3ème Conférence Internationale sur les Sciences Appliquées et l'Innovation (CISAI-2025)
Proceedings Book Series –PBS- Vol 25, 22-30

Predicting of non-exhaust PM1 emissions under real driving conditions

Tasnim Harizi [1], Ines Belkacem [2,3], Ali Helali [4,5], Salah Khardi [6], Khalifa Slimi [2]

[1] PhD student, Higher institute of transport and logistics of Sousse, University of Sousse. Email,

tassnimharizi12@gmail.com

[2] Laboratory of Thermal and Energy Systems Studies (LESTE), Monastir, National Engineering school of Monastir, University of Monastir, Ibn Eljazzar Street, 5019, Tunisia Email: khalifa_slimi@yahoo.fr

[3] Department of logistics, Higher Institute of Transport and Logistics, University of Sousse, Sousse, Tunisia ,

Email: ines.belkacem2014@gmail.com

[4] Mechanical Laboratory of Sousse, National Engineering School of Sousse, University of Sousse, Tunisia.

[5] Department of Transport Technology and Engineering, Higher Institute of Transport and Logistics, University of Sousse, Tunisia.

Email:alilehlali@gmail.com

[6] Environment, Planning, Safety and Eco-design Laboratory, Department of Mobility, Plan- ning and the Environment, Gustave Eiffel University, University of Lyon - Lyon, France

Email: salah.khardi@univ-eiffel.fr

Abstract

Urban air pollution, particularly from non-exhaust particle emissions, is an increasingly critical issue due to its adverse effects on public health and environmental quality. This study focuses on modeling of non- exhaust particle mass concentrations (PMC), which induced particularly from four sources; known as brake wear, tire abrasion, road surface fric- tion, and road dust resuspension. Measurements were obtained using a GRIMM Aerosol Spectrometer during real-world urban driving condi- tions in the city of Lyon, France. To predict PM_1 concentrations, three machine learning models—Random Forest (RF), Support Vector Regres- sion (SVR), and K-Nearest Neighbors (KNN)—were applied using $PM_{2.5}$ and alveolic particle data as input features. Model performance was eval- uated based on standard statistical metrics. Results show that the SVR model provided the best predictive performance, which characterized by a high coefficient of determination R2=0.93, and lower error tends to- wards zero. This study confirms the effectiveness of artificial neural net-

work approaches for accurately modeling urban non-exhaust emissions and offers valuable tools for urban air quality management in order to en- sure a sustainable mobility.

Keywords: Modeling, Machine learning, non-exhaust emissions, air quality, KNN, SVR, RFR

1 Introduction

Air pollution remains one of the most pressing environmental challenges of the 21st century, with serious consequences on public health, ecosystems, and the climate. Among various pollutants, particulate matter (PM)—notably PM1, PM2.5, and PM10—has attracted growing attention due to its significant health ef- fects and its influence on climate forcing (Silva et al., 2020; Belkacem et al., 2020). Exposure to fine particles has been associated with increased morbidity and mortality, particularly from cardiovascular and respiratory diseases (Belkacem et al., 2022).

In the road transport sector, PM emissions originate from both exhaust and non-exhaust sources. Exhaust particles result from incomplete fuel combustion, while non-exhaust particles come from four sources such as;brake wear, tire abra- sion, road surface wear, and road dust resuspension caused by vehicle movement (Amato et al., 2014; Grigoratos& Martini, 2020). Although strict regulations have contributed to reducevehicle exhaust emissions,while, the non-exhaust

ISSN: 2961-6611

emissions (NEE) are in continuous elevation and not yet regulated for ultrafine particles,due to urban traffic growth and the proliferation of heavier electric vehicles (OECD, 2020; Amato, 2018).

Modeling of PM emissions is essential for several reasons: First, for evaluating the exposure level and designing effective mitigation strategies. Second, it can in- form urban planners and policymakers about the impacts of traffic on air quality. This can lead to better zoning laws, traffic management strategies, and the promo- tion of sustainable transportation options (e.g., public transit, cycling, walking). Third, contribute to broader climate change mitigation efforts. Traditional disper- sion models such as AERMOD or OSPM are often used but are limited by simpli- fied assumptions that can introduce uncertainties in urban environments (Mishra et al., 2022). Recent developments in machine learning (ML) have offered promising alternatives, allowing for better modeling of complex and nonlinear relationships between several variables (eg. meteorological variables, vehicle characteristics and driver behavior) and PM concentrations (Silva et al., 2020; Khan et al., 2023).

