustness and interpretability. By applying a more comprehensive and cautious approach to model evaluation, researchers can enhance the reproducibility, transparency, and utility of their findings for informing environmental decision-making.

## CRediT authorship contribution statement

**Souichi Oka:** Writing – original draft. **Takuma Yamazaki:** Investigation. **Yoshiyasu Takefuji:** Project administration, Supervision, Writing – review & editing.

## Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Declaration of competing interest

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

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envres.2025.122656.

## Data availability

No new data were generated or analyzed in support of this research.

## References

Adler, A.I., Painsky, A., 2022. Feature importance in gradient boosting trees with cross-validation feature selection. Entropy 24, 687. https://doi.org/10.3390/e24050687.

Alaimo Di Loro, P., Scacciatelli, D., Tagliaferri, G., 2023. 2-step gradient boosting approach to selectivity bias correction in tax audit: an application to the VAT gap in Italy. Stat. Methods Appl. 32, 237–270. https://doi.org/10.1007/s10260-022-00643-4.

Bilodeau, B., Jaques, N., Koh, P.W., Kim, B., 2024. Impossibility theorems for feature attribution. Proc. Natl. Acad. Sci. 121, e2304406120. https://doi.org/10.1073/pnas.2304406120.

Faragalli, A., et al., 2025. Do machine learning methods solve the main pitfall of linear regression in dental age estimation? Forensic Sci. Int. 367, 112353. https://doi.org/10.1016/j.forsciint.2024.112353.

Fisher, A., Rudin, C., Dominici, F., 2019. All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. J. Mach. Learn. Res. 20, 177. https://doi.org/10.48550/arXiv.1801.01489.

Huang, X., Marques-Silva, J., 2024. On the failings of Shapley values for explainability. Int. J. Approx. Reason. 171, 109112. https://doi.org/10.1016/j.ijar.2023.109112.

Kumar, I., Scheidegger, C., Venkatasubramanian, S., Friedler, S., 2021. Shapley residuals: quantifying the limits of the Shapley value for explanations. Adv. Neural Inf. Process. Syst. 34, 26598–26608.

Lenhof, K., Eckhart, L., Rolli, L.M., Lenhof, H.P., 2024. Trust me if you can: a survey on reliability and interpretability of machine learning approaches for drug sensitivity prediction in cancer. Brief. Bioinform. 25, bbae379. https://doi.org/10.1093/bib/bbae379.

Lipton, Z.C., 2018. The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery. ACM Queue 16, 31–57. https://doi.org/10.1145/3236386.3241340.

Liu, Y., et al., 2025. Source apportionment of PM10 particles in the urban atmosphere using PMF and LPO-XGBoost. Environ. Res., 121659 https://doi.org/10.1016/j.envres.2025.121659.

Musolf, A.M., Holzinger, E.R., Malley, J.D., Bailey-Wilson, J.E., 2022. What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics. Hum. Genet. 141, 1515–1528. https://doi.org/10.1007/s00439-021-02402-z.

Okoye, K., Hosseini, S., 2024. Correlation tests in R: pearson cor, Kendall's tau, and Spearman's rho. In: Okoye, K., Hosseini, S. (Eds.), R Programming: Statistical Data Analysis in Research. Springer Nature, pp. 247–277. https://doi.org/10.1007/978-981-97-3385-9_12.

Salles, T., Rocha, L., Gonçalves, M., 2021. A bias-variance analysis of state-of-the-art random forest text classifiers. Adv. Data Anal. Classif. 15, 379–405. https://doi.org/10.1007/s11634-020-00409-4.

Touw, W.G., Bayjanov, J.R., Overmars, L., et al., 2013. Data mining in the life sciences with random forest: a walk in the park or lost in the jungle. Brief. Bioinform. 14, 315–326. https://doi.org/10.1093/bib/bbs034.

Ugirumurera, J., Bensen, E.A., Severino, J., Sanyal, J., 2024. Addressing bias in bagging and boosting regression models. Sci. Rep. 14, 18452. https://doi.org/10.1038/s41598-024-68907-5.

Wood, D., Papamarkou, T., Benatan, M., Allmendinger, R., 2024. Model-agnostic variable importance for predictive uncertainty: an entropy-based approach. Data Mining Knowl. Discovery 38, 4184–4216. https://doi.org/10.1007/s10618-024-01070-7.

Xie, Y., Jing, Z., Pan, H., Xu, X., Fang, Q., 2025. Redefining the high variable genes by optimized LOESS regression with positive ratio. BMC Bioinf. 26, 104. https://doi.org/10.1186/s12859-025-06112-5.

Yu, H., Hutson, A.D., 2024. A robust Spearman correlation coefficient permutation test. Commun. Stat. Theory Methods. 53, 2141–2153. https://doi.org/10.1080/03610926.2022.2121144.

Zhang, J., Wu, X., Hoi, S.C.H., Zhu, J., 2020. Feature agglomeration networks for single stage face detection. Neurocomputing 380, 180–189. https://doi.org/10.1016/j.neucom.2019.10.087.

Souichi Oka[a,*], Takuma Yamazaki[a], Yoshiyasu Takefuji[b]

[a] *Science Park Corporation, 3-24-9 Iriya-Nishi Zama-shi, Kanagawa, 252-0029, Japan*

[b] *Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo, 135-8181, Japan*

[*] Corresponding author.

*E-mail addresses:* souichi.oka@sciencepark.co.jp (S. Oka), tyamazaki@sciencepark.co.jp (T. Yamazaki), takefuji@keio.jp (Y. Takefuji).