www.nature.com/scientificreports

scientific reports

Check for updates

OPEN

An innovative virtual sensing system for the vehicle-centric evaluation of emissions in the sustainable mobility transition

Silvia Carla Strada✉, Antonio Pagliaroli & Sergio Matteo Savaresi

The objective of the research is to develop a new methodology capable of quantifying the emissions of internal combustion engine cars based not only on their technology but also on the dynamic behavior of the individual vehicle. No complicated exhaust measurement apparatus is needed. Employing just a telematic device equipped with GNSS and an inertial measurement unit, which can be easily attached to the vehicle, we demonstrate how individualised, accurate climate-altering emissions result by using figures extracted from European test databases, appropriately combined with real travelled distances, speeds, and driving style of each individual car. Pollutant emissions are, in this paper, also assessed though in a more general manner, and this represents a starting point for a future complete vehicle-centric emissions evaluation. This research demonstrates that circulating car emissions have to be assessed considering not only the Euro class and type of fuel, but also how the vehicle is driven. In this work a new, implementation oriented methodology is proposed in order to properly merge average European standard emission data with car-based telematic measures to highlight a new paradigm for regulating in a more conscious way the transition of the internal combustion engine car park towards a large scale, sustainable mobility ecosystem. This approach would place greater responsibility on the driver, without leaving anyone behind and without compelling the car scrapping solely based on the Euro class of its engine. The key contribution is ultimately to establish an easy to handle methodology enabling targeted and behavior-based traffic management policies. In fact, through a massive telematics recent dataset, the proposed virtual emission sensing system reveals that an extensive and inefficient use of eco-friendly vehicles produces much more greenhouse gas emissions than older vehicles driven in an environmentally responsible way.

In the light of the increasingly radical effects of climate change, the urgency of environmental sustainability has become paramount in shaping economic and social development. The mobility sector is poised for transformative changes[1–3] and the global community has recognized the imperative of mitigating climate change and reducing traffic emissions. Within this context, the key objectives of enhancing energy efficiency in transportation and transitioning towards zero-emission mobility have arisen as critical endeavors in aligning with the Paris Agreement targets of limiting temperature increase to well below $2°C$[4]. Specifically, the EU has taken on a global leadership role in mitigating climate change and developing low-carbon energy technologies in the transportation sector[5]. Currently, transportation accounts for nearly a quarter of Europe's greenhouse gas emissions and it is the leading cause of air pollution[6]. Among various transportation modes, passenger vehicles (cars and buses) contribute for a 45.1%, followed by trucks (29.4%), aviation (11.6%), shipping (10.6%), pipelines (2.2%), and rails (1%)[7]. Furthermore, road transport specifically contributes to more than a third of NO_x emissions in Europe and serves as the dominant source of polluting gases in urban and heavy traffic areas[6]. To address these challenges and go towards zero-emission mobility, the European Commission has formulated the Strategy for Sustainable and Smart Mobility, which aims to guide European transport towards a sustainable and smart future, with the goal of achieving a reduction of transport emissions by 90% by 2050, as stated in the European Green Deal[2,8].

To expedite the monumental transition towards sustainable mobility, policymakers and researchers must adopt innovative methodologies that intricately address the dynamic interplay of technology, behavioral shifts, and policy frameworks. A body of studies underscores the pivotal role of transitioning from internal combustion

Department of Electronics, Information and Bioengineering, Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milano, Italy. ✉email: silvia.strada@polimi.it

engines (ICEs) to electric vehicles (EVs), considering also the role of hybrid vehicles[9] and a novel shared mobility paradigm in substantially influencing greenhouse gas mitigation[10,11]. Indeed, hybrid and electric vehicles also produce both greenhouse gas emissions and pollutants, taking into account the so-called "well-to-wheel" emissions, which are estimated across the entire process of energy flow, from raw material extraction to battery/vehicle production to ready-to-use car (the global electric vehicle fleet emits half the amount of carbon dioxide that would have been emitted from an equivalent fleet of internal combustion engine vehicles[12]). Nonetheless, these emissions are not easily evaluable yet, in particular those related to battery assembly from raw materials. Similarly, standard emissions for hybrid vehicles are still in need of verification. Furthermore, motorized transport, in particular road private cars, will remain dependent on internal combustion engines for a quite long time. Considering the functional limitations of electric vehicles in terms of range autonomy, and on the other hand, the need to cover high mileage to breakeven the high purchase cost within a manageable timeframe, the transition to electric cars is slower than expected. Moreover, the complexities introduced by inflation, consumer preferences, the financial burdens associated with developing a more environmentally friendly electricity grid, and disruptions in the supply chain necessitate a judicious equilibrium between advancements in technology and the continued presence of internal combustion engine vehicles[13–16]. In some way, the complete, inclusive transition to a fully electric vehicle fleet will only occur concurrently with their shared use, likely enabled by driverless technology. Until then, many privately owned cars with internal combustion engines will still be in circulation, and their sustainable management, with innovative regulatory tools, will be of fundamental importance. Previous studies, as comprehensively reviewed by Singh and Kathuria[17], have extensively explored driver behavior, encompassing characteristics such as aggressiveness, smoothness, and more. Employing various data-gathering methods, studies have offered diverse approaches to evaluate driver behavior, with many relying on smartphone sensors[18,19] or retrieving information from the On-Board Diagnostic interface alongside additional sensors like GNSS from mobile phones[20]. However, it is crucial to recognize that these smartphone-based methods, while suitable for everyday tasks, lack the precision required for quality road positioning[18,19], thus limiting their ability to detect and quantify the true impact of driver behavior on vehicle emissions and fuel consumption in real-world scenarios.

To address this critical knowledge gap, we established a groundbreaking partnership with UnipolTech Spa, the telematics branch of Unipol Group Spa, a prominent insurance company based in Bologna, Italy, to embark on a pioneering analysis. Leveraging an extensive and unparalleled dataset collected from black box devices permanently installed in 8,457 privately owned vehicles across Italy, our study encompasses over 11 million trips, spanning approximately 190 million kilometers. These black box devices, equipped with Inertial Measurement Units (IMU) and GNSS receivers, provide continuous spatio-temporal measurements of each vehicle's behavior, enabling transcend the confines of conventional emission evaluations that predominantly rely on limited criteria such as standards and registration year. By leveraging this granular and real-world dataset, the aim is to unlock deeper insights into the complex interplay between driving behavior and its direct impact on fuel consumption and emissions.