Numerous studies have demonstrated the effectiveness of ML techniques— such as Random Forests, Gradient Boosting, and deep learning models—in predicting PM2.5 and PM1 concentrations with high accuracy in urban contexts (Kumar et al., 2022; Wang et al., 2021). Furthermore, hybrid approaches that inte- grate satellite-derived data like aerosol optical depth (AOD) with ground-based measurements and ML algorithms have improved spatial estimation of PM, espe- cially in areas with limited air quality monitoring networks (Li et al., 2023). The objective of this study is to forecast vehicle non-exhaust particle mass concentra- tions PMCs, using three different methods of Artificial Neural Networks, known as; Random Forest Regression (RFR), Support Vector Regression (SVR) and K- nearest neighbor (KNN), basing on continuous real-world measurements. Little studies interestedto modelthe non-exhaust particulate emissions using the ANN (e.g. Belkacem et al., 2022). The exactitude of the ANNs depends on both the quantity and quality of data. Adjusted variables greatly taken into consideration in the prediction of each ANN model, such as the number of neighbors, neighbor weighting and the distance metric for KNN model.

PMCs are collected on urban road of Lyon, France, using GRIMM analyser se- ries 1.108 and modeled by the ANN. The performance of the proposed ANN s models were assessed by several statistical parameters; Mean absolute deviation (MAD), Root mean squared error (RMSE), the coefficient of determination (R^2) and mean squared error (MSE). Finally, this research tends to compare the effi- ciency of the three ANN chosen models.

The remainder of this paper is structured as follows; the next section describes the followed methodology for estimating the vehicle non-exhaust PMCs. Then, section 3 presents the most important results and discussion. Finally, some con- cluding remarks are presented in the latest section.

2. Sampling and Measurement Strategies

In July 2019, a field measurement campaign was conducted near Lyon, France, to investigate the non-exhaust particulate emissions under real-world driving con- ditions. These particulate emissions were produced mainly from four sources, e.g., road dust resuspension, tire and brake wear, and road surface abrasion and meas- ured under realdriving conditions (Belkacem et al.,2022).

Real-time particle mass concentration measurements were carried out during real driving conditions (driver behaviors, road traffic conditions), with distances ranging from 32 to approximately 75 km. Sampling was conducted either in the morning or evening, depending on the weather conditions (for more details on ex- perimental data, see in Belkacem et al., 2022).

3. Study Area

Lyon is the second-largest metropolitan area and the third most populous city in France — is in the southeast-central region of the country, within the Auvergne- Rhône-Alpes region. The city, with a population of 515,695 inhabitants (INSEE, 2016), is intersected by a dense urban road network. This makes its peripheral roadways a suitable and strategic site for evaluating the environmental impact of high-speed traffic. During the campaign, a 172 km route was driven exclusively on urban road, offering a representative context to analyze emissions under condi- tions characterized by high speeds and continuous vehicle flow.

4. Methodology

ISSN: 2961-6611

This study started with the collection of the instantaneous vehicle speed and non-exhaust particulate matter (PM1) emissions using a V7 Pro GPS device and a GRIMM Series 1.108 Aerosol Spectrometer, respectively (Belkacem et al., 2022).

ANN is an efficient tool to identify complex patterns and non linear relation- ships in large datasets, making them well suited for modeling dynamic environ- mental systems such as air quality. (Nielsen 2015).

Three ANN models, known as SVR, KNN and RFR were trained for predicting the PM1using the same input variables (PM2.5 and Alveolic). The main objective was to evaluate and compare the predictive performance of the Random Forest Regressor, Support Vector Regression (SVR) and Kneighbors Regressor models against actual PM1 GRIMM measurements. The training process was set when the model error tends towards 0 and the coefficient of determination (R2) tends to 1. A schematic representation of the study's methodology is provided in Figure 1.

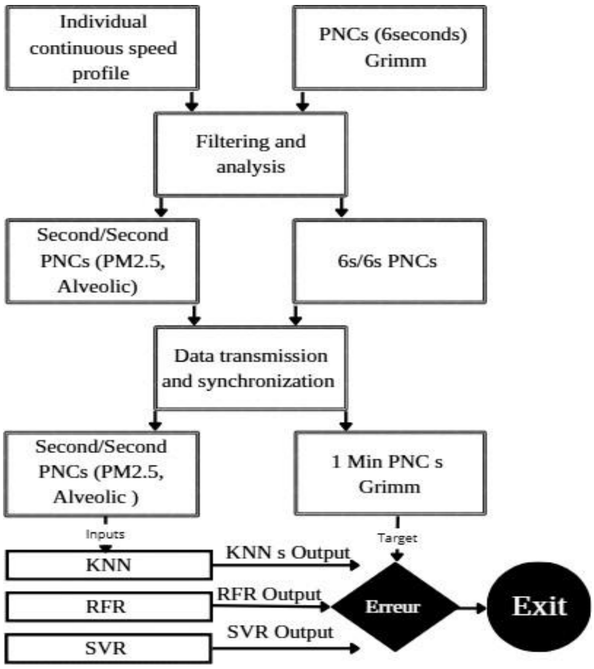

Figure 1. Flow chart methodology of non-exhaust emissions

In this study, several statistical parameters were used in order to evaluate the performance of the model; such as: the mean absolute error (MAE), root mean square error (RMSE) and correlation coefficient (R)

