DOI: 10.61782/fa.2025.0566

MICROSCOPIC AND DYNAMIC ROAD TRAFFIC NOISE MODEL COUPLED WITH A VEHICLE RECOGNITION AND TRACKING VIDEO ANALYSIS TOOL

Domenico Rossi[1] **Antonio Pascale**[2] **Aurora Mascolo**[1]
Margarida C. Coelho[2] **Claudio Guarnaccia**[1]
[1] Department of Civil Engineering, University of Salerno, Italy
[2] Department of Environment and Planning / Centre for Mechanical Technology and Automation (TEMA), University of Aveiro, Portugal & LASI - Intelligent Systems Associate Laboratory, Portugal

ABSTRACT

The usage of road traffic noise models to assess the contribution of traffic flow to environmental noise is a common procedure. However, it is often complicated to obtain accurate input data for model calibration, development, and validation.

This study presents a method for gathering key input parameters for road traffic noise modelling, including vehicle count, classification, and speed, through video recordings of traffic on a national road. The method is based on a custom Python script that employs image processing techniques to determine the position and speed of vehicles while categorizing them as either light or heavy, refining a previous work developed by some of the authors.

The collected data are integrated with the CNOSSOS-EU noise emission model to estimate sound power levels of each vehicle. The pressure level at the receiver is then calculated by summing the contribution of all the vehicles within the influence range, as a function of time. This allows us to reproduce the time history of the road traffic noise signal and to compare it with the sound level meter recording, as well as to calculate the equivalent level on any time basis, including those required by regulation.

*Corresponding author: drossi@unisa.it

Keywords: *road traffic noise, video analysis, vehicle tracking, dynamic microscopic noise model, field measurement validation*

1. INTRODUCTION

The European Environment Agency estimated that 75% of the population resides in cities and urban areas, with projections indicating an increase to 84% by 2050 [1]. Concurrently, transportation activities are expected to grow significantly, exceeding 6000 billion passenger-kilometers travelled annually. As a direct consequence, more than one-quarter of the European population is exposed to harmful noise levels daily, primarily generated by road traffic, highlighting the need for noise mitigation strategies [1-3]. Noise mitigation can be considered a sustainable goal in urban development as it significantly improves environmental quality while also generating socioeconomic effects. For instance, studies have shown a correlation between urban noise levels and average housing rental prices [4]. More broadly, noise exposure has a severe impact on public health, contributing to a substantial burden of disease (deprivation of sleep, lack of concentration, hearing loss, elevated blood pressure) [5-6], as quantified on Disability-Adjusted Life Years, and diminishing overall quality of life [7]. In this framework, the importance of a precise assessment of noise is very clear. Two ways for the assessment of road traffic noise can be implemented: direct measurement of noise levels or their simulation. While direct measurement provides precise assessment, its implementation is not always feasible or economically viable. Moreover, it is impractical during the design phase of new road infrastructure. Noise levels can be estimated using Road Traffic Noise Models (RTNM), which are well-established in the literature. Among the various models, it is worth mentioning CoRTN, SonRoad, NMPB, RLS90 and Harmonoise. In response to the European Noise Directive and to meet the need for a standardized procedure for noise level evaluation and noise mapping, the CNOSSOS-EU model was developed [8-10]. A comprehensive description of principal RTNMs, their peculiarities, and the history of their development can be found in [11]. Generally speaking, RTNMs require input data in terms of traffic volumes by categories, average traffic flow speed, and traffic conditions (e.g., free flow or congestion scenarios). The quality of these input data plays a critical role in the performance of RTNMs, highlighting the need for precise data collection methods.