In this article, we present the pioneering data-driven methodology of fine virtual sensing of emissions calibrated for each individual vehicle, combining its technology with its actual behavior. By extracting, from raw black box tracking data, features such as distances traveled, average speeds on different road types, hard accelerations and braking, we develop an innovative method to combine them with the experimental HBEFA database, to assess real-world fuel consumption, CO_2 emissions, and NO_x emissions in a vehicle-centric way. Our study offers a paradigm shift in quantifying the true environmental impact of individual vehicles, overcoming the perspective of average emission classes. By shedding light on the influence of driving patterns on emissions and consumption, we empower policymakers with a novel emission virtual sensing tool to craft evidence-based transportation policies that regulate limited urban access zones, incentive sustainable driving habits, and steer the renewal of vehicle fleets. Ultimately, our work seeks to transform the discourse on sustainable mobility by proposing a simple yet effective and deep clustering methodology. This innovative approach classifies vehicles into distinct groups based on their environmental impact, enabling policymakers to identify the best and worst performers. By harnessing the power of existing black box data, our findings pave the way for a more sustainable future, wherein transportation systems are optimized for environmental stewardship while meeting the mobility needs of society, by redefining what are now called black boxes as "green" boxes.

In the following sections, we present our data-driven virtual sensing methodology, discuss the implications of our analysis, and outline potential policy interventions that can sustain a more concrete and precise assessment of private car emissions, beyond widespread common places.

Results

In this section, we present the system for one's car individual evaluation of emissions, starting from the existing vehicle information and Euro class rating (from Euro 0 to Euro 6, the most recently defined standard). All vehicles on European roads, since 1992, must have an Euro class rating which represents the results of tests carried out by vehicle makers to simulate the levels of harmful emissions produced in certain driving conditions. Leveraging the extensive dataset, collected from vehicles equipped with black box devices between January and September 2022 and immatriculated all over Italy, we develop a methodology for virtually measuring environmental impact and consumption for the individual vehicle, no longer for a whole class of vehicles. Needless to say, the issue is quite significant because it is impossible to physically measure a car's emissions with standard onboard instruments. To quantitatively assess the environmental implications of driving behavior, we define and develop three adimensional key performance indicators (KPIs):

- P_{cons} (fuel consumption): This KPI encapsulates a range of influential factors, including average speed, road type, kilometers traveled, vehicle type, and driving style.

- P_{CO_2} (CO_2 emissions): A comprehensive KPI that considers various contributors to carbon emissions, such as those associated with average speed, kilometers traveled, road type, fuel type, and vehicle characteristics.
- P_{NO_x} (NO_x emissions): This KPI reflects the nitrogen oxide emissions, taking into account factors like vehicle type, Euro Class, and kilometers traveled.The KPIs are scaled from 0 to 1, with 0 signifying optimal environmental performance (low consumption and emissions) and 1 representing the worst case (high consumption and emissions). In providing a holistic perspective, these KPIs unveil the intricate interplay between driving behavior and its direct effects on fuel consumption and emissions. Our new methodology shows how to use known and well established models and combine them with the driving data and vehicles characteristics to generate a customized measure of environmental impact of any car. The innovation in the method we will illustrate lies in the proposition of an off-the-shelf algorithmic and hardware tool that can accurately monitor the emissions of any car in real-time.

Fuel Consumption Analysis

Assessing the real fuel consumption of vehicles is crucial for gaining an accurate understanding of their environmental impact and for designing effective sustainability measures. In this respect, we introduce the first KPI, P_{cons}, which represents the output of the proposed model of virtual sensing fuel consumption, properly merging together various additional and non standard influential factors, such as average speed, road type, kilometers traveled, driving style, and vehicle characteristics.

Based on the experimental relationship between average speed and fuel consumption for road vehicles[21], reported graphically in Fig. 1a, it is made clear, and is something not quite intuitive at first glance, that internal combustion engine cars consume fuel at the lowest rate when driven at speeds of 50-75 km/h, namely in the "Green Speed" range. At lower speeds, fuel consumption rapidly worsens, mainly due to the increasing rolling resistance force. At higher speeds, more intuitively, more fuel is consumed because of the need to overcome aerodynamic drag. To deepen our analysis, we navigate the complexity of evaluating fuel consumption across three distinct contexts: urban roads, highways, and mixed roads (encompassing diverse road types within the same journey, including extra-urban roads). Utilizing publicly available information, we first refer to manufacturers' declared fuel consumption data for each road typology, as determined through the European mandated Worldwide Harmonised Light Vehicles Test Procedure (WLTP)[22]. The WLTP is divided into four phases that correspond to different traffic conditions with varying average speeds: low, medium, high, and extra high. This cycle draws

