Greenhouse gas emissions and road infrastructure in Europe: A machine learning analysis

Transportation Research Part D 139 (2025) 104602

Contents lists available at ScienceDirect

Transportation Research Part D

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

Greenhouse gas emissions and road infrastructure in Europe: A machine learning analysis☆

Cosimo Magazzino[a,b,*], Alberto Costantiello[c], Lucio Laureti[c], Angelo Leogrande[c], Tulia Gattone[d,e]

[a] Department of Political Science, Italy, Roma Tre University
[b] Economic Research Center, Western Caspian University, Baku, Azerbaijan
[c] LUM University "Giuseppe Degennaro", Casamassima, Italy
[d] Department of Economics, John Cabot University, Rome, Italy
[e] Syracuse University in Florence, Florence, Italy

A R T I C L E I N F O

JEL Codes:
Q51
Q52
R40
R42
R48

Keywords:
Greenhouse gas emissions
Road transportation
Panel data
Machine Learning
Europe

A B S T R A C T

This paper explores the determinants of greenhouse gas (GHG) emissions in Europe, focusing on transportation-related variables. By combining classical econometric models with Machine Learning (ML) techniques, we analyze data spanning from 2013 to 2021. The empirical findings highlight the complex relationship between newer passenger cars and GHG emissions, noting the significant impact of their production and increased usage. Conversely, the adoption of alternative fuel vehicles is found to significantly reduce emissions. This is further supported by ML models, which emphasize the critical role of car density and alternative fuel vehicles in determining emissions. Policy implications suggest the need for targeted interventions, including the promotion of electric and hybrid vehicles, enhancements in transportation infrastructure, and the implementation of economic incentives for clean technologies.

1. Introduction

The transportation sector's significant contribution to greenhouse gas (GHG) emissions has been extensively documented (Albuquerque et al., 2020), particularly in Europe, where road infrastructure is diverse and where mopeds and motorcycles are popular means of transportation (Ntziachristos et al., 2006). Road transportation, encompassing a broad range of vehicles and extensive networks, plays a critical role in economic activities but also poses substantial environmental challenges due to its significant emissions

Abbreviations: ANN, Artificial Neural Networks; BIC, Bayesian Information Criterion; DPD, Dynamic Panel Data; GHG, Greenhouse Gas; MAE, Mean Absolute Error; ML, Machine Learning; MOPEDS, Mopeds with Non-Petrol Engines; MSE, Mean Squared Error; POLS, Pooled Ordinary Least Squares; PDPs, Partial Dependence Plots; OR, Length of Other Roads Outside Built-Up Areas; OVER125, Motorcycles with a Displacement Greater than 125cc; PC1000, Passenger Cars per Thousand Inhabitants; PC10TO20, Passenger Cars 10 to 20 Years Old; PCLESS2, Passenger Cars Less Than 2 Years Old; RMSE, Root Mean Squared Error; SVM, Support Vector Machines; TRAM, Number of Trams; WLS, Weighted Least Squares.

☆ All potential errors and opinions are ours. The standard disclaimers apply.

* Corresponding author.
E-mail addresses: cosimo.magazzino@uniroma3.it (C. Magazzino), costantiello@lum.it (A. Costantiello), laureti@lum.it (L. Laureti), leogrande.cultore@lum.it (A. Leogrande), tulia.gattone@johncabot.edu (T. Gattone).

https://doi.org/10.1016/j.trd.2025.104602
Received 30 July 2024; Received in revised form 3 December 2024; Accepted 8 January 2025
Available online 13 January 2025
1361-9209/

(Al-Foraih & Sreekanth, 2021). The European road network, including major highways, urban roads, and rural byways, supports various types of vehicular traffic, from personal cars to heavy-duty trucks, each contributing differently to the emissions profile (Clairotte et al., 2020). The complexity of this network and the diversity of vehicle types, ranging from highly efficient electric vehicles to older ones, which are inferior in terms of exhaust emissions and fuel consumption, necessitate nuanced strategies for reducing GHG emissions (Lejda et al., 2021).

Different vehicle categories have distinct roles and exhibit varying fuel efficiencies and emissions profiles, adding layers of complexity to emissions reduction efforts (Alhindawi et al., 2020). Furthermore, well-to-wheel emissions, including those from plug-in electric vehicles, contribute to the release of nitrogen oxides and particulate matter (Gan et al., 2021). This diversity and intricacy underscore the necessity for strategies to effectively mitigate emissions within the transportation sector and across other economic sectors, as uniform policies may fail to address the unique dynamics inherent in transportation (Pietrzak & Pietrzak, 2020).

