Road Traffic Noise Predictions by means of L_{10} Modelling with a Multilinear Regression Calibrated on Simulated Data

Domenico Rossi*, Aurora Mascolo, Claudio Guarnaccia
Department of Civil Engineering,
University of Salerno,
Via Giovanni Paolo II, 132, Fisciano, I-84084,
ITALY

*Corresponding Author

Received: February 26, 2023. Revised: April 15, 2023. Accepted: May 23, 2023. Published: June 27, 2023.

Abstract— Estimation of road traffic noise is fundamental for the health of people living in urban areas, and it is usually assessed based on field-measured data. Real data may not always be available, anyway, and for this reason, predictive models play an important role in the evaluation and controlling of the noise impact. In this contribution, the authors present a multilinear regressive model calibrated on simulated noise levels instead that on real measured ones, correlating percentile noise levels to independent traffic variables. The model efficiency is then evaluated on two field measurement datasets by analyzing data statistics and error metrics. Results show that the model provides good results in terms of mean error (less than 1 dBA on average) even if slight underestimations and overestimations are present. The presented model, then, can be used to assess the impact of road traffic noise anytime field measurements are not available, or even predict it when designing new road infrastructures.

Keywords—Noise assessment, Road traffic noise, Percentile Levels, Multilinear regression.

I. Introduction

ENVIRONMENTAL impact of the road network is a relevant topic in an urban context reconciliation because these infrastructures affect human health, especially in relation to air pollution and road traffic noise. Actually, continuous noise exposure has a large impact on people's quality of life. Traffic noise in particular is linked to many health conditions such as high blood pressure, hearing loss, cardiovascular problems, etc., [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12]. According to World Health Organization studies, about 30% of the population of the European Union is exposed to day and night average traffic noise levels of 55 or more dBA, [13]. Therefore, the evaluation of noise pollution nearby existing or planned road infrastructure must be executed. This assessment can be achieved both by a wide range of experimental activities and by software simulations. In particular, the mathematical modelling of the sound sources and of the propagation in the area under study needs to be very accurate, in order to provide reliable results. The development of Traffic Noise prediction Models (TNMs) began in the 50s', with the usage of statistical approaches, calibrating predictive formulas on big datasets collected with field measurements, [14], [15]. A comprehensive review of the main literature approach can be found in [16], [17], covering the most used TNMs so far, while in [18], the authors reported a resume of some innovative approaches, such as cellular automata, [19], machine learning techniques, [20], stochastic models, [21], among the others. The main inputs of regressive and statistical models usually are traffic flow, vehicle type, the distance between source and receiver, and sometimes, speed. Moreover, many additional parameters can be taken into account in the TNM, as coefficients in the predictive formula or as additive correction terms, [22], [23], [24], [25], [26], [27], [28], [29], [30].

In this paper, a multilinear regressive model is presented and calibrated on simulated data rather than on field-measured ones. Based on the original idea published in a work of [31], the here presented model calculates $L_{10,h}$ values and then investigates the correlation between the independent variables and $L_{10,h}$ values by a multilinear regression technique. After the calibration, a test is performed on two real field measurement datasets, in order to estimate the performances of the model and to conclude about the possible extension of this approach on different case studies, also variating the simulated calibration dataset. The results will show that such an approach is effective, especially in the standard traffic flow and speed conditions, also compared to other literature models, making the model suitable for road traffic noise assessment, particularly when field measurements are not available.

II. Material and Methods

The aim of the work is to computationally generate a model for road traffic noise assessment and to validate it. The first step of this work is to calibrate the multilinear regressive model to be applied to the simulated dataset. Following the work of [31], this calibration is performed on a simulated dataset, built with the aim of including several possible traffic conditions. The variables chosen are the hourly traffic flow *Q* [veh/h], the percentage of heavy vehicles - generally defined by a weight greater than 3.5 tons – *P* [%], the distance between the lane center and the receiver *d* [m], and the mean speed of the flow *v* [km/h]. Through a random generation function, then, a dataset of 200 rows is obtained, and for each of the rows $L_{eq,h}$ is calculated using the formula provided in [31], deprived of some terms which were not useful for this specific application:

$$L(t') = 10\log_{10}\left[\sum_{i=1}^{Q} 10^{0.1\left[L_{wi}\left(v_i(t')\right)\right]}\left(\frac{d_0^2}{\left(\frac{1}{3.6}\int_{t'}^{t_{0i}} v_i(t)dt\right)^2 + d_i^2 + (\Delta h)^2}\right) + \left(\frac{I_e}{I_0}\right)\right] \quad (1)$$