$$\mathbf{MAE} = \frac{\sum_{i=1}^{n} |\mathbf{Target}_i - \mathrm{Input}_i|}{\mathbf{n}} \quad (1)$$

$$\mathbf{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\mathbf{Target}_i - \mathrm{Input}_i)^2}{\mathbf{n}}} \quad (2)$$

$$\mathbf{MSE} = \frac{1}{n}\sum_{i=1}^{n} (\mathbf{Target}_i - \mathbf{Output}_i)^2 \quad (3)$$

$$\mathbf{MAD} = \frac{\sum|\mathbf{Target}_i - \mathbf{Output}_i|}{\mathbf{n}} \quad (4)$$

Where (i) = 500

ISSN: 2961-6611

5. Results and Discussion

5.1. Random Forest Regressor

The Random Forest Regressor (RFR) algorithm requires the specification of two key hyperparameters: the number of trees in the ensemble (ntree), typically set to a default value of 500, and the number of predictors randomly selected at each split (mtry). According to Anurag et al. (2023), these parameters generally have only a limited impact on overall model performance. In this study, the RF model was applied to predict PM1 concentrations using PM2.5 and Alveolic as input variables. While the default value for ntree was retained, the optimal mtry was selected using the tune RF() function from the random Forest package in R. Further optimization involved tuning three additional hyperparameters: the number of trees (n_estimators), the maximum depth of the trees (max_depth), and the minimum number of samples required to split an internal node (min_samples_split). A grid search procedure was employed to identify the best combination of these parameters, resulting in an optimal configuration of n_estimators = 100, max_depth = 10, min_samples_split = 10, and min_samples_leaf = 2. This setup produced strong predictive performance, achieving a coefficient of determination (R^2) of 0.8793 and a relatively low root mean square error RMSE of 1.13 (Fig 2). This result indicate that the (RFR) model an efficient tool to establish the relationship between the PM1 and both PM2.5 and alvealic. The increase of the number of trees beyond a certain point yields minimal improvements, as the model's performance tends to stabilize (Anurag et al., 2023).

5.2. K-Nearest Neighbors

The K-Nearest Neighbors (KNN) algorithm is non-parametric and instance- based learning method. It was employed to predict PM1 concentrations using PM2.5 and Alveolic as input variables. As described by Altman (1992) and Coo- mans and Massart (1982), KNN operates by identifying the K nearest samples in the training dataset based on a specified distance metric and estimates the output for a test sample by averaging the target values of these neighbors often using a distance-weighted approach to give higher importance to closer points. In this study, model tuning focused on selecting the optimal number of neighbors (n_neighbors), a parameter that critically influences both accuracy and robustness. The best performance was achieved with n_neighbors = 5, yielding a coefficient of determination (R^2) of 0.8819 and the RMSE of 1.58, along with the lowest ob- served MAE during validation (Fig 3). While this indicates a good level of predic- tive accuracy, the model displayed sensitivity to local variations in the data, espe- cially during peak events. This overfitting tendency is characterized of KNN's

instance-based nature, which does not construct a global model, but instead relies on the distribution of nearby data points.

5.3. Support Vector Regression (SVR)

Support Vector Regression (SVR) is a machine learning technique grounded in the principle of structural risk minimization (Drucker et al. 1996). It leverages kernel functions to map input data into a high-dimensional feature space, where a linear regression is performed. This approach allows SVR to model complex and non-linear relationships. The SVR has demonstrated strong performance in nu- merous regression tasks, particularly in environmental modeling and air quality prediction (Ayodeji and Liu, 2018). In this study, SVR was employed to estimate PM1 concentrations using PM2.5 and Alveolic as input variables. The model's ef- fectiveness largely depends on the proper tuning of three key hyperparameters: the regularization parameter (C), the epsilon margin (ε), and the kernel coefficient (γ). After optimization, the best performance was achieved with C = 100, epsilon =0.2, and gamma = 0.1. Under this configuration, SVR achieved an R^2 of 0.9357 and a low RMSE of 0.31(table 1)