Regional variations further complicate the development of effective emission reduction strategies. Given the heterogeneity of socio-economic features and transportation patterns, effective policies, characterized by high traffic volumes and industrial activities, may not apply to other regions (He et al., 2021). For instance, regions characterized by higher income, increased vehicle ownership, and greater transport volumes exhibit higher emission rates compared to those undergoing fuel transition, experiencing increased fuel prices, or marked by higher population density and urbanization rates (Lim et al., 2020). Additionally, the impact of the national car and supplier industry, ecological modernization, and fuel tourism influence dieselization rates across Europe (Cames & Helmers, 2013). These differences underscore the importance of localized environmental impacts and the need for tailored strategies that reflect specific regional conditions and priorities.

The originality of this study lies in its innovative approach to analyzing the relationship between GHG emissions and road transport in Europe, offering a comprehensive perspective that integrates environmental and transportation data. Unlike previous research that examined emissions in isolation, this study explores the interconnectedness of emissions and transportation infrastructure, providing a nuanced understanding of how policies and systems affect environmental outcomes. Additionally, this study advances the field by employing Machine Learning (ML) algorithms, enhancing the ability to forecast future emission trends based on current and historical data. ML techniques can handle complex datasets with multiple interacting variables, uncovering patterns that traditional statistical methods might miss (Shao et al., 2023; Breiman, 1996).

In conclusion, this study emphasizes the need for a comprehensive understanding of the relationship between road transportation and GHG emissions in Europe. By using advanced econometric and ML techniques, it provides valuable insights into the factors driving emissions and offers a detailed analysis to inform the development of effective and sustainable transportation policies. The findings highlight the importance of integrating environmental and transportation policies to achieve sustainable outcomes and contribute to the global effort to combat climate change.

The article is organized as follows: Section 2 reviews the relevant literature, Section 3 details the data and the economic strategy employed, Section 4 reports the main empirical findings and their discussion, Section 5 discusses the robustness checks via artificial neural networks, and Section 6 concludes the study.

2. Literature review

The complexity of road transportation in Europe and its environmental implications necessitate a detailed and multidisciplinary analysis. Following the COVID-19 pandemic, studies highlight the need for an integrated approach, considering transportation market maturity, budget constraints, and equity considerations, to understand the intricate dynamics between transportation and GHG emissions (Kéry et al., 2021). Employing augmented optimization models is essential to assess the economic, energy, and environmental impacts of policy interventions on emissions, integrating diverse data sources for a comprehensive view (Rauf & Umer, 2023).

A thorough analysis of the technologies available for decarbonization, the infrastructure required to support electrification, and policy options is crucial for developing effective regulations and addressing environmental challenges in Europe (Creutzig et al., 2012). These studies underscore the importance of an interdisciplinary approach, combining economic, technological, and environmental data from different sectors to develop more successful strategies (Omranian et al., 2023). However, this research aims to address a gap in understanding the specific regional impacts and the effectiveness of tailored policy measures.

Technological innovations play a relevant role in reducing emissions from road transportation by enhancing efficiency and minimizing overall environmental impact. For example, the integration of lithium-ion batteries and fuel cell hybrid power systems in electric vehicles can significantly enhance urban transport sustainability (Casals et al., 2016). Hadi and Al Hasibi (2022) propose an integrated system for supplying electricity and hydrogen for road transportation, aiming to reduce emissions through the synergistic use of renewable energy. Additionally, Lee et al. (2023) emphasize the benefits of vehicle-to-everything communications for traffic management, which indirectly reduces emissions through improved fuel efficiency. These technological advancements are crucial for achieving significant reductions in emissions from road transport (Aminzadegan et al., 2022) and meeting Europe's ambitious emission reduction targets. Yet, the research needs to delve deeper into the implementation barriers and real-world effectiveness of these technologies, particularly in diverse urban and rural settings.

Across European countries, adopting advanced technologies and novel methodologies for calculating the carbon footprint is essential for achieving emission reduction targets and promoting sustainable transport (Roukounakis et al., 2020). Cohesive and comprehensive renewable energy policy systems are critical for developing zero-emission GHG targets while maintaining air quality. Although maritime transport generally contributes relatively little to the transport sector's CO_2 emissions footprint, the volume of freight transported necessitates localized policy frameworks for effective emission reduction strategies (Nusa & Kodak, 2023). Local governments play a key role in promoting sustainable practices through integrated policy frameworks that address emissions and broader environmental impacts. They are also essential in developing comprehensive emission inventories across the country (Gogeri & Gouda, 2024).