In this formulation L_{wi} is the single vehicle source power level, as a function of the speed v_i, d_0 is a reference distance used for the computation of L_{wi}, according to the REMEL approach, [32]. Δh is the height of the receiver, I_e and I_0 indicate, respectively, the background noise intensity and the hearing threshold. Namely, in the formulation here proposed, a degree of asymmetry between the source and receiver is neglected with respect to the original formulation proposed in [31]. The hourly equivalent level L_{eq} so simulated is then converted to L_{10} using the formula suggested by the UK regulation, [26]:

$$L_{eq,h} = 0.94 \times L_{10,\mathrm{h}} + 0.77 \quad (2)$$

The choice to transform the L_{eq} in L_{10} follows the approach used and allows us to perform a comparison on the dataset provided in [24], as well as to compare the model's results with the prediction provided by the UK model CoRTN, [33]. It is important to note that the chosen equivalent level is functional to the needs of the research and that it can be changed by simply implementing the formulas on the model itself.

The final database to be calibrated, then, is made of 200 rows times 5 columns corresponding to independent variables: *Q*, *v*, *d*, *P*, and a column of simulated $L_{eq,h}$. The multiple regression between all the independent variables with $L_{eq,h}$ values is the core of the calibration of the model itself. Multilinear regression has been performed in Python by using *numpy*, *pandas*, and *statsmodel* packages, respectively for numerical analysis, dataset creation and management, and linear regression technique application. Linear regression, in detail, has been applied with the Ordinary Least Squares method. The goodness of the linear regression is assessed by some statistical parameters: residual standard error, Multiple R squared, and adjusted R squared *p*-value. Considering previous observations (data not shown) and also given the remarks of [31], the linear regression technique has been observed to have validity within fixed ranges of values of the independent variables. The details about the ranges of the variables are reported in Table 1.

TABLE 1. Range of variation of the variables

Variable	Lower limit	Upper limit
Q [veh/h]	10	1000
d [m]	5	200
V [km/h]	25	75
P [%]	0	20

Once the dataset has been built and the multilinear coefficients have been estimated, a second linear regression is performed, in order to tune the slope coefficients of each of the four variables and of the overall function X, defined as in formula (3), where $C_{(Q)}$, $C_{(V)}$, $C_{(P)}$ and $C_{(d)}$ are the slopes of the multilinear regression. Please note as *P* and *d* variables are added to a positive integer value to avoid a negative argument of the logarithm.

$$X = C_{(Q)} Log_{10} Q + C_{(V)} V + C_{(P)} Log\left(P+5\right) + C_{(d)} Log(d+15) \quad (3)$$

The intercept and slopes of the new linear regression are used to compute the final predictive formula of the model proposed in this paper as in equation (4). In this formula, C_X is the slope of linear regression.

$$L_{10,sim} = Intercept_X + C_X X \quad (4)$$

The so-built model has been validated on two datasets coming from different sites, in order to check the generality of the approach and to highlight possible limitations of the presented approach.

The validation of the model has been evaluated on real field measurements coming from two different sites. The first site is a straight carriageway in Patiala (India), without any significant slope, where road traffic noise has been recorded during good weather conditions. Available data refer to 84 hourly L_{eq} levels, together with values of *Q*, *v*, and *P*. The second dataset includes measurements taken from the Long Term Monitoring Station project of Universitè Gustave Eiffel, France. Such a dataset is made of more than 30000 15-minute L_{eq} levels. For the scope of this work, the authors selected a limited range of values between August and September 2007, converting them into hourly L_{eq} values. In order to perform this conversion, the authors selected only the hours that included all the required inputs to feed the model and that presented no missing data in the L_{eq} column. This selection reduced the dataset to 100 hours.

The model has been validated on both datasets by simulating the $L_{10,h}$ from the independent variables and then comparing it with the measured ones. The goodness of the model is assessed by analyzing the error distributions and by evaluating two main statistical parameters of the errors: the *MAE* (Mean Absolute Error) and the *RMSE* (Root Mean Square Error).