In recent years, automated procedures for collecting traffic volumes and speed data have been developed, alongside the widespread use of dedicated sensors in urban areas. These stations, often installed for other purposes (public security, private areas monitoring), can provide a large amount of data, often in real-time, which can simultaneously serve as input data for RTNMs. Moreover, when these stations have the potential to record noise levels, they offer the opportunity to provide ground truth value for models' validation. For noise level monitoring in urban cities, stations equipped with cameras turn out to be appealing since input data for RTNMs can be extracted from their records. Furthermore, video analysis techniques offer the ability to extract individual vehicle speed, providing a higher quality of data compared to average speed information, which could potentially improve RTNMs estimations. An approach in this sense was provided in [12] where the EAgLE – Equivalent Acoustic Level Estimator was proposed for the first time, in a highway environment. This procedure was successively improved in terms of vehicle detection and speed estimation, leading to EAgLE 2.0, and applied in a different case study [13]. In a recent publication [14], the case study presented in [12] was revisited by enhancing the model to the EAgLE 3.0 version, with an improved video-tracking procedure for vehicle identification. This improvement was driven by the integration of Artificial Intelligence techniques, together with a more sophisticated and modern code approach. The present contribution aims to apply EAgLe 3.0 to the case study proposed in [13] and to estimate the noise time history (sound pressure level versus time). The main novelty of the procedure is related to the detection of the trajectories of each vehicle. In such a way the noise contribution at the fixed receiver can be calculated as a function of time, providing a much finer estimation of the overall noise contribution and paving the way for a real-time application of the model.

2. MATERIALS AND METHODS

2.1 Case study

Data on road traffic noise levels and vehicle volumes are the same as those reported in [13] and were collected along a segment of the N235 National Road in Aveiro, Portugal, where the speed limit is 80 km/h. The N235 serves as a key transportation corridor connecting two major cities in central Portugal, Aveiro and Coimbra. The monitoring site was established on an 8-meter-high bridge (40º 36' 53'' N, 8º 38' 5'' W), where a Braun Germany Champion 4K camera and a Class 1 sound level meter (Rion NL-52) were deployed. Both instruments were mounted on a tripod positioned 1.2 m above the bridge surface, corresponding to a total height of 9.2 m above the road infrastructure. The camera was configured to record video at a resolution of 1280 x 720p at 30 frames per second and was oriented to capture vehicle trajectories in both traffic directions. Prior to data collection, the sound level meter was calibrated using a reference signal of 94 dB at 1000 Hz. A windscreen was placed over the microphone to mitigate potential wind interference, although wind speeds during the measurement period remained below 5 m/s. A-weighted sound pressure levels were recorded at 100 ms intervals. Data collection was conducted on a clear, sunny weekday between 11:00 AM and 1:00 PM, encompassing both the transition from off-peak to peak traffic hours. The research team continuously monitored the instruments and traffic conditions throughout the data acquisition period. No anomalies were observed, and no extraneous noise sources other than vehicular traffic were detected. Following data acquisition, the two-hour video was split into 24 5-minute videos, that were analysed to extract vehicle trajectories, specifically the instantaneous speed of each vehicle within a designed region of interest. These trajectories were subsequently used to estimate the noise level history, defined as the temporal variation of the sound pressure level, which was then compared with the signal recorded by the sound level meter, serving as the ground truth reference.

2.2 Implementation of the model: software and resources

The procedure here described has been implemented in Python programming language (version 3.13.2), and consists of four different parts, as reported in Figure 1.

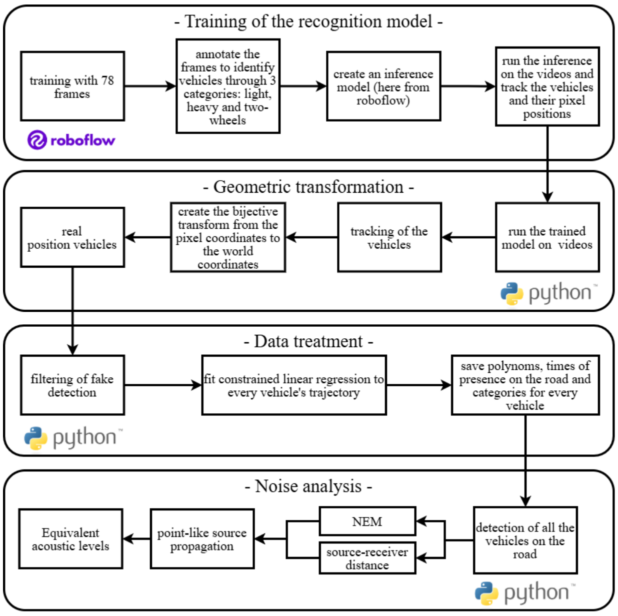

Figure 1: Workflow of the experimental procedure.