However, the large-scale and distributed nature of vehicle emissions makes inventory methods particularly challenging, which is why ML models using indirect top-down estimation of road transport emissions can address this challenge (Mukherjee et al., 2021). These alternative approaches and user-generated datasets (Li, 2021) are vital for effective policy implementation and sustainable development in regions where GHG emissions related to international trade are difficult to quantify. Further research is necessary to explore the interplay between regional policies and their effectiveness in reducing emissions.

Advanced predictive models are crucial for forecasting future emission trends and planning effective mitigation strategies. For instance, an important topic in the empirical debate is calculating the trade-off between the decarbonization potential of the road transportation sector and its critical metal requirements from the demand-side perspective, with current estimates suggesting that electric vehicles by 2050 might increase lithium, nickel, cobalt, and manganese demands by more than 1,000% (Zhang et al., 2023). Therefore, integrating ML algorithms is pivotal for providing accurate predictions that inform effective emission reduction strategies. These advanced techniques allow for more precise forecasting and better policy planning, which is critical for addressing the dynamic and complex nature of GHG emissions from road transport in Europe. Recent studies also emphasize the role of uncertainty in current on-road emission and energy consumption modeling approaches (Wang et al., 2020). Despite some advances, gaps remain in integrating these models with real-time data and in their application across different European regions.

Policy strategy in this field necessitates a comprehensive and effective well-to-wheel emission analysis and quantification, optimal route design, integration of sustainable transport modes, and clean energy generation (Vallarta-Serrano et al., 2023). For instance, evaluating the mitigation effect of replacing highway goods transport with conventional railways, as examined by Lin et al. (2021), through scenario analysis by Matsuhashi & Ariga (2016) using population cell density categories, could yield groundbreaking insights in energy research. Despite various solutions for road transport, electric vehicles stand out as a promising option, strongly supported by car manufacturers and research entities, to gradually replace conventional vehicles (Atabay et al., 2022; Mebarki et al., 2020). Conducting a lifecycle analysis is crucial for assessing the full environmental impact of electric transportation systems, including emissions from the production, operation, and disposal of vehicles. These assessments, which encompass socioeconomic and spatial determinants, are essential for identifying the most effective strategies to minimize the environmental footprint of transportation (Baur et al., 2015). Nevertheless, there is a pressing need to standardize these assessments and promote their widespread adoption across Europe.

Balancing economic growth with environmental protection is essential for road freight transportation, as highlighted by Domagała and Kadłubek (2022) and Axsen et al. (2020). Optimizing logistics, improving fuel efficiency, and adopting greener technologies, using both open or closed transfer tubes to the dilution tunnel configurations, can reduce the environmental footprint of road transport while supporting economic growth, with population density playing an antagonistic role between road density and environmental sustainability (Din et al., 2023; Giechaskiel, 2020). These complexities and shifting variable roles require advanced tools, like system dynamic modeling and scenario analysis, capable of properly evaluating CO_2 emissions limits and supporting sustainable development across Europe (Al-Osaimi et al., 2020; Zervas, 2010). Nonetheless, more research is needed to explore the cost-effectiveness and scalability of these green logistics solutions.

The role of international cooperation and knowledge sharing, for comprehensive policy options like subsidizing environmentally friendly technologies, developing green transport infrastructure, and enacting decarbonizing regulations, is also paramount in the pursuit of sustainable transportation solutions (Habib et al., 2021). European countries can benefit significantly from collaborative efforts to standardize policies and share best practices in reducing GHG emissions from road transportation. This is particularly important for the development and implementation of technologies and policies that address the specific needs of diverse regions and the environmental efficiency of high-capacity transportation, from densely populated urban centers to rural areas with limited road infrastructure (Palander et al., 2021). Sharing technology and policy insights in economic sectors such as energy, industry, buildings, transport, and land use can help accelerate the transition to low-emission transportation systems, contributing to broader environmental sustainability goals (Lamb et al., 2021).

In summary, this review highlights the importance of integrating diverse methodologies to understand the factors influencing emissions and develop localized policy frameworks. These insights are crucial for creating effective strategies to reduce emissions and promote sustainable transportation in Europe. The review underscores the necessity of combining technological innovation with robust policy frameworks to address the complex challenges of GHG emissions from road transportation. By leveraging advanced predictive models, fostering international cooperation, and promoting region-specific solutions, Europe can achieve its ambitious goals of reducing emissions and advancing sustainable transportation practices. Filling these gaps is essential for developing a holistic approach that effectively addresses the multifaceted challenges of road transportation emissions.