III. Results and Discussion

The simulated model is calibrated by a multilinear regression, to assess the contribution of each independent variable to the composition of the $L_{10,h}$. The results of the multilinear regression are listed in Table 2.

TABLE 2. Results of the multilinear regression

	Variable	Estimated value	Standard deviation
Coefficients	$Log_{10}(Q)$	10.4535	1.49938
	V	0.0348	0.04062
	$Log_{10}(P+5)$	3.4105	5.92760
	$Log_{10}(d + 15)$	-1.9515	6.33512
Intercept		44.26899	3.11606

It is possible to notice how Q is the most affecting variability of the model, meaning that the equivalent noise level profoundly changes when changing the number of passing vehicles at a given time. Moreover, the only negative correlation is the one of the distance, reflecting the well-known phenomenon of the reduction of the noise level at increasing distance from the noise emitter. Multilinear regression has been evaluated with different parameters, listed in Table 3. Figure 1 visualizes the regression of all the independent variables in the multilinear regression model.

TABLE 3. Goodness fit parameters

Residual standard error	R-squared	Adjusted R-squared	p-value
1.335	0.9214	0.9198	$< 2.2*10^{-16}$

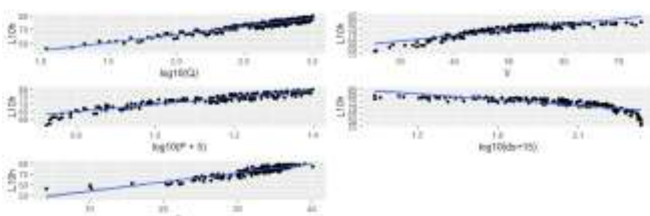

Figure 1. Multilinear regression results. In each plot of the figure the correlation between a single independent variable and L10,sim is shown.

The second step of the calibration is a new linear regression between $L_{10,sim}$, and the overall function X, as described in formula (3). Results and statistic parameters of such linear regression are listed in Table 4, and lead to the final expression of the X function as in (5):

$$X = 10.4535 Log_{10}(Q) + 0.0348(V) + 3.4105 Log_{10}(P+5) - 1.9515 Log_{10}(d+15) \quad (5)$$

TABLE 4. Results of second the linear regression

		Estimated value
Coefficients	Intercept	45.5075
	Slope	0.8510
Regression parameters	Residual standard error	2.013
	R squared	0.8185
	Adjusted R squared	0.8176
	p-value	$<2.2*10^{-16}$

Analysis of residuals indicates that the model has a mean residual value basically null ($2.33*10^{-16}$), with a standard deviation of 1.32. Skewness and Kurtosis are respectively -0.84 and -0.49. In Figure 2 the distribution of the population of such residuals is presented.

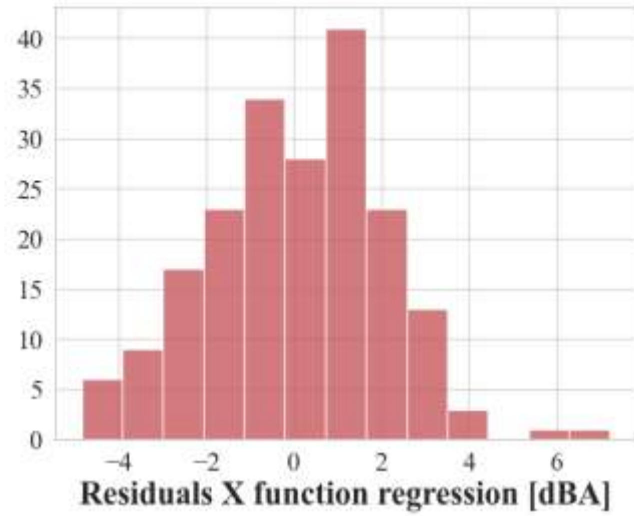

Figure 2. Distribution of residuals of X function regression

Once the second linear regression has been performed, the final value of $L_{10,h}$ is simulated by using formula 4), resulting in the following:

$$L_{10,h} = 45.5075 + 8.8960 \log_{10}(Q) + 0.0296(V) + 2.9024 \log_{10}(P+5) - 1.6608 \log_{10}(d+15) \quad (6)$$