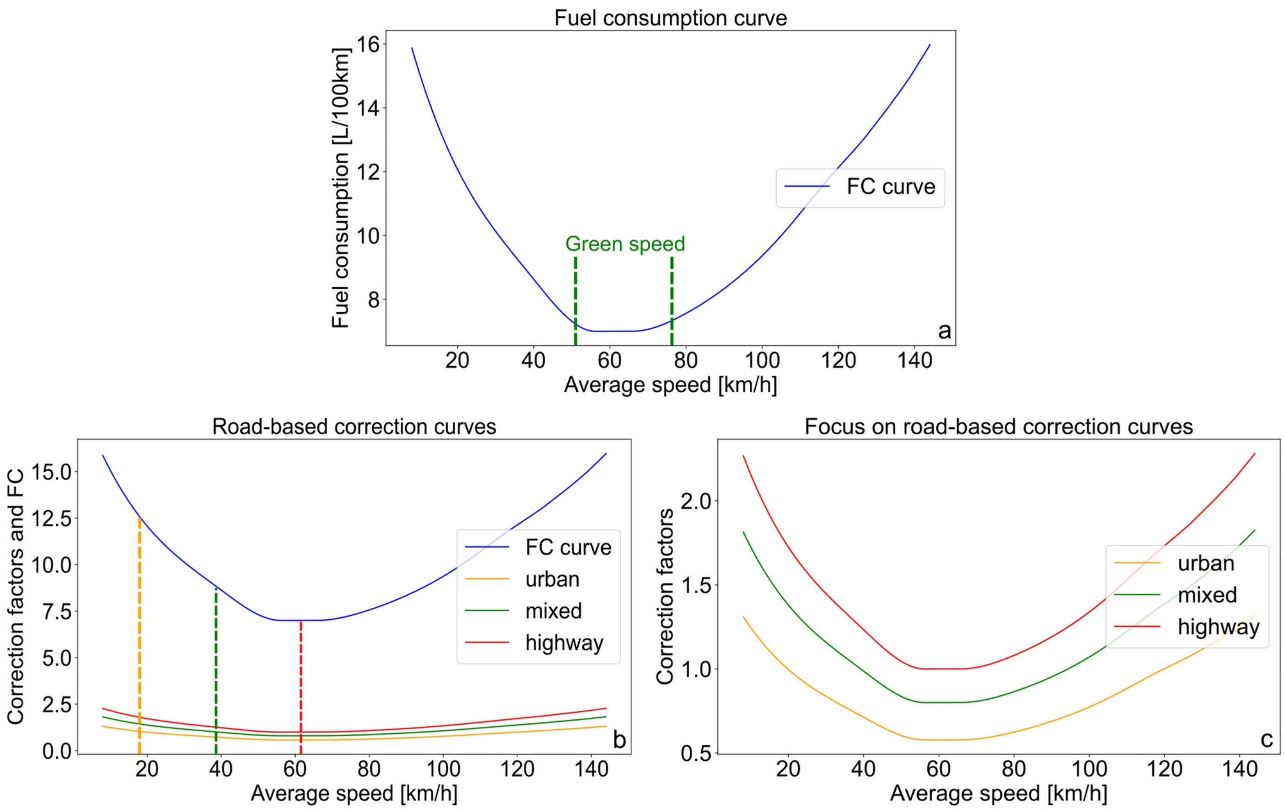

Fig. 1. **a** Mapping Average Speed vs. Fuel Consumption(FC), pinpointing the optimal "Green Speed" range (50-75 km/h). **b** Normalized Consumption Curves are shown for urban, mixed, and highway contexts by normalizing the original FC curve with respect to the Fuel Consumption Values at 18.9 km/h, 39.2 km/h, and 60.8 km/h, which correspond to the speeds defined by the Worldwide Harmonized Light Vehicles Test Procedure (WLTP) cycle for urban roads, mixed roads, and highways, respectively. **c** A Close-up presentation of the three curves.

from real-world driving data, representing approximately one million kilometers of driving[22]. It also considers the Power-to-Mass Ratio (PMR) and classifies vehicles into three categories. In our study, we concentrate on vehicles of class 3 (PMR 34KW/ton), which aligns with the majority of private cars in Europe.

Given that our dataset provides detailed information about the type of roads traveled (urban, mixed, or highway), we determine the average speeds of all trips on these road types to be 21.2 km/h, 41.96 km/h, and 71.48 km/h, respectively. These values led us to reasonably assume that manufacturers' WLTP declared fuel consumption values are designed around the average speed of the WLTP Class 3 cycle. In fact, for urban and mixed roads, the average speeds derived from the low and medium cycles with stops of the WLTP test are 18.9 km/h and 39.2 km/h, respectively (very similar to the above data driven values of 21.2 km/h and 41.96 km/h). These speeds choice aligns with the real-world conditions where drivers often encounter frequent stops and starts in urban areas due to traffic lights, stop signs, and so on. In contrast, when assessing highway road fuel consumption, we employ the average speed corresponding to the high-speed cycle without stops of the WLTP test, which stands at 60.8 km/h (similar to the above data driven value of 71.38 km/h).

Based on the above assumptions on the average velocities employed by the car manufacturers, three adimensional correction curves are constructed for urban, mixed, and highway roads normalizing the general function, shown in Fig. 1a. For the urban context, a curve modeling the relationship between average speed and a new fuel consumption correction factor is derived by normalizing the general function with respect to the consumption value (12.15 L/100km) corresponding to the speed of 18.9 km/h. Methodologically this means generating a new function, shown in yellow in Fig. 1b and magnified in Fig. 1c, dividing all the y-values in Fig. 1a by 12.15 L/100km. Thus, when a vehicle maintains an average speed of 18.9 km/h on urban roads, the car's consumption matches the manufacturer's declared urban value; deviations from this speed, above or below, determine fuel consumption variations according to the correction factor curve shown in Fig. 1b. Similar procedures are applied to derive the curves for mixed roads and highways, with normalization based on consumption values of 8.75 L/100km for mixed roads and 7 L/100km for highways, corresponding to average speeds of 39.2 km/h and 60.8 km/h, respectively. This simple but key correction procedure ensures that the derived curves of Fig. 1b and Fig. 1c accurately reflect the declared consumption values by manufacturers for trips conducted at the average speeds defined by the WLTP cycle.

This process yields three unique, adimensional speed-based correction functions that establish the relationship between the average speed for each road type and the fuel consumption correction, in a way that allows us to refine the fuel consumption calculation, commonly based on the distance traveled solely, according to the actual average driving speed for each road type and whether it is more or less close to the "Green Speed" range.

Using the normalized consumption curves, shown enlarged in Fig. 1c, as scaling factors, P_{cons} is formulated as follows:

$$P_{cons} = D_U \cdot F_U \cdot C_U + D_M \cdot F_M \cdot C_M + D_H \cdot F_H \cdot C_H \tag{1}$$

where D_U, D_M, and D_H [km] represent the distances traveled by the car on urban, mixed, and highway roads, respectively, F_U, F_M, and F_H [dimensionless] are the scaling factors corresponding to the vehicle's average speed on these road types and C_U, C_M, and C_H [L/km] denote the car manufacturer's declared fuel consumption on the same urban, mixed, and highway ride's segments, respectively.

Since black boxes allow to take into account aggressive or calm driving behavior, we incorporate this into the P_{cons} indicator too, in order to fine-tune even better the consumption indicator. The data provided on harsh events, recorded for each vehicle, include information about accelerations and decelerations higher than |0.3|g (g-force), in the range from -2 g to 2 g. We then categorize braking and acceleration harsh events based on their intensity, Average [0.3-0.5]g, Heavy [0.5-0.8]g, Very Heavy [0.8-1]g, and Extreme [1-2]g. Each category is assigned a weight that holds higher with higher g-values, to penalize more the more aggressive behavior. We then calculate a weighted sum for each car by multiplying the number of harsh events in each of the four categories by the corresponding weight. This sum is then divided by the total distance traveled, yielding the harsh events against distance traveled ratio (*HR*), expressed as:

$$HR = \frac{W_A \cdot N_A + W_H \cdot N_H + W_{VH} \cdot N_{VH} + W_E \cdot N_E}{DT} \tag{2}$$

In our model the weights W_A, W_H, W_{VH}, and W_E have values 0.1, 0.3, 0.6, 1, and are coupled to the average, heavy, very heavy, and extreme category, respectively. Similarly, N_A, N_H, N_{VH}, and N_E denote the number of registered harsh events for each vehicle and each category, and *DT* [km] is the total distance traveled by each car.

The resulting *HR* values are normalized against the maximum ratio, representing the most aggressive driver. The normalized $\overline{HR}$ falls in the range [0,1] and we define an additional adimensional factor, *DS* (Driving Style), to penalize aggressive driving in P_{cons}, following other studies[23,24] that indicate a worsening of the consumption by a percentage from 20% to 40 % with aggressive driving. In our case, we decide conservatively, to penalize the most aggressive driver by 30 %: $\overline{HR}$=1 implies *DS*=1.3 (30% higher fuel consumption due to driving style), while a non aggressive driver (no harsh events) has a $\overline{HR}$=0 associated to *DS*=1 (no modification of fuel consumption).