The findings from this review emphasize that a multifaceted approach, incorporating both computational technological advancements and comprehensive policy measures, is essential for effectively mitigating the environmental impacts of road transportation in Europe. Future research should continue to explore the integration of renewable energy systems, the development of low-emission vehicle technologies, and the implementation of sustainable infrastructure improvements (Mukherjee et al., 2023; Ho et al., 2022). Furthermore, the role of behavioral changes and local public engagement in promoting sustainable transportation practices should not be overlooked, as they are crucial for ensuring the long-term success of emission reduction strategies (Handayani et al., 2019; Santopietro et al., 2022). Addressing these gaps will enhance the effectiveness of emission reduction strategies and support the overarching goals of sustainable development.

3. Data and empirical strategy

Data for this study were obtained from the Eurostat database, which contains comprehensive and reliable statistics relevant to European countries. The dataset includes a range of variables that are crucial in examining GHG emissions in Europe, focusing on the period 2013–2021. The analysis encompasses a total of 30 European countries: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and Switzerland.

The selected key variables for the econometric model are summarized in Table 1.

The selection of variables in this analysis is crucial for understanding the multifaceted impacts of transportation on GHG emissions across European countries. Each variable is strategically chosen to represent different aspects of transportation infrastructure, vehicle use, and their subsequent environmental implications.

- GHG **Emissions (GHG)**: GHG emissions serve as the dependent variable in this analysis, representing the total net emissions of GHG. This metric is pivotal as it encapsulates the comprehensive environmental impact of transportation activities and infrastructure across various European regions. When considered alongside exhaust emissions of nitric oxide (NO_x), nitrogen dioxide (NO_2), and $PM_{2.5}$, it provides a critical measure of environmental performance and policy efficacy in addressing climate change (Kousoulidou et al., 2008).
- **Passenger Cars Less Than 2 Years Old (PCLESS2)**: This variable denotes the number of passenger cars under two years old. Although newer cars typically feature more advanced fuel-efficient technologies and lower operational emissions, these vehicles' production and distribution processes contribute substantially to GHG emissions. Thus, despite their operational efficiency, the net environmental impact includes significant emissions from manufacturing (Miravete et al., 2018). This variable is expected to show a complex relationship with GHG emissions, balancing the lower emissions during use against the high emissions from production or poor pollution abatement strategies (Costagliola et al., 2014). As shown in Fig. A.1 in the Appendix, the relationship between GHG and new passenger cars in Europe is complex and multifaceted.
- **Passenger Cars 10 to 20 Years Old (PC10TO20)**: This variable captures the number of older passenger cars, which are generally less fuel-efficient and produce higher emissions due to outdated technologies. Older vehicles often lack advanced emission control technologies and are more likely to be poorly maintained, leading to higher emissions over their operational lifetime. Consequently, a positive correlation with GHG emissions is anticipated, reflecting their propensity to emit more pollutants (Hankey & Marshall, 2010). The relationship between GHG and vehicles aged between 10 and 20 years is illustrated in Fig. A.2 in the Appendix
- **Number of Trams (TRAM)**: This variable reflects the total number of trams in operation within a region and indicates the extent to which a region has invested in efficient and environmentally friendly public transportation alternatives. Due to their high capacity and reliance on electricity, often derived from renewable sources, trams contribute to reduced GHG emissions by decreasing dependence on private cars and mitigating urban traffic congestion (Sarigiannis et al., 2017; Nanaki et al., 2017). Consequently, a negative relationship between the number of trams and GHG emissions is expected. Fig. A.3 in the Appendix shows the relationship between GHG and trams in European countries.
- **Passenger Cars per Thousand Inhabitants (PC1000)**: This variable measures car ownership density, which can signify more extensive vehicle use and potentially lead to increased emissions. However, regions with higher car ownership might also possess advanced infrastructure that mitigates emissions through better traffic management and lower per capita car usage (Martin & Shaheen, 2011). Thus, the relationship with GHG emissions can be complex, depending on the balance between ownership and infrastructure efficiency (Luo et al., 2017). The relationship between GHG emissions and passenger cars per thousand inhabitants is depicted in Fig. A.4 in the Appendix.
- **Length of Other Roads Outside Built-Up Areas (OR)**: This variable measures the extent of road infrastructure beyond urban areas. Extensive road networks can facilitate efficient transport and alleviate urban congestion, potentially leading to lower GHG emissions if they reduce overall vehicle idling and congestion (Barrington-Leigh & Millard-Ball, 2017). However, increased road length can also encourage higher vehicle usage, potentially offsetting the benefits if not managed properly with strategies such as

Table 1
Variables' description.