by which the model is validated through two different datasets of real field measurement, by comparing the measured levels of $L_{10,h}$, and the one simulated with the model. The levels simulated with the proposed model are plotted versus the corresponding measured values in Figure 3, to check the overall performances in terms of overestimation and underestimation. For the Saint Berthevein dataset, the authors observed a slight overestimation, since all the points are above the bisectrix of the graph. For the Patiala dataset, on the contrary, the model seems to systematically underestimate the real value, since all the points fall below the bisectrix. These over- and underestimations may be due to the specificities of the two datasets, and the detailed study will be the object of further work. Overall, anyway, the error metrics indicate a Mean Error value lower than 1 dBA (0.82 and 0.93 for Saint Berthevin and Patiala respectively). Interestingly, the other models considered for comparison have higher mean errors (1.27 and 1.72 for CORTN, 2.53 and 0.96 for the model proposed by [24]). All values of error metrics can be found in Table 5. As the last validation analysis, we analyzed the violin plot distribution of predicted equivalent values for all the models, as shown in Figure 4 and Figure 5. Graphs show how the proposed model fits the measured values for the Patiala dataset, only missing the highest portion of the distribution. As for the Saint Berthevin dataset, the violin distribution of the data well replicates the one of the measured data, whereas other models don't, or have a mean value significantly shifted from the real one.

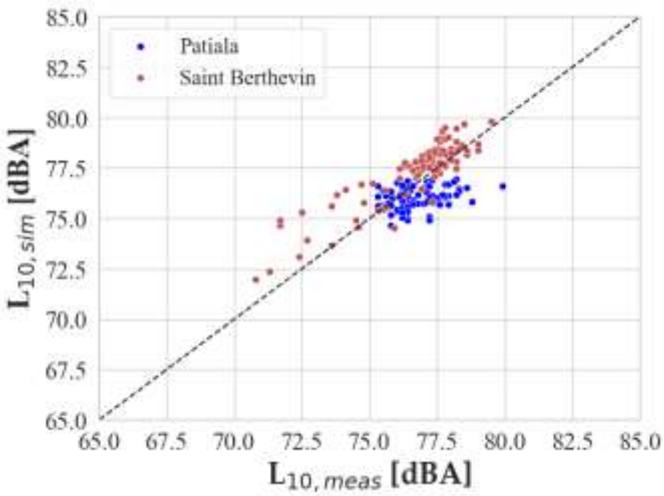

Figure 3. Scatter plot of simulated data versus the corresponding measurements for Patiala (blue circles) and Saint Berthevin (red circles) datasets

TABLE 5. Error metrics of all the compared models

	Patiala dataset		Saint Bertheven dataset	
	MAE [dBA]	RSME [dBA]	MAE [dBA]	RSME [dBA]
Presented model	0.93	1.16	0.82	1.04
CoRTN	1.72	1.97	1.27	1.50
Model of [31]	0.96	1.19	2.53	2.78

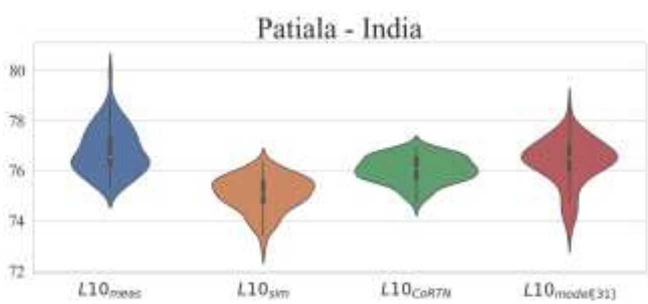

Figure 4. Distribution of $L_{10,h}$ values measured in Patiala compared with the distributions simulated with the proposed model, with the CoRTN model, and with the model proposed in [31].

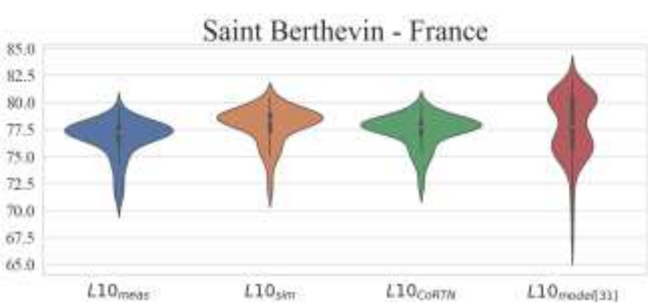

Figure 5. Distribution of $L_{10,h}$ values measured in Saint Bertheven compared with the distributions simulated with the proposed model, with the CoRTN model, and with the model proposed in [31]