As for the training of the recognition model, Python communicates with Roboflow's API thanks to Roboflow's libraries supervision 0.19.22 and inference 0.9.22. Roboflow is a set of computer vision datasets used to train models for automatic recognition and tracking of objects [15-16]. It works easily within the Python environment and takes advantage of models like YOLO for object identifications. In the here described procedure, YOLO 11 [17] has been used for the vehicles' identification. The geometric transformation part has been performed with the *cv2* package of Python. The data treatment and the noise analysis parts have been performed with the following packages: pandas, *cv2*, NumPy, matplotlib, and bottleneck. The entire procedure was performed on a desk computer with the following specifications: 3.50 GHz CPU, 16 GB of RAM, 64-bit operating system.

2.3 Training of the recognition model

The Roboflow platform was used to train the model used for the recognition of the vehicles on video. Following the instructions on the Roboflow documentation, a subset of frames has been chosen out of the 24 videos. In detail, 78 frames have been selected from the 24 videos, containing 134 vehicles, divided into three categories: light-duty vehicles (121), heavy vehicles (9), and two wheels (3). Such apparent unbalance of the dataset is due to the real presence of many light-duty vehicles on the recorded section of the road, where very few heavy vehicles were present. Please note that the authors decided to treat medium vehicles as light ones. To train the model, such a subset of pictures has been split into three different sections: the training part (68 images), the validation part (6 images), and the test part (4 images). Pictures have been automatically annotated by assigning to each vehicle a category: such annotation has been manually validated by the authors. Once all the pictures had been correctly annotated, brightness adjustments and size rescaling were performed on all of them. Then the dataset has been used to train the model: YOLO11 has been chosen for object recognition. Figure 2 shows a picture used for training the model, with two light vehicles and a heavy vehicle correctly annotated.

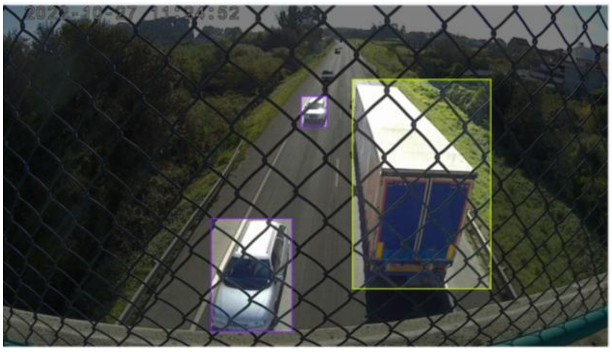

Figure 2: Annotation of pictures for model training.

2.4 Geometric transformation

A geometric transformation has been implemented to convert images from pixel-measured distances to real-world ones. To do that, the authors followed a procedure described in the documentation of Roboflow's, but completely implemented offline using Python.

The idea is that, by using a homography matrix H, it is possible to pass from the image plane (O_1,x,y) to the real-world plane (O_2,X,Y) [18].

$$\begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} = H \times \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \qquad (1)$$

The existing homography (equation 1) matrix has been applied by using the built-in function of the *cv2* Python package. Of course, real-world coordinates must be furnished to permit the correct transformation: such coordinates have been retrieved from the measuring tool of Google Maps, as reported in Figure 3.

Figure 3: Google Maps® image used for the generation of the homography matrix and the identification of the sensitive area for the vehicles' recognition.