Acronym	Definition	Years	Obs.	Countries
GHG	Net Greenhouse Gas Emissions	2013–2021	270	Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland
PCLESS2	Passengers' cars less than 2 years		270	
PC10TO20	Passengers' cars 10 to 20 years		270	
TRAM	Number of Trams		270	
PC1000	Passenger cars per thousand inhabitants		270	
OR	Length of other roads outside built-up areas		270	
MOPEDS	Mopeds with non-petrol engines		270	
OVER125	Motorcycles with a displacement greater than 125		270	

electrifying cars and buses and renewing older trucks (Minet et al., 2021). Moreover, suburban residents tend to be less sensitive to the built environment policies than urban residents (Tao & Cao, 2024). Fig. A.5 in the Appendix illustrates the relationship between GHG emissions and the length of roads outside built-up areas.

- **Mopeds with Non-Petrol Engines (MOPEDS):** This variable indicates the number of mopeds using alternative, typically cleaner, fuels. Mopeds with non-petrol engines are expected to have a negative relationship with GHG emissions, given their lower emissions than petrol-powered counterparts. Still, great differences in terms of emissions are observed across mopeds of different engine types (Adam et al., 2010). The shift towards electric or hybrid vehicles reflects a broader trend towards more sustainable transportation options contributing to reduced emissions (Gialos et al., 2022; Camargo-Caicedo et al., 2021). The relationship between GHG emissions and non-petrol mopeds in Europe is shown in Fig. A.6 in the Appendix.
- **Motorcycles with Engine Displacement Greater than 125cc (OVER125):** This variable represents the number of larger motorcycles. Despite the general efficiency of motorcycles compared to cars, larger models might have higher emissions per kilometer, particularly if they are high-performance vehicles with higher fuel consumption (Giechaskiel, 2020). This variable is expected to show a complex relationship with GHG emissions, potentially indicating a negative correlation due to higher fuel efficiency and reduced congestion, but a positive correlation if these vehicles are extensively used and contribute to higher total emissions (Hankey & Marshall, 2010). Fig. A.7 in the Appendix demonstrates the relationship between GHG emissions and motorcycles over 125 cm^3 in Europe.

The descriptive statistics for these variables, as illustrated in Fig. 1, provide a foundational understanding of data distribution and central tendencies across different European countries.

The histograms and density plots highlight substantial variability in GHG emissions among European countries, showing a right-skewed distribution where a few countries have very high emissions due to intensive industrial activities and heavy reliance on fossil fuels.

The spread of PCLESS2 across Europe highlights economic disparities and consumer preferences. Conversely, PC10TO20 is bimodal, suggesting significant numbers of older vehicles in some countries, driven by economic challenges and lenient emission regulations. In addition, TRAM varies significantly. Moreover, PC1000 indicates a balance between car ownership and robust public transport systems, while OR varies, with some countries investing heavily in rural road networks to alleviate urban congestion. Furthermore, the shift towards more sustainable urban transport solutions is evident from the distribution of MOPEDS and OVER125, which help reduce the transport sector's environmental impact in Europe.

The descriptive statistics in Fig. 1 not only highlight variability across key transportation and GHG emission variables but also reveal significant disparities in infrastructure and technology adoption across European countries. These patterns suggest that emissions are influenced by diverse economic, technological, and policy factors, which manifest in the clustering of countries with similar profiles. For instance, the skewness in PCLESS2 might reflect variations in consumer preferences and economic capacity, while