Table 1. Statistical parameters performance of RFR, SVR and KNN models

MODEL	R^2	MSE	RMSE	MAD
Random Forest	0.8793	1.2969	1.1388	0.7241
SVR	0.9357	0.0943	0.3071	0.2507

ISSN: 2961-6611

KNN	0.8819	2.5011	1.5815	1.0568

ISSN: 2961-6611

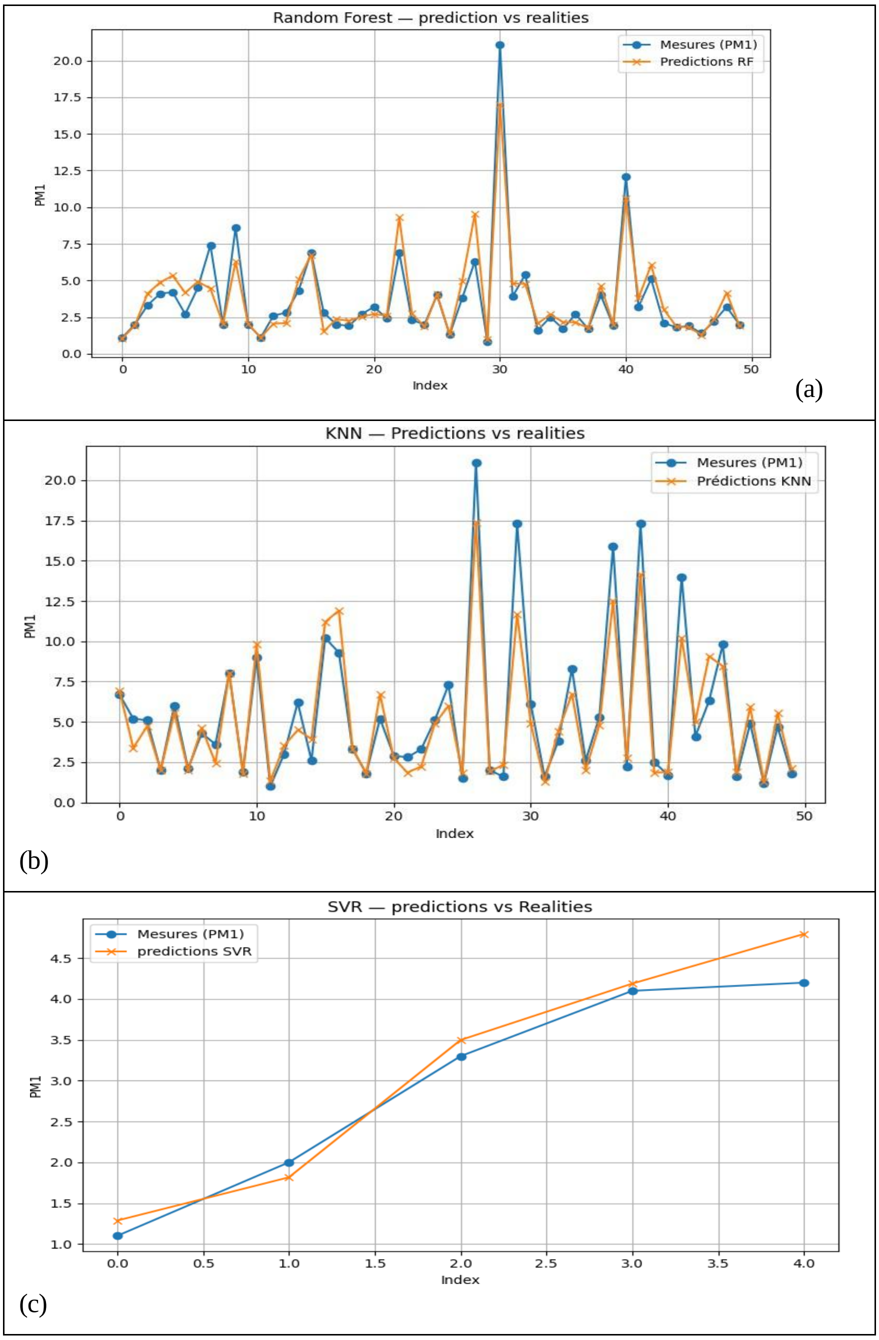

Figure 2. prediction of non-exhaust PM1 with SVR, KNN and RFR models

ISSN: 2961-6611

This study aimed to model PM1 concentrations using two input variables: PM2.5 and the Alveolic fraction. An initial correlation analysis revealed a strong linear relationship between PM1 and both predictors, indicating that PM2.5 and Alveolic are reliable proxies for estimating PM1. This justified their use as input features in the regression models.
Three machine learning algorithms were evaluated: K-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Regression (SVR). The KNN model, which relies on the proximity between observations, achieved a coefficient of determination (R^2) of 0.8819, reflecting good explanatory power. However, its error metrics (RMSE = 1.58; MAE = 1.05) revealed a tendency to overfit local fluctuations, particularly during peak pollution events—an inherent limitation of instance-based models that do not generalize underlying patterns. The RF model delivered similar R^2 performance (0.8793) but achieved a lower RMSE (1.13), highlighting its strength in capturing nonlinear relationships through ensemble learning and decision tree aggregation. SVR, on the other hand, outperformed both KNN and RF with a significantly higher R^2 of 0.9357 and the lowest RMSE of 0.31, demonstrating strong generalization capabilities and better handling of com- plex patterns in the data.

Conclusion

This study has highlighted the relevance and effectiveness of machine learning approaches in predicting vehicle non-exhaust particulate emissions under real- world conditions. By applying and comparing three artificial neural network-based models—Random Forest, Support Vector Regression, and K-Nearest Neighbors— the research demonstrated the ability of these algorithms to model the complex re- lationships between PM concentrations and related variables. Among the tested models, one method clearly outperformed the others in terms of accuracy and gen- eralization, showing strong potential for practical applications in environmental monitoring.
The findings support the use of data-driven techniques as valuable tools for ad- dressing current limitations in traditional modeling approaches. As urban traffic and the contribution of non-exhaust sources continue to rise, accurate predictive models are crucial for informing public health decisions and guiding sustainable urban planning strategies. The integration of such advanced modeling techniques can enhance air quality management and contribute to the development of targeted mitigation policies.