As visible in Figure 3, tracking procedure of the vehicles identified by Roboflow's trained YOLO model has not been performed on the whole image, but only on a sensitive area. The choice of such sensitive areas was arbitrary, but it was mainly driven by two reasons. First, farther vehicles from the video recording device were more frequently misrecognized by the model, as the number of pixels representing the object was reduced in number. Additionally, far vehicles contribute less to the equivalent noise level compared to closer ones. Secondly, it has been considered the average permanence time of vehicles within the sensible area (given the average speed according to the limits on the analysed road segment). Since the speed limit on the road is fixed at 80 km/h, a sensible area of 40 meters in length has been chosen. Considering a vehicle exactly running at 80 km/h, i.e. about 22.2 m/s, it would cover a 40-meter distance in about 1.8 s. Since the videos have been recorded at 30 frames per second, the vehicles would stay in the sensitive areas for about 54 frames. This amount is enough to ensure a good tracking of the vehicles itself and also the subsequent smoothing procedure to build the trajectory and estimate the speed. Furthermore, in an extreme and unlikely scenario, such as a vehicle traveling at 130 km/h, it would remain in the sensitive areas for about 33 frames, a duration that is still considered sufficient.

2.5 Data treatment

The tracking procedure resulted in a list of positions of any vehicle flowing into the sensitive area. Nonetheless, some fake detections were present, mainly due to the accuracy of the model. In order to remove such data, the list has been filtered based on two principles. First, vehicles tracked for less than 8 frames were excluded. This threshold has been chosen referring to the same principle of the choice of the sensitive area. The remaining data have been used to estimate positions and speeds of the vehicles. The estimated speeds were then additionally filtered, as will be reported in the results section, providing the final dataset to be used for computing the noise contribution of each vehicle.

2.6 Noise estimation

The sound power level of each vehicle has been computed referring to the CNOSSOS-EU noise emission model: speed (*v*) and category of the vehicles have been the input variables. According to the model, then, the power level of each single vehicle has been computed considering the rolling and propulsion contributions to the noise power level. The power levels of each vehicle are then propagated at the receiver (sound level meter). According to the classic point-like source formula, i.e. free-field propagation, the pressure levels have been computed as in equation 2.

$$L_{p,j}(t) = L_{W,j} - 20Log_{10}(d_j) - 10Log_{10}(4\pi) \qquad (2)$$

with j indicating the j-th vehicle and d_j the distance between the vehicle and the sound level meter. The overall noise was then obtained by summing the contribution of all the vehicles in the sensitive area at the time t:

$$L_p(t) = 10Log_{10}\left(\sum_j 10^{\frac{L_{p,j}(t)}{10}}\right) \qquad (3)$$

Finally, the equivalent noise level can be computed in the time interval under analysis.

3. RESULTS AND DISCUSSIONS

3.1 Vehicle detections and tracking generation

After training the Roboflow model for vehicle detections, it has been applied to all 24 videos, each 5 minutes in length. The results of automatic recognition have been filtered according to the specifications described in Section 2 and reported in Table 1, together with the manual counting of the vehicles [13]. Roboflow's recognition model works quite well in identifying flowing vehicles and assigning them the correct category. In the column "Diff" the total difference is shown between the two counting procedures. Anyway, it is also visible how Roboflow greatly fails in the vehicles' recognition in two videos, namely videos 2 and 7, highlighted in italics in Table 1. Roboflow's procedure did not work properly on these videos because they have some corrupted frames. Since, as described in Section 2, vehicle recognition and counting are performed by frames, when a frame is corrupted the process stops, only reporting the number of vehicles identified and tracked up to the last uncorrupted frame. For this reason, the authors removed video 2 and video 7 from the subsequent analysis.

The existing discrepancies in the other videos mean that some vehicles are not properly recognized and tracked. The main reason may be found in the presence of the metallic grid in the videos, which makes the process difficult. Although Roboflow's model has been trained with grid images, the inference model tracks objects once they have been recognized, following their trajectory over time. Each time the vehicle image passes through a section of the metallic grid, the tracking model can lose the object itself. This is particularly true for distant objects, that are smaller than the near ones, as will be shown later. Missing vehicles could also be due to very high speeds not permitting a proper number of frames over which carry the recognition.

3.2 Real-world trajectories and speed

Using the homography matrix described in Section 2, the pixel coordinates of the vehicles are transformed in real-world coordinates, to obtain trajectories.