IV. Conclusions

In the here presented contribution a model for predicting $L_{10,h}$ levels produced by road traffic, based on four independent variables, is presented. The calibration of the model performed on simulated – and not measured – data makes the model feasible to be used even if field measurements are not available. The calibration of the model involves two linear regression techniques, one multivariate, and one univariate, and the final prediction formula is obtained by combining the resulting coefficients. Comparison with the other two literature models for noise assessment is performed in the last part of the paper, indicating the presented model is well-performant on average, even if slight overestimations and underestimations have been found in the two datasets. All in all, the here presented model provides good predictions of road traffic noise, with an error always lower than 1 dBA, and fits experimental data better, on average, than other commonly used methods.

Acknowledgments

The authors acknowledge the support of Dolores Durante for writing the first code in the R-project and performing the data analysis during her bachelor thesis work. This research was partially supported by the Sustainable Mobility Center (MOST) - Spoke 7: Connected Networks and Smart Infrastructure (PNRR - Missione 4 Componente 2 Investimento 1.4 - project No. CN 00000023).

References

[1] W. Babisch, Road traffic noise and cardiovascular risk, Noise Health, 10, 27-33 (2008).

[2] T. Bodin, M. Albin, J. Ardö., E. Stroh, P. Östergren and J. Björk, Road traffic noise and hypertension: results from a cross-sectional public health survey in southern Sweden, Environmental Health, 8, 1-10 (2009).

[3] F.J. Langdon, Noise nuisance caused by road traffic in residential areas: Parts I and II, Journal of Sound and Vibration, 47, 243-282 (1976).

[4] E. Petraitis, M. Pranskeviciu, R.L. Idzelis and P. Vaitiekunas, Predictive modelling of environmental noise levels in Lithuanian urban areas, Environmental Engineering and Management Journal, 10, 1935-1941 (2011).

[5] J. Selander, M.E. Nilsson, G. Bluhm, M. Rosenlund, M. Lindqvist, G. Nise and G. Pershagen, Longterm exposure to road traffic noise and myocardial infarction, Epidemiology, 20, 272-279 (2009).

[6] K. D. Kryter Annoyance from aircraft and ground vehicle noise, Journal of the Acoustical Society of America, 72, 1222-1242 (1982).

[7] C. Tomozei, A. Astolfi, V. Nedeff and C. Lazar, Noise sources characterization inside and outside a factory, Environmental Engineering and Management Journal, 11, 701-708 (2012).

[8] Y. Cai, R. Ramakrishnan and K. Rahimi, Long-term exposure to traffic noise and mortality: A systematic review and meta-analysis of epidemiological evidence between 2000 and 2020, Environmental Pollution, 269, 116222 (2021)

[9] J. O. Klompmaker, G. Hoek, L. D. Bloemsma, A. H. Wijga, C. van den Brink, B. Brunekreef, E. Lebret, U. Gehring and. N.A.H. Janssen, Associations of combined exposures to surrounding green, air pollution and traffic noise on mental health, Environment international, 129, 525-537 (2019)

[10] E. Murphy and E.A. King, Environmental noise pollution: Noise mapping, public health, and policy, Elsevier, 2022

[11] R. Biel, Rita, C. Danieli, M. Shekarrizfard, L. Minet, M. Abrahamowicz, J. Baumgartner, R. Liu, M. Hatzopoulou and S. Weichenthal, Acute cardiovascular health effects in a panel study of personal exposure to traffic-related air pollutants and noise in Toronto, Canada, Scientific Reports, 10(1), 1-12 (2020)

[12] A. Seidler, J. Hegewald, A.L. Seidler, M. Schubert and H. Zeeb, Is the whole more than the sum of its parts? Health effects of different types of traffic noise combined, International journal of environmental research and public health, 16(9) 1665 (2019)

[13] F. Ibili, E.K. Adanu, C. Adams, S.A. Andam-Akorful, S.S. Turay and S. A. Ajayi, T raffic noise models and noise guidelines: A review, Noise & Vibration Worldwide, 53(1-2), 65-79 (2022)

[14] R. B. Ranpise and B.N. Tandel, Urban road traffic noise monitoring, mapping, modelling, and mitigation: A thematic review, Noise Mapping, 9(1), 48-66 (2022)