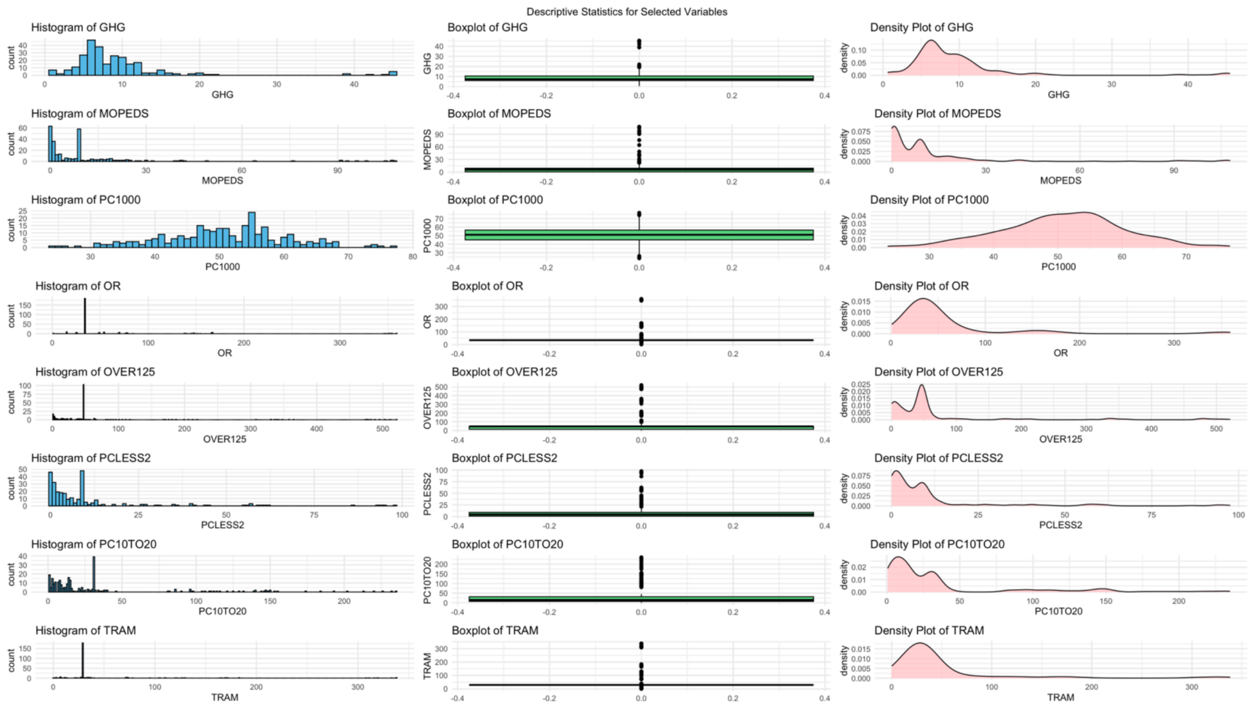

Fig. 1. Descriptive statistics.

the bimodal distribution of PC10TO20 may emphasize the coexistence of outdated vehicle fleets alongside newer technologies.

The empirical correlations between the variables and GHG emissions, as shown in Fig. 2, largely align with theoretical expectations.

The correlation between PCLESS2 and GHG emissions is −0.019, slightly deviating from the expected positive relationship, suggesting an impact that is nuanced. The value for PC10TO20 is −0.075, which was unexpectedly negative, demonstrating that factors other than age and technology of vehicles affect emissions. The negative value of TRAM at −0.034 further supports the theoretical expectation that trams reduce emissions. Overall, Fig. 2′s correlation plot clarifies nuanced relationships existing between transportation variables and GHG emissions. The weak and sometimes unexpected correlations, like that between PC10TO20 and GHG emissions, point to the need for nonlinear relationships to be included in this research. The observed negative correlations, especially for variables such as TRAM, reinforce theoretical expectations about the role of sustainable transport systems in reducing emissions.

In our analysis of GHG emissions across European countries, we employed both trend analysis and clustering using the k-Means algorithm, a robust method for grouping data based on similarities. The k-Means algorithm, an unsupervised ML technique, is widely used in environmental and economic research for its ability to reveal underlying patterns in complex datasets (Jain, 2010). Before diving into empirical models, clustering helps in understanding the heterogeneity among the countries by categorizing them into groups with similar emission profiles, facilitating targeted policy recommendations, and comparative analysis.

The clustering process was optimized using the Silhouette Coefficient, a measure of how similar an object is to its own cluster compared to other clusters, to determine the most suitable number of clusters (Kaufman & Rousseeuw, 1990). The optimal number of clusters was identified by comparing Silhouette coefficients for different values of k, ranging from two to eight. This approach ensures that the clustering captures meaningful groupings rather than arbitrary divisions, enhancing the interpretability and relevance of the analysis. The results indicated that a four-cluster solution (k = 4) was optimal, providing the highest Silhouette coefficient, as depicted in Fig. 3.

Fig. 3 shows the relationship between the number of clusters (k) and the Silhouette coefficient, underlying how k = 4 yields the highest clustering quality. This optimal cluster size was therefore selected for our analysis. Therefore, we divide the 30 European countries into four distinct clusters based on GHG emissions from 2013 to 2021 to enable a fine-grained understanding of the patterns of emissions and their drivers.

The clustering results are as follows:

- **Cluster 0** includes Belgium, Czechia, Denmark, Germany, Greece, Cyprus, Netherlands, Austria, Poland, Slovenia, and Finland. These countries are characterized by significant, yet stable, GHG emissions, suggesting effective regulatory frameworks and advanced technologies that mitigate environmental impacts despite high industrial activity. In particular, it records an increase of rather stable average emissions, as its values increase from 10.01 in 2013 to 10.35 in 2018 and later decreased to 8.65 in 2020 but