Further studies could explore the incorporation of additional variables, such as meteorological data or traffic patterns, and extend the modeling framework to oth- er geographic areas to assess its venerability in the context of macroscopic varia- bles. These approaches highlight the value of advanced data-driven tools in sup-porting public health policies and air pollution control in order to ensure sustaina- ble environment and mobility.

References

Silva, L. F., Schneider, I. L., Artaxo, P., Núñez-Blanco, Y., Pinto, D., Flores, É. M., ... & Dotto, G. L. (2022). Particulate matter geochemistry of a highly industrialized region in the Caribbean: Basis for future toxicological studies. *Geoscience Frontiers*, *13*(1), 101115.

Belkacem, I., Khardi, S., Helali, A., Slimi, K., & Serindat, S. (2020). The influence of ur- ban road traffic on nanoparticles: Roadside measurements. *Atmospheric Environ- ment*, *242*, 117786.

Belkacem, I., Helali, A., Khardi, S., & Slimi, K. (2022). Investigations on vehicle non- exhaust particle emissions: real-time measurements. International Journal of Environ- mental Science and Technology, 19(12), 11749-11762.

Wu, H., Wang, T., Wang, Q. G., Cao, Y., Qu, Y., & Nie, D. (2021). Radiative effects and chemical compositions of fine particles modulating urban heat island in Nanjing, Chi- na. *Atmospheric Environment*, *247*, 118201.Amato, F. (Ed.). (2018). *Non-exhaust emissions: an urban air quality problem for public health; impact and mitigation measures*. Academic Press.

Amato, F., Cassee, F. R., Van Der Gon, H. A. D., Gehrig, R., Gustafsson, M., Hafner, W.,
... & Querol, X. (2014). Urban air quality: the challenge of traffic non-exhaust emis- sions. *Journal of hazardous materials*, *275*, 31-36.

ISSN: 2961-6611

Grigoratos, T., & Martini, G. (2014). Non-exhaust traffic related emissions. Brake and tyre wear PM. *Report EUR*, *26648*.

OECD (2020), Non-exhaust Particulate Emissions from Road Transport: An Ignored Envi- ronmental Policy Challenge, OECD Publishing, Paris, https://doi.org/10.1787/4a4dc6ca- en.

Ayodeji, A., & Liu, Y. K. (2018). Support vector ensemble for incipient fault diagnosis in nuclear plant components. *Nuclear Engineering and Technology*, *50*(8), 1306-1313.

Altman, N. S.: An introduction to kernel and nearest-neighbor nonparametric regression, Am. Stat., 46, 175–185, 1992.

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. J., and Vapnik, V. N.: Support vector regression machines, in: Advances in Neural Information Processing Systems 9, NIPS 1996, MIT Press, 155–161, 1997.

ISSN: 2961-6611