For the same reasons impairing the vehicles' recognition, trajectories generation was disturbed by the presence of the grid. In the reconstruction of vehicles' trajectories, some points abruptly diverge from the actual straight trajectory for some frames. When reporting in real-world coordinates, such shifts are interpreted as sudden and non-realistic movements, like the vehicle moving instantaneously from one carriage to another or moving backwards for a few frames or jumping forward of meters.

Basically, the Roboflow detection model cannot assure a fluid and proper following of the object when it passes behind the grid metal. This leads to errors in reconstructing the proper trajectory and, consequently, in the calculation of the speed. Speed, in fact, is calculated as the finite derivative of Y coordinate over time, being Y the axis parallel to the road direction. Figure 4 reports the speed of 4 randomly sampled vehicles tracked by the model. The speeds have sudden variations, due to the problems in tracking mentioned above, suggesting the need for a smoothing process. The greatest variations are observed for low Y coordinates, i.e. far from the bridge, highlighting a lower accuracy of the tracking algorithm in that region of the sensitive area. Since a free flow was observed, i.e. no congestion occurred during the measurement campaign, the authors performed a linear regression on the computed speed over time. The regression has been performed only in the range $20<Y<40$, which ensures a more stable tracking of the vehicles.

Table 1: Manual and Roboflow counting for light-duty vehicles (LDV), heavy-duty vehicles (HDV), and two-wheel motorbikes and scooters (TW).

	Manual Counting			Roboflow counting			Diff
Vid	LDV	HDV	TW	LDV	HDV	TW	
1	70	0	1	72	0	2	-3
2	*90*	*6*	*0*	*71*	*4*	*0*	*21*
3	71	2	0	73	2	0	-2
4	79	5	0	82	4	0	-2
5	79	2	1	79	3	1	-1
6	68	4	0	73	2	0	-3
7	*65*	*5*	*2*	*11*	*0*	*1*	*60*
8	70	7	0	74	6	0	-3
9	66	1	0	67	1	0	-1
10	64	5	1	67	3	1	-1
11	66	4	0	69	2	0	-1
12	80	2	1	82	1	1	-1
13	84	3	0	85	3	0	-1
14	89	1	0	90	0	0	0
15	62	2	0	63	1	0	0
16	86	1	0	88	0	0	-1
17	108	0	0	108	0	0	0
18	77	1	1	77	1	1	0
19	58	0	1	64	0	2	-7
20	83	1	1	85	0	0	0
21	70	1	0	72	0	0	-1
22	110	0	1	110	0	1	0
23	84	4	0	90	0	0	-2
24	78	1	1	79	0	0	1

The speed versus time obtained with the regression line is then considered the real speed of the vehicle and it is used to calculate the source power level as a function of time. This procedure assumes that speed is linear, i.e. acceleration can be constant (linearly increasing or decreasing speed) or zero (constant speed), neglecting the presence of vehicular jerk, i.e. the derivative of the acceleration. This approximation was supported also by a probe vehicle, equipped with an "On Board Diagnostic" (OBD) instrument measuring the real-time speed and GPS coordinates, which reported an almost constant speed over 14 transits performed in the test site, with very little variations [13].

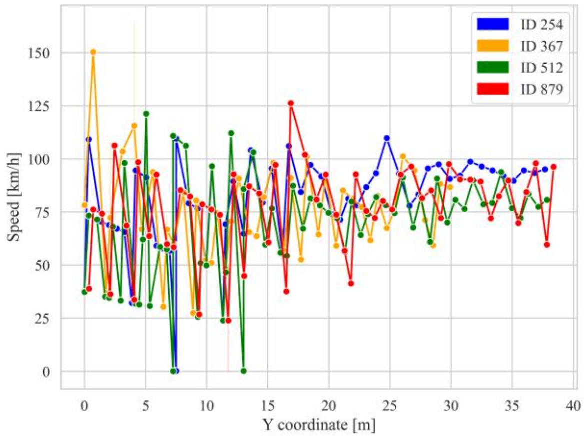

Figure 4: Speed versus Y coordinate for 4 sample vehicles.