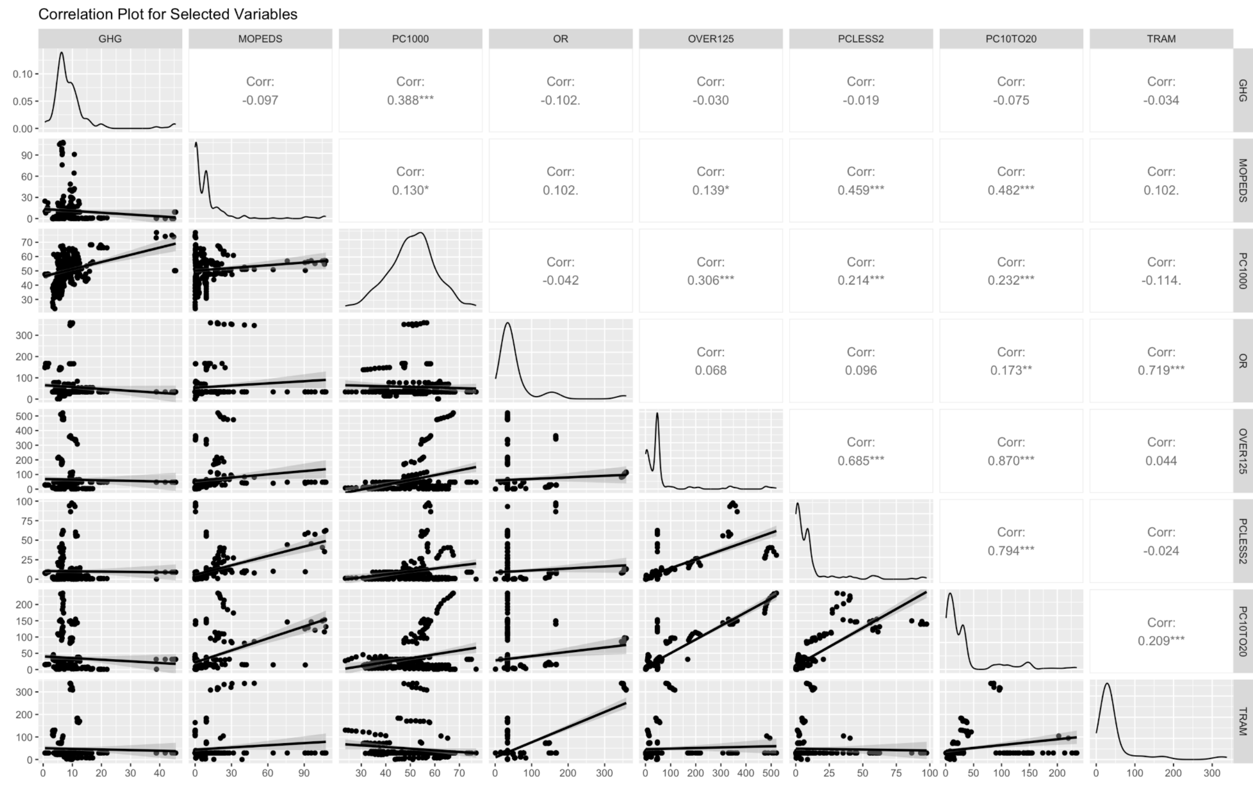

Fig. 2. Correlation plot.

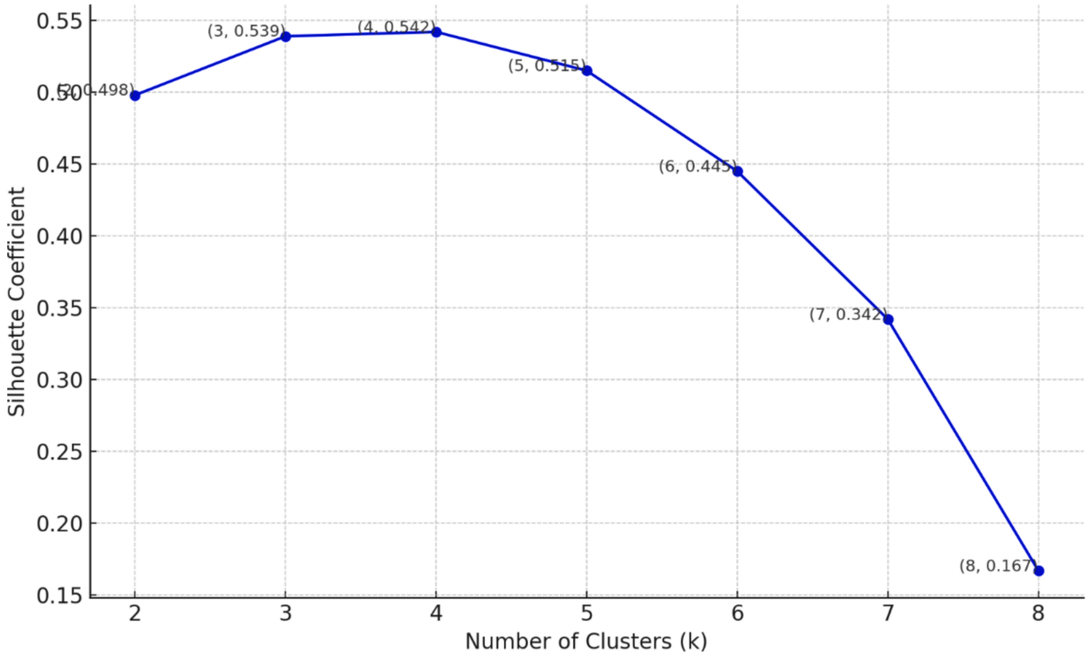

Fig. 3. The relationship between the number of k and the Silhouette coefficient.