[15] WHO, Transport-related health effects with particular focus on children, (World Health Organization Geneva, Switzerland, 2004)

[16] J. Quartieri, N.E. Mastorakis, G. Iannone, C. Guarnaccia, S. D’ambrosio, A. Troisi and T.L.L. Lenza “A review of traffic noise predictive models” in Recent Advances in Applied and Theoretical Mechanics, 5th WSEAS International Conference on Applied and Theoretical Mechanics (MECHANICS’09)

[17] N. Garg and M. Sagar "A critical review of principal traffic noise models: Strategies and implications." in Environmental Impact Assessment Review 46 (. Puerto De La Cruz, Tenerife, Canary Islands, Spain December 2014) pp. 68-81

[18] C. Guarnaccia, Advanced Tools for Traffic Noise Modelling and Prediction, WSEAS Transactions on Systems, Issue 2, 12, 121-130 (2013)

[19] J. Quartieri, N.E. Mastorakis, G. Iannone and C. Guarnaccia, C Cellular automata application to traffic noise control, in 12th WSEAS International Conference on Automatic Control, Modelling and Simulation, ACMOS '10, 2010, pp. 299-304

[20] D. Singh, A.B. Francavilla., S. Mancini., C. Guarnaccia, Application of machine learning to include honking effect in vehicular traffic noise prediction, Applied Sciences, 11(13), 6030 (2021)

[21] C. Guarnaccia, J. Quartieri, C. Tepedino and E.R. Rodrigues, A time series analysis and a non-homogeneous Poisson model with multiple change-points applied to acoustic data, Applied Acoustics, 114, 203-212 (2016).

[22] I. D. Griffiths and F.J Langdon, Subjective response to road traffic noise, Journal of Sound and Vibration 8, 16-32 (1968).

[23] Calculation of Road Traffic Noise, London, United Kingdom Department of Transport Welsh, HMSO (1988)

[24] M. A. Burgess, Noise prediction for urban traffic conditions, related to measurements in Sydney Metropolitan Area, Applied Acoustics, 10, 1-7 (1977).

[25] G. B. Canelli, K. Gluck, S.A. Santoboni, A mathematical model for evaluation and prediction of mean energy level of traffic noise in Italian towns, Acustica, 53, 31-53 (1983).

[26] Department of Transport, (1988), Calculation of Road Traffic Noise, HMSO, UK.

[27] Centre Scientifique et Technique du Batiment, Etude theorique et experimentale de la propagation acoustique, Revue d’Acoustique n.70 (1991),

[28] W. M. To, C. Rodney, W. Ip, A multiple regression model for urban traffic noise in Hong Kong, The Journal of the Acoustical Society of America, 112, 551-556 (2002).

[29] Harmonoise, (2004), Imagine project funded by EC under the sixth framework program, On line at: http://www.imagine-project.org

[30] E. C. Paz, and P.H.T. Zannin, Urban daytime traffic noise prediction models, Environmental Monitoring and Assessment, 163, 515-529 (2010).

[31] S. Afandizadeh, and H. Gharehdaghli. A new steady-state traffic noise model for estimating L10 (h) on free flow roads using Reference Energy Mean Emission Levels Building and Environment 196 107685 (2021).

[32] R. L. Wayson, T.W.A. Ogle and W. Lindeman, Development of Reference Energy Mean Emission Levels for Highway Traffic Noise in Florida, Transportation Research Record 1416, 82–91 (1993).

[33] P. G. Abbott and P. M. Nelson, Converting the UK traffic noise index LA10,18h to EU noise indices for noise mapping, Project Report PR/SE/451/02, (2002).

Contribution of individual authors to the creation of a scientific article (ghostwriting policy)

Conceptualization: D. Rossi, C. Guarnaccia
Data curation: D. Rossi, A. Mascolo, C. Guarnaccia
Methodology: D. Rossi, C. Guarnaccia
Software: D. Rossi, A. Mascolo, C. Guarnaccia
Supervision: C. Guarnaccia
Visualization: D. Rossi, A. Mascolo
Writing - original draft: D. Rossi
Writing - review & editing: all the co-authors

Sources of funding for research presented in a scientific article or scientific article itself

This study was carried out within the MOST – Sustainable Mobility National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1033 17/06/2022, CN00000023). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.