3.3 Road traffic noise simulation results

Once the speed of each vehicle has been estimated as described above, the sound power level and the pressure level for each vehicle at the receiver have been estimated as a function of time according to equations 2. Then, the total pressure level at the receiver has been obtained summing the sound pressure levels – L_p produced by vehicles that occupy a position in the sensitive area at the same time. This procedure assumes that the noise produced by the vehicles is instantaneously propagated at the receiver, neglecting a possible small delay related to the speed of sound. Anyway, this assumption is reasonable, considering that the sensitive area is 40 m and sound takes approximately 0.1 s to cover this distance.

The simulated L_p has been compared to measured pressure levels in Figure 5 for videos 9, 10, and 11. The orange dotted line, that is the simulated L_p, follows quite well the blue line, i.e. the measurements. This means that the trend and part of the local oscillations of the measured level time history have been correctly retrieved. Missing points in the simulated series mean that no vehicles were tracked in those time periods, as confirmed by the low values of the measured level. However, local underestimations and overestimations are observed, as well as missing peaks related to failure in the detection of the vehicles, but, on average, the model produces a good approximation of ground truth values as a function of time.

Having the time histories of simulated L_p, it has been possible to estimate the continuous equivalent level on each 5-minute video and to compare with measured values. Results are reported in Table 2, excluding videos 2 and 7, that have been discarded in the recognition phase.

A general agreement between the simulations and measurements is shown. It can be noted that the model has some discrepancies, mainly because it tends to be quite stable over the videos. This is probably due to the cleaning and smoothing of data performed during the tuning of the process and needs to be further investigated.

The total measured L_{eq} over the two hours is 69.4 dBA, while the L_{eq} simulated with the EAgLE 3.0 model, is 68.5 dBA. When averaging on a longer time interval, discrepancies tend to reduce, giving hints about the potentialities of this approach for calculating common noise descriptors, such as L_{day}, L_{night}, L_{den}. In the authors' opinion, this result is promising, given the possible improvements that can be additionally achieved by further development of the modelling and finer tuning of the tracking algorithm.

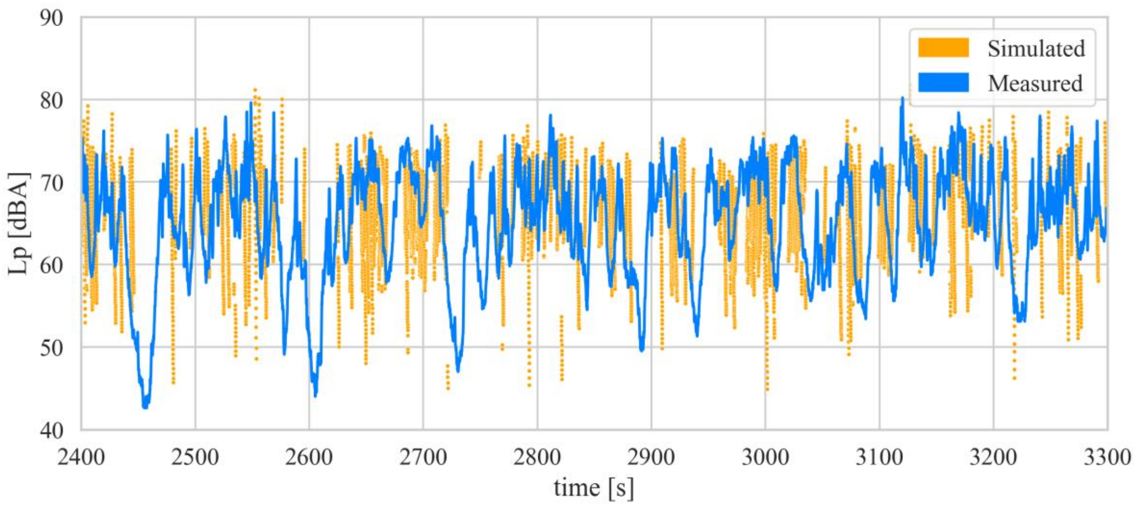

Figure 5: Time history of measured and simulated pressure levels as a function of time, in videos 9, 10, and 11.

Table 2: Comparison between measured and simulated equivalent level per each of the 22 videos, excluding videos 2 and 7, discarded in the tracking phase.