This methodology brings to the innovative virtual measure of the final fuel consumption KPI, the latter being generated for each individual car:

$$P_{cons} = (D_U \cdot F_U \cdot C_U + D_M \cdot F_M \cdot C_M + D_H \cdot F_H \cdot C_H) \cdot DS \tag{3}$$

Ultimately P_{cons} for each vehicle is normalized against the maximum value of P_{cons} that corresponds to the vehicle that has the higher consumption. This normalization ensures that the KPI for each vehicle ranges between 0 (vehicle not traveling and not consuming fuel) and 1 (indicating high consumption).

Carbon dioxide emissions analysis

After an extensive literature review on average speed vs. CO_2 emission [g/km] relationship, it is confirmed that they are linked by a convex polynomial relationship, similar but different in form for petrol and diesel, due to their distinct combustion processes[23,25–27]. This functional relationship clearly indicates how there is an intermediate speed range in which emissions reach a minimum level. In this work, we refer to the convex polynomial curves (for diesel and petrol vehicles) presented in[23] and derived from the Handbook Emission Factors for Road Transport (HBEFA), a reliable and well known source for emissions average values calculated on real fleets.

We firstly employ the HBEFA database to verify if greenhouse gas emissions are dependent on the Euro class of the vehicle, in addition to the type of fuel. The HBEFA-based experimental convex polynomial relationship between average speed and CO_2 emissions turns out to be in practice invariant with respect to the Euro class of the car[28]. For the purposes of the method, which aims to be a practical and simple tool for regulating traffic also based on the driving style of the car, this is an important simplifying factor. Differently from fuel consumption, the car makers do not distinguish CO_2 emissions on different type of roads. Therefore, we have only two normalized CO_2 emissions functions, one for diesel and one for petrol. To obtain the dimensionless correction factor function, we identify the speed of approximately 70 km/h corresponding to the minimum consumption (both for diesel and for petrol) and normalize the functions with respect to their corresponding y-value. This procedure is carried over both for the diesel and for petrol curves. In Fig. 2, as an example for the diesel case, we show the procedure and the obtained normalized curve (in yellow in Fig. 2b and Fig. 2c), which shows that, as it should be, a unity dimensionless correction factor corresponds to the 70 km/h speed.

This speed-based normalization procedure leads to the definition of the type of fuel based CO_2 emissions KPI:

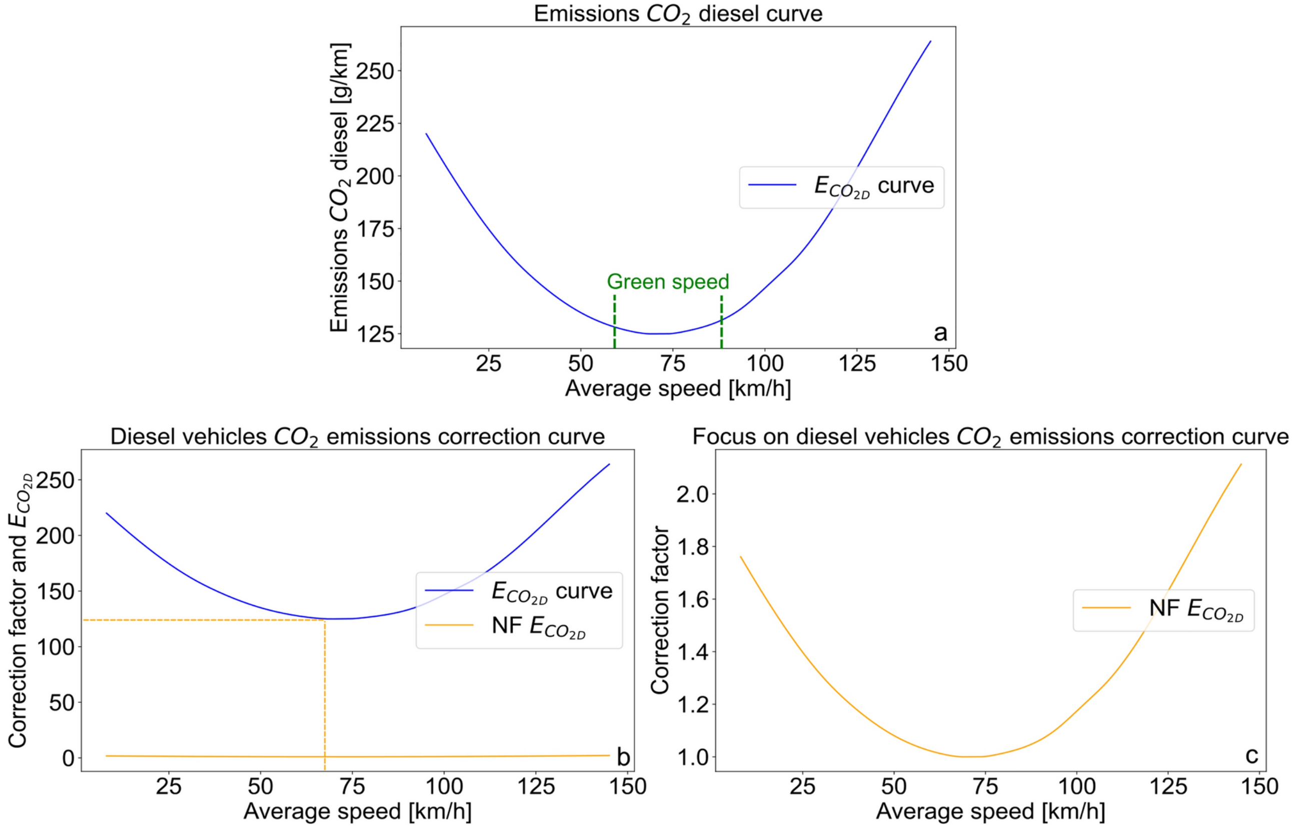

Fig. 2. **a** Mapping Average Speed vs. CO_2 emissions of Diesel Vehicles ($E_{CO_{2D}}$). **b** Normalized Emissions Curve (NF $E_{CO_{2D}}$) is obtained by normalizing the $E_{CO_{2D}}$ curve against the minimum emissions value associated with speeds falling within the "Green Speed" range. This normalization is conducted under the assumption that the declared CO_2 emissions value provided by car manufacturers corresponds to this specific speed (69.28 km/h), with a reference value of 124.94 g/km. **c** A Close-up examination of the normalized emission curve for diesel vehicles.