Video	Measured $L_{eq,5min}$ [dBA]	Simulated $L_{eq,5min}$ [dBA]
1	68.1	68.4
3	69.1	68.3
4	70.0	68.6
5	69.2	68.8
6	69.9	68.9
8	69.9	68.5
9	68.6	68.5
10	68.4	67.9
11	69.4	68.8
12	69.1	68.0
13	69.4	68.2
14	70.0	68.9
15	68.2	67.8
16	69.6	68.4
17	70.5	68.5
18	69.4	68.7
19	67.9	69.3
20	69.6	68.6
21	68.7	68.4
22	70.8	68.7
23	69.6	68.3
24	68.9	68.0

The error has been calculated as the difference between measured and simulated equivalent levels, over the 22 videos. The error is quite always positive, suggesting that the model has a systematic underestimation, probably due to missing vehicles recognition and the absence of ambient noise. The latter aspect will be investigated, for instance adding the measured L_{90} or L_{95} to the simulations, as an estimation of the background noise. The mean error is 0.81 dBA and the standard deviation among the 22 videos is 0.74 dBA.

The relationship between errors of the model and the total flow observed in each 5-minute video can be explored in Figure 6. The correlation between traffic flow and error is quite clear, since the trend and the peaks of flow and error curves mostly correspond. In several cases, the peaks in traffic flow coincide with peaks in the error (e.g., videos 17 and 22), suggesting that the model performances decrease with high-flow conditions.

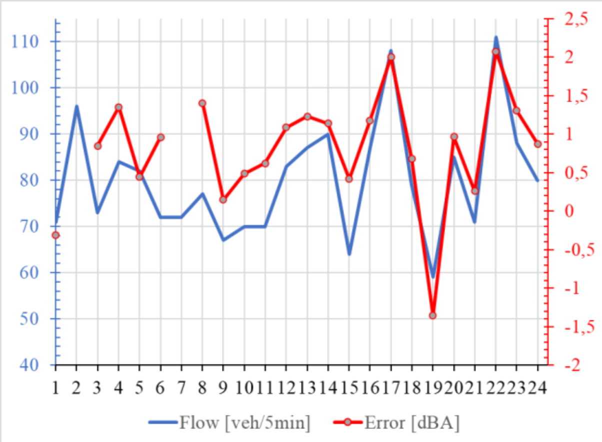

Figure 6: Overlap between total flow (blue line) and error of EAgLE 3.0 model (red line, secondary axis) in each 5-minute video, except videos 2 and 7, discarded in the tracking phase.

4. CONCLUSIONS

In this paper, a vehicle classification and tracking video analysis tool, based on AI image recognition, has been implemented and coupled with a microscopic and dynamic road traffic noise model. The AI algorithm was able to recognize most of the vehicles recorded with a camera in an extra-urban road case study. The classification highlighted some criticisms due to the presence of a safety grid in the image of the video, which affected also the tracking of the vehicles. Anyway, by removing videos with corrupted frames, performing data cleaning, and a smoothing of the calculated speed, it was possible to achieve a reasonable estimation of single-vehicle kinematics. Using the CNOSSOS-EU noise emission model and the free field propagation formula, the pressure level produced by each vehicle was calculated at the receiver and then aggregated to obtain the simulated time history and the continuous equivalent levels. The comparison between measurements and simulations provided good results in terms of mean and dispersion of the error, as well as in calculating the equivalent continuous level on the entire dataset.

The presented approach improves the results obtained in previous applications of the EAgLE model and, given the microscopic and dynamic nature of the model, it provides a solid basis for real-time noise assessment. After a proper tuning of the classification and tracking algorithm, as well as a period of validation with a sound level meter, the estimation of sound levels via the EAgLE approach is possible in any site in which a camera is installed. Anyway, limitations are still present, especially in the classification and tracking algorithm, which needs a large database of image frames, covering all the possible vehicles, and all the light and weather conditions. The possibility of other sources contributing to the overall equivalent level should also be faced, unless the target of the study is to assess only the road traffic noise. The future challenge will be testing the model in urban areas, in which speed varies much more than in highways and extra-urban roads.