- P_{CO_2} for diesel vehicles (D):

$$P_{CO_{2D}} = \left(E_{CO_{2D}} \cdot F_{U_D} \cdot D_U + E_{CO_{2D}} \cdot F_{M_D} \cdot D_M + E_{CO_{2D}} \cdot F_{H_D} \cdot D_H\right) \cdot DS \tag{4}$$

where F_{U_D}, F_{M_D}, F_{H_D} are the adimensional corrective factors based on the average speed on urban, mixed, and highway roads, respectively (the function is one, and we consider the different average speeds on each road type). D_U, D_M, D_H [km] are the distances traveled on urban, mixed, and highway roads, and $E_{CO_{2D}}$ [g/km] is the CO_2 declared emissions by car manufacturers of a diesel vehicle. *DS* is defined in the same manner as for P_{cons}, based on studies that reveal a potential increase of around 30% in CO_2 emissions with an aggressive driving style[29,30].

- P_{CO_2} for petrol vehicles (P):

$$P_{CO_{2P}} = \left(E_{CO_{2P}} \cdot F_{U_P} \cdot D_U + E_{CO_{2P}} \cdot F_{M_P} \cdot D_M + E_{CO_{2P}} \cdot F_{H_P} \cdot D_H\right) \cdot DS \tag{5}$$

where factors and variables are the same as above, with subscript P that indicates petrol vehicles.

In conclusion, the normalization of P_{CO_2} for each vehicle is carried out relative to the maximum P_{CO_2} value observed among all vehicles. This normalization guarantees that the resulting KPI for each vehicle falls within the range of 0 (vehicle not traveling and not emitting CO_2) to 1 (indicating high CO_2 emissions).
It is important to note, even though we use HBEFA, that an analysis based on COPERT, another EU standard vehicle emissions calculator, would have provided very similar results, as it makes use of HBEFA data. It is important to emphasize that these emission calculation tools are related to different fleets and various scenarios, and they cannot directly model the behavior of each individual vehicle. This study demonstrates the potential of these tools in building up the behavior of each individual vehicle, paving the way for more specific and personalized emission monitoring systems.

Nitrogen oxides emissions analysis
During vehicle engine operation, combustion generates heat that catalyzes the binding of nitrogen (*N*) and oxygen (O_2) to form nitric oxide (*NO*) or nitrogen dioxide (NO_2), both of which are encompassed under the term nitrogen oxides (NO_x)[31]. The available information concerning NO_x emissions has been affected by biases in car manufacturers' declarations since 2017 (dieselgate) and for a long time diesel polluting real emissions where far above the limits,[31,32]. Anyway, the impact of car technology and Euro class is the most significant factor affecting nitrogen oxides emissions and particulate matter. For the aforementioned reasons, we build here a simple model that relates the NO_x emissions to the distance traveled, referring the interested reader to our forthcoming studies on the dependence on speed and driving style. Creating a *KPI* that considers NO_x emissions presents an additional challenge as the manufacturer is not required to indicate the values of individual vehicle pollutant emissions. Therefore, our methodology assigns to each car the NO_x emissions highest value corresponding to its Euro class[34], for which we possess information derived from the vehicle's model and year of registration. The calculation of P_{NO_x} is as follows:

- P_{NO_x} for diesel vehicles (D):

$$P_{NO_{xD}} = E_{NO_{xD}} \cdot D_T \tag{6}$$

where D_T is the total distance traveled, $E_{NO_{xD}}$ corresponds to the NO_x worst emissions of a diesel vehicle (based on Euro class).

- P_{NO_x} for petrol vehicles (P):

$$P_{NO_{xP}} = E_{NO_{xP}} \cdot D_T \tag{7}$$

where D_T is the total distance traveled, $E_{NO_{xP}}$ corresponds to the NO_x worst emissions of a petrol vehicle and is based on the Euro class.

Concluding the analysis, the normalization of P_{NO_x} for individual vehicles is performed against the highest observed P_{NO_x} value among all vehicles. This normalization guarantees that the resulting KPI for each vehicle lies within the range of 0 (vehicle not traveling and not emitting NO_x) to 1 (indicating high NO_x emissions).
It is worth mentioning here that we are developing, based on HBEFA and validated by experiments from a partner certification company, an innovative vehicle-centric method for pollutant emissions measurement considering driving behavior as well, which we will publish in the near future. We are sure that this study will be able to deepen the analysis on NO_x emissions and their dependencies from driving style.

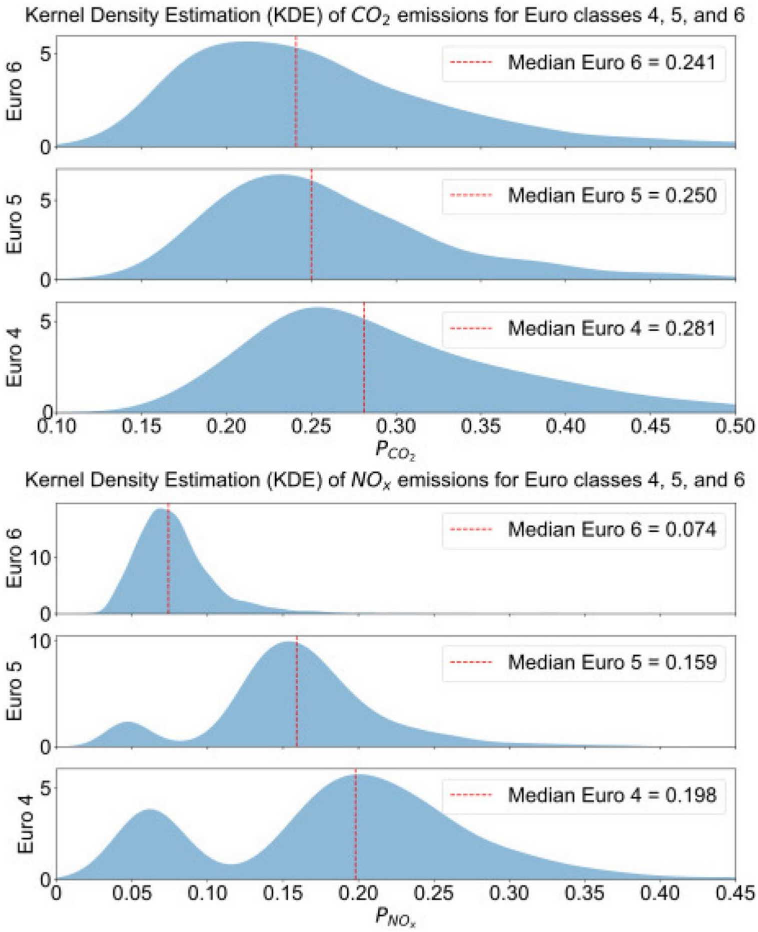

Fig. 3. Kernel Density Estimations of P_{CO_2} and P_{NO_x} for Euro 4, Euro 5, and Euro 6 vehicles. Each sub-image corresponds to a specific Euro class, illustrating emissions distribution normalized to a scale of 0 to 1. Median values are indicated for each distribution, offering insights into the environmental performance of vehicles within each Euro class. For detailed emission values, please refer to the Table 1. Notably, in the P_{NO_x} distribution, there are two distinct peaks in Euro 4 and Euro 5, due to the limits on NO_x emissions that are greater than 3 times for Diesel Vehicles (0.08 g/km, 0.06 g/km, 0.06 g/km for Petrol Euro 4, 5, 6, respectively, while 0.25 g/km, 0.18 g/km, 0.08 g/km for Diesel Euro 4, 5, 6)[34].

Euro	Median CO_2	Emissions CO_2[Kg]	Median NO_x	Emissions NO_x[g]
Euro 4	28.10	3255.936	19.80	4580.684
Euro 5	25.00	2892.697	15.90	3681.985
Euro 6	24.10	2800.602	7.40	1700.362

Table 1. Emission values relative to median KPIs.

Determining the actual environmental impact of vehicles

Our study introduces a novel set of KPIs to provide a robust and finer foundation for assessing the true environmental impact of vehicles, without the need of bulky and expensive equipment. Before delving into the data-driven results comprehensive analysis, it's essential to emphasize the innovative nature of the presented KPIs and their interconnectedness.

Remarkably, a substantial correlation (with a coefficient of $R > 0.95$) emerges between the fuel consumption KPI and the CO_2 emissions one. This correlation aligns with the conclusions of prior scientific investigations[23,35], reinforcing the consistent existence of a linear relationship between fuel consumption and climate altering emissions. In light of this correlation, we purposefully focus the following discussion on P_{CO_2} and P_{NO_x}.

Taking these KPIs into consideration, our analysis unveils intriguing insights. We find that among the vehicles in our dataset, there are 59 vehicles with Euro 3, 1283 with Euro 4, 1563 with Euro 5, and 5552 with Euro 6 standard. The Kernel Density Estimation (KDE) of cars depending on P_{CO_2} and P_{NO_x} for Euro 4, 5, and 6 vehicles is depicted in Fig. 3, omitting Euro 3 class due to its limited representation in our sample. Notably, Euro class alone does not provide a complete picture of a vehicle's environmental impact: while it's commonly assumed that higher Euro standards correlate with lower emissions, our analysis uncovers a more complex reality.

Analyzing the median values, Table 1, of Euro 4, 5, and 6 estimates, we observe a common expectation that a higher Euro standard is generally associated with lower emissions. However, the analysis uncovers a significant nuance that challenges this conventional wisdom. It is evident that not all vehicles conform to the straightforward notion that a lower Euro class implies lower emissions. Some Euro 6 vehicles, indeed, exhibit significantly higher emissions compared to Euro 4 ones. Correspondingly, several Euro 4 vehicles demonstrate lower emissions compared to their Euro 6 counterparts. These discrepancies are attributable to factors beyond the Euro class, such as distance traveled, road types, average speeds, driving style, and vehicle characteristics, as encapsulated in the presented KPIs.

In light of these results, policies advocating the renewal of vehicle fleets by scrapping those with lower Euro classes, aimed for example at restricting access to urban areas, may be overly simplistic and fail to effectively address the genuine environmental impact of vehicles.

Our holistic approach, as demonstrated in this paper, enables in-depth analyses of the emissions of each individual vehicle by evaluating all available information through black boxes. Exploiting the full potential of our model and the developed KPIs, we considered the four Euro classes present in our dataset: Euro 3, 4, 5, and 6.

To illustrate the insufficiency of Euro classes alone in determining a vehicle's pollution level, we adopted a clustering approach using P_{CO_2} and P_{NO_x} as features. Employing the K-Means clustering method with 4 clusters, mirroring the number of Euro classes in the considered dataset, we classified vehicles into four environmental impact groups. The resulting classification, shown in Fig. 4, significantly differes from the distribution of Euro classes in the original dataset (59 Euro 3, 1283 Euro 4, 1563 Euro 5, 5552 Euro 6).

Specifically, the worst environmental performers group contains 636 vehicles, the next cluster comprises 1688 vehicles, the following includes 2001 vehicles, and the last group, considered the least environmentally impactful, includes 4132 vehicles. Notably, this classification allows for a more nuanced impact assessment than Euro classes alone, providing a concrete understanding of individual behavior.

Moreover, our model transcends the creation of generic groups and facilitates assessments at different granularities of vehicle environmental impact. It enables individualized environmental evaluations, empowering policymakers to set emission limits for different pollutants in each vehicle. This approach challenges conventional restrictions based solely on Euro classes. In instances where a vehicle might be deemed too polluting under existing policies due to its Euro class, our study's in-depth analysis allows for a re-evaluation, potentially allowing it to continue circulating based on its individual environmental performance.

Discussion

Our aim is to present a methodology, lightweight and easily implementable, to virtually sense with precision climate altering and polluting emissions on an individual vehicle basis. The method challenges the conventional approach to assessing the environmental impact of vehicles based on the widely used Euro class, which proves insufficient as the sole determinant of a vehicle's environmental impact. Our new algorithm KPIs provide a comprehensive framework that enables traffic control policies, which currently mostly rely on the Euro class solely, to consider also virtual measurements (without expensive sensors) of emissions based on the actual behavior of each individual vehicle in terms of distance, speed, and driving style. The current paradigm of assessing cars environmental impact fails to acknowledge that an older vehicle could meet environmental limitations if used judiciously, with limited kilometers, at eco-friendly speeds, and employing an optimal driving style. Hence, since the affordability of a new full electric car might still be a barrier for many, the new proposed way of considering emissions allows for the inclusion of social and economic aspects, which will be important to evaluate during this transition in mobility, so that it is not destructive and overly economically impactful on individuals.