5. ACKNOWLEDGMENTS

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' opinions, neither the European Union nor the European Commission can be considered responsible for them.
M. C. Coelho and A. Pascale acknowledge the support of the UID 00481 Centre for Mechanical Technology and Automation (TEMA).

6. REFERENCES

[1] European Commission. Directive 2002/49/EC Relating to the Assessment and Management of Environmental Noise, 2002.
[2] World Health Organization. Environmental noise guidelines for the European region. Regional Office for Europe: World Health Organization; 2018.
[3] European Environmental Bureau. Noise Pollution. Available online: https://eeb.org/work-areas/air-and-noise-pollution/ (accessed on 20 June 2024).
[4] P. Morano, F. Tajani, F. Di Liddo, and M. Darò: "Economic evaluation of the indoor environmental quality of buildings: The noise pollution effects on housing prices in the city of Bari (Italy)", *Buildings*, vol. 11. No 5, pp 213, 2021.
[5] A. Mehrotra, S.P. Shukla, A.K. Shukla, M.K. Manar, S.K. Singh, and M. Mehrotra: "A Comprehensive Review of Auditory and Non-Auditory Effects of Noise on Human Health", *Noise and Health*, vol 26. no. 121, pp. 59-69, 2024.
[6] D. Hugh, and I. Van Kamp: "Noise and cardiovascular disease: A review of the literature 2008-2011", *Noise and Health*, vol. 14, no. 61 pp. 287-291, 2012.
[7] M. Wrótny, and J. Bohatkiewicz: "Traffic noise and inhabitant health—A comparison of road and rail noise". *Sustainability*, vol. 13, no. 13, pp. 7340, 2021.
[8] S. Kephalopoulos, M. Paviotti, F. Anfosso-Lédée, "Common noise assessment methods in Europe (CNOSSOS-EU)", *Publications Office of the European Union*, 2012.
[9] S. Kephalopoulos, M. Paviotti, F. Anfosso-Lédée, D. Van Maercke, S. Shilton & N. Jones: "Advances in the development of common noise assessment methods in Europe: The CNOSSOS-EU framework for strategic environmental noise mapping", *Science of the Total Environment*, 482, 400-410, 2014.
[10] A. Kok, A. van Beek, "Amendments for CNOSSOS-EU: description of issues and proposed solutions", RIVM Lett Rep, 2019.
[11] C. Guarnaccia, A. Mascolo, P. Aumond, A. Can, and D. Rossi: "From early to recent models: A review of the evolution of road traffic and single vehicles noise emission modelling", *Current Pollution Reports*, vol. 10, no. 4, pp. 662-683, 2024.
[12] C. Guarnaccia: "EAgLE: Equivalent Acoustic Level Estimator Proposal", *Sensors*, vol. 20, no. 701, 2020.
[13] A. Pascale, E. Macedo, C. Guarnaccia, and M.C. Coelho: "Smart mobility procedure for road traffic noise dynamic estimation by video analysis", *Applied Acoustics*, vol. 208, pp. 109381, 2023.
[14] C. Guarnaccia, U. Catherin, A. Mascolo, and D. Rossi: "Improvement and Validation of a Smart Road Traffic Noise Model Based on Vehicles Tracking Using Image Recognition: EAgLE 3.0". *Sensors*, vol. 25, no. 6, pp. 1750, 2025
[15] Available online: https://docs.roboflow.com/api-reference/introduction (accessed on 15 March 2025).
[16] Available online: https://blog.roboflow.com/estimate-speed-computer-vision/ (accessed on 18 March 2025).
[17] Y. H. Lee, and H.J. Kim: "Comparative Analysis of YOLO Series (from V1 to V11) and Their Application in Computer Vision". *Journal of the Semiconductor & Display Technology*, vol. 23, no. 4, pp. 190-198, 2024.
[18] A. Agarwal, C.V. Jawahar, and P.J. Narayanan: "A Survey of Planar Homography Estimation Techniques", Tech. Rep. IIIT/TR/2005/12; Centre for Visual Information Technology: Hyderabad, India, 2005.