Our proposed methodology embraces a "vehicle-centric" emission virtual measurement process, allowing each vehicle to be classified combining average values from trustworthy road transport emission estimation

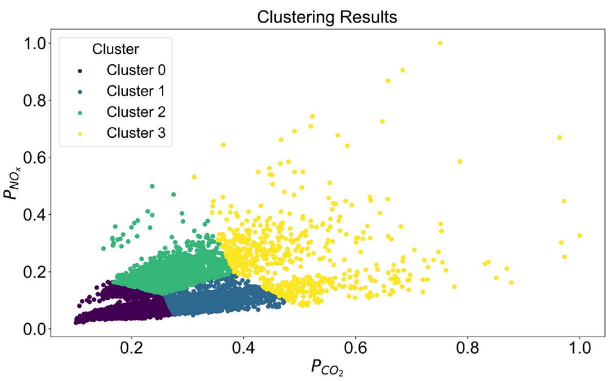

Fig. 4. Cluster Analysis of Vehicle Emissions Based on P_{CO_2} and P_{NO_x} KPIs: assessing 8457 vehicles across four distinct clusters reveals varying environmental performance. Cluster sizes: 4132 (Cluster 0: environmental leaders), 2001 (Cluster 1: low impact), 1688 (Cluster 2: moderate impact), and 636 (high impact).

models with specific vehicle behavior information, requiring only the use of a simple telematics black box equipped with GNSS and IMU. Thereby, cars' greenhouse gas and pollutant emissions can be individually compared to standards set by policymakers seeking to devise effective strategies for transportation rulemaking during this transition period in road transportation. The European Commission, in fact, aims to achieve a remarkable 90% reduction in transport emissions by 2050. This proposed pioneering approach is crucial for ensuring an accurate and as fair as possible transition phase, placing a decisive new responsibility on the driver, since he or she can directly "measure" his or her driving manner effects on emissions, so allowing that, if his or her driving is virtuous, even the old car will not necessarily have to be right away decommissioned.

Methods

Characteristics of GNSS receivers used for data collection

The GNSS receivers employed in this study are embedded within a specialized small device, referred to as telematic black box, featuring several key attributes crucial for data collection and analysis. The about 5 Mio boxes circulating in Italy are equipped also with an Inertial Measurement Unit (IMU), enhancing the precision and reliability of gathered data. The IMU, working at a max frequency of 200Hz, supplements GNSS-derived positioning information, usually sampled at 10Hz, by providing details regarding device orientation and acceleration. The black boxes predominantly utilize modules such as Quectel Mc60 or U-blox, or other variants from different brands. Data acquisition is made through the reception of NMEA (National Marine Electronics Association) messages, specifically in the $GNRMC format. These messages, collected approximately every 2 kilometers, contain essential data including UTC time, latitude, longitude, instantaneous speed and heading. The 2 kilometers interval ensures that data transmission from sensors restricts the centralized server memory usage while maintaining proportional data collection aligned with the study's objectives. Acceleration data, indicative of 'harsh events,' is asynchronously recorded when acceleration values exceed |12 km/(h.s)| (0.3g), providing valuable insights into vehicle behavior. The GNSS receivers and in general the black boxes are continuously powered, drawing energy directly from the vehicle's battery or, in some instances, from an onboard battery source. This uninterrupted power supply ensures seamless data collection throughout the whole observation period.

Dataset

Our dataset consists of around 11000 private vehicles and 25Mio trips between January and September 2022 in Italy. The dataset is not publicly available, and has been made accessible through a joint research project with UnipolTech (Unipol Group), the technology company of the second insurance group in Italy.

The starting dataset consists of GPS measurements collected at $1\ Hz$ by firmly installed telematic black boxes. Every 2km, the sample, denoted as *event*, is sent to a company's server and contains the following information: Id vehicle, Event Type (0 start, 1 active, 2 stop), Timestamp, Latitude, Longitude, Distance Traveled since the previous event, Instantaneous Speed, Heading GPS, Road Type. Exploiting Id vehicle and Event Type, we are able to aggregate, for each vehicle, this event-based information into a complete trip-based one. To achieve this goal, sequences of start-active-stop events (0,1,...,2) are identified and categorized as a trip. This leads to a trips-based dataset containing the necessary information for the aim of this study: Id Vehicle, Id Trip, Time Start, Time End, Latitude Start, Longitude Start, Latitude End, Longitude End, Distance, Duration, Road Type, Average Speed.

In addition to trip-specific information, we have another raw dataset, which records asynchronously for each vehicle (identified by Id vehicle), the acceleration and braking events exceeding the threshold of |0.3|g. In this way, the trips-based dataset is augmented with information on all harsh events.

Furthermore, we also have the original vehicle information - car model, manufacturer, type of fuel, and year of registration - that can be linked to the trip-based dataset via the Id Vehicle field. This information enables access to public databases concerning Euro class, declared consumption and emissions.

Data preprocessing

In our analysis, we focus on internal combustion engine vehicles powered by either petrol or diesel engines. We apply specific pre-processing criteria to the dataset to ensure its relevance and feasibility for our study. The following criteria are employed for data cleaning:

1. Trips with a total distance shorter than 500 m are excluded.
2. Trips with a total time shorter than 5 minutes or longer than 12 hours are removed.
3. Trips with an average speed lower than 5 km/h or higher than 150 km/h are considered unfeasible and are eliminated.Furthermore, vehicles with more than 70% of trips eliminated during this phase are also excluded from the analysis. As a result of the above criteria, the dataset is refined to encompass 8457 vehicles (87% diesel, 13% petrol) and approximately 11Mio trips.

Data availability

The dataset used and analysed during the current study cannot be made publicly available since it is obtained from third parties, within a funded research collaboration with UnipolTech Spa. The research contract contains a confidentiality clause stating that Politecnico di Milano is obligated to observe secrecy towards any person not authorized by UnipolTech Spa. This applies to data, information, knowledge, documents identified as confidential that have been communicated by UnipolTech Spa under this contract. Data may eventually be made available from the corresponding author in a partial but exhaustive version, on reasonable request.

Received: 1 February 2024; Accepted: 9 October 2024
Published online: 16 October 2024

References

1. Mobility strategy. Tech. Rep., European Commission (2021).
2. Sustainable and smart mobility strategy - putting european transport on track for the future. Tech. Rep., European Commission (2020).
3. Tsavachidis, M. & Petit, Y. L. Re-shaping urban mobility - key to europe's green transition. *Journal of Urban Mobility* **2**, 100014. https://doi.org/10.1016/j.urbmob.2022.100014 (2022).
4. Scott V., G. O. The challenge of carbon dioxide removal for eu policy-making. *Nat Energy* **3**, 350–352, https://doi.org/10.1007/s001090000086 (2018).
5. Neuhoff, K. et al. *Staying with the leaders: Europe's path to a successful low-carbon economy* (Tech. Rep, Climate Strategies, 2014).
6. Transport and environment report 2021. Tech. Rep., European Environment Agency (2022).
7. Ritchie, H. Cars, planes, trains: where do co2 emissions from transport come from? *Our World in Data* (2020).
8. The eu transport sector and its contribution to reaching climate neutrality. Tech. Rep., European Commission (2022).
9. Plötz, P., Funke, S. A., Jochem, P. & Wietschel, M. Co2 mitigation potential of plug-in hybrid electric vehicles larger than expected. *Scientific Reports* **7**, 16493. https://doi.org/10.1038/s41598-017-16684-9 (2017).
10. Morfeldt, J. & Johansson, D. J. A. Impacts of shared mobility on vehicle lifetimes and on the carbon footprint of electric vehicles. *Nature Communications* **13**, 6400. https://doi.org/10.1038/s41467-022-33666-2 (2022).
11. Woody, M., Keoleian, G. A. & Vaishnav, P. Decarbonization potential of electrifying 50% of u.s. light-duty vehicle sales by 2030. *Nature Communications* **14**, 7077, https://doi.org/10.1038/s41467-023-42893-0 (2023).
12. Global ev outlook 2020. Tech. Rep., IEA (2020).
13. Arriola, C. *et al.* Challenges to international trade and the global economy. *OECD Trade Policy Papers*[SPACE]https://doi.org/10.1787/5c561274-en (2023).
14. Repowereu: A plan to rapidly reduce dependence on russian fossil fuels and fast forward the green transition*. Tech. Rep., European Commission (2022).
15. Zhang, C., Zhao, X., Sacchi, R. & You, F. Trade-off between critical metal requirement and transportation decarbonization in automotive electrification. *Nature Communications* **14**, 1616. https://doi.org/10.1038/s41467-023-37373-4 (2022).
16. Singh, M., Yuksel, T., Michalek, J. J. & Azevedo, I. M. L. Ensuring greenhouse gas reductions from electric vehicles compared to hybrid gasoline vehicles requires a cleaner u.s. electricity grid. *Scientific Reports* **14**, 1639, https://doi.org/10.1038/s41598-024-51697-1 (2024).
17. Singh, H. & Kathuria, A. Profiling drivers to assess safe and eco-driving behavior - a systematic review of naturalistic driving studies. *Accident Analysis & Prevention* **161**, 106349. https://doi.org/10.1016/j.aap.2021.106349 (2021).
18. Warren, J., Lipkowitz, J. & Sokolov, V. Clusters of driving behavior from observational smartphone data. *IEEE Intelligent Transportation Systems Magazine* **11**, 171–180. https://doi.org/10.1109/MITS.2019.2919516 (2019).
19. Tsavachidis, M. & Petit, Y. L. Re-shaping urban mobility - key to europe's green transition. *Journal of Urban Mobility* **2**, 100014. https://doi.org/10.1016/j.urbmob.2022.100014 (2022).
20. Massoud, R., Bellotti, F., Berta, R., De Gloria, A. & Poslad, S. Eco-driving profiling and behavioral shifts using iot vehicular sensors combined with serious games. In *2019 IEEE Conference on Games (CoG)*, 1–8, https://doi.org/10.1109/CIG.2019.8847992 (2019).
21. Nasir, M., Kalam, M. A., Masum, B. & Md. Noor, R. Reduction of fuel consumption and exhaust pollutant using intelligent transport system. *The Scientific World Journal*[SPACE]https://doi.org/10.1155/2014/836375 (2014).
22. Tutuianu, M. et al. Development of the world-wide harmonized light duty test cycle (wltc) and a possible pathway for its introduction in the european legislation. *Transportation Research Part D: Transport and Environment* **40**, 61–75. https://doi.org/10.1016/j.trd.2015.07.011 (2015).
23. Fontaras, G., Zacharof, N.-G. & Ciuffo, B. Fuel consumption and co2 emissions from passenger cars in europe - laboratory versus real-world emissions. *Progress in Energy and Combustion Science* **60**, 97–131. https://doi.org/10.1016/j.pecs.2016.12.004 (2017).
24. Thomas, J., Huff, S., West, B. & Chambon, P. Fuel consumption sensitivity of conventional and hybrid electric light-duty gasoline vehicles to driving style. *SAE International Journal of Fuels and Lubricants* **10**, https://doi.org/10.4271/2017-01-9379 (2017).
25. Aydın, M. & Soruşbay, C. A methodology to creating semi-artificial city driving cycle: case study of istanbul. *Journal of Engineering and Applied Science* **69**, 77. https://doi.org/10.1186/s44147-022-00132-1 (2022).
26. Barth, M. J. & Boriboonsomsin, K. *Real-world carbon dioxide impacts of traffic congestion* (UC Berkeley University of California Transportation Center, Tech. Rep., 2010).
27. Ngure, S., Ochieng, F. & Mailutha, J. Fuel consumption model for road distribution of petroleum products in kenya. *IOSR Journal of Environmental Science, Toxicology and Food Technology* **8**, 56–60. https://doi.org/10.9790/2402-08145660 (2014).
28. Schmied, M. The european handbook of emission factor for road transport (hbefa) (2014).
29. Shahariar, G. H. et al. Impact of driving style and traffic condition on emissions and fuel consumption during real-world transient operation. *Fuel* **319**, 123874. https://doi.org/10.1016/j.fuel.2022.123874 (2022).
30. Adamidis, F. K., Mantouka, E. G. & Vlahogianni, E. I. Effects of controlling aggressive driving behavior on network-wide traffic flow and emissions. *International Journal of Transportation Science and Technology* **9**, 263–276. https://doi.org/10.1016/j.ijtst.2020.05.003 (2020).
31. Explaining road transport emissions. *Tech* (Rep, European Environment Agency, 2016).
32. Air pollutant emissions standards. https://ec.europa.eu/commission/presscorner/detail/en/MEMO_15_5705 (2015).
33. Prati, M. V., Costagliola, M. A., Zuccheroso, A. & Napolitano, P. Assessment of euro 5 diesel vehicle nox emissions by laboratory and track testing. *Environmental Science and Pollution Research* **26**, https://doi.org/10.1007/s11356-019-04486-7 (2019).
34. Emission standards. https://dieselnet.com/standards/eu/ld.php (2022).
35. Mickūnaitis, V., Pikūnas, A. & Mackoit, I. Reducing fuel consumption and co2 emission in motor cars. *TRANSPORT* **22**, 160–163. https://doi.org/10.3846/16484142.2007.9638119 (2007).

Acknowledgements

The research is supported by UnipolTech Spa, a private company that designs, develops, and delivers innovative services based on telematics, Internet of Things and Big Data in the service of insurance and mobility. UnipolTech Spa is part of Unipol Group and collaborates with our Politecnico di Milano research team intensively on the mobility ecosystem research. We thank UnipolTech for supporting this studies.

Author contributions

Designed the study: S.M.S., S.C.S., A.P.; developed and performed the analysis: A.P., S.C.S.; drafted the paper: A.P., S.C.S.; contributed to writing: A.P., S.C.S.; all authors have read and approved the final manuscript.

Declarations

Competing interests

The authors declare no competing interests. Correspondence and requests should be addressed to S.C.S.

Additional information

Correspondence and requests for materials should be addressed to S.C.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.