manage a partial recovery in 2021 at 9.07. The medians are quite similar to the mean, therefore, in this case, a prior assumption of symmetric distribution of emission among the countries within the cluster can be assumed. Hence, such a peak in 2018 and the fall back by 2020 may relate to some relevant policy intervention or changes in industrial outputs during this period.

- **Cluster 1** consists solely of Iceland, reflecting its unique energy mix heavily reliant on geothermal and hydroelectric power. Although these sources are renewable, they pose environmental challenges due to geological and infrastructural factors. Starting with 45.5 in 2013, the figures have fallen gradually to 38.9 by 2021. This shows that both means and medians of the year have the same value for explaining that there is no extreme variation in Iceland's emission figure. In light of this, a continuing fall provides evidence for striving toward renewable energy by Iceland and, hence, reductions in reliance on fossil fuels.
- **Cluster 2** comprises Estonia, Ireland, and Luxembourg, which show a trend of declining emissions likely due to proactive policy implementation and shifts towards sustainable energy sources. These countries have moderate emissions, with an average peak of 17.37 in 2018, then declined to 13.4 in 2020, then slightly recovered to 14.17 in 2021. Its medians, except for very few years or most years in the clusters, are lower in value compared with the means for this cluster. Hence, the distribution shape is right-skewed, indicating that very few countries showed high emissions. The decline beyond 2018 represents either the implementation of emissions control policies or economic factors affecting industrial activity. This cluster highlights the potential of targeted environmental strategies.
- **Cluster 3** is the most heterogeneous cluster, including Bulgaria, Spain, France, Croatia, Italy, Latvia, Lithuania, Hungary, Malta, Portugal, Romania, Slovakia, Sweden, Norway, and Switzerland. These countries exhibit varying emission trends driven by different stages of economic development, energy dependencies, and transitional energy policies. The mean for this cluster varies from 5.56 in 2013, peaking at 5.93 in 2017, then bottoming out to 4.89 in 2020, before a partial recovery in 2021 to 5.28. In all cases, however, the medians are higher than the means, which reflects that the underlying distribution for these data is left-skewed; that is, some countries have much lower emissions compared to the cluster average. Some countries in this cluster are advancing in renewable energy adoption and economic transitions, while others are in the early stages or experiencing fluctuations in progress.

Detailed descriptive statistics for each cluster are provided in the online Supplementary Materials. Overall, the data underlines variable success in emissions reduction across clusters, likely driven by differences in policy implementation, economic factors, and energy resource utilization. Smaller overall reductions in emissions, or very flat trends in some clusters, may also reflect the need to adopt more enhanced policy frameworks, while larger reductions reflect possibilities for effective targeted environmental strategies. The trends go further to show what was expected from the impacts of the COVID-19 pandemic on emissions, since most clusters have recorded a noticeable decline in 2020 – a reflection of reduced industrial activities and transport emissions during the year. This was while, in 2021, a relatively steady recovery across clusters could suggest going back to pre-pandemic levels of industrial and economic activity, raising questions about the plausibility of sustained and adaptive climate policy maintaining this unique period of reduction. The temporal trends in Fig. A.8 in the Appendix, showing a slight decline in all clusters, may indicate broader influences that commonly affect these regions. However, the sharp separation between the clusters, most notably the distance between C1 and the rest, speaks volumes of each cluster's standing out uniquely from the rest. This order reflects not only the relative positions of the clusters but also hints at the diverse dynamics shaping the outcomes within each group (Fig. A.8).

Following this discussion, the econometric model specified in this study is as follows, with the estimated equation being:

$$GHG_{it} = \alpha + \beta_1(PC1000)_{it} + \beta_2(OR)_{it} + \beta_3(MOPEDS)_{it} + \beta_4(OVER125)_{it} + \beta_5(PCLESS2)_{it} + \beta_6(PC10TO20)_{it} + \beta_7(TRAM)_{it} + \text{Country}_i + \text{Year}_t + \epsilon_{it}$$

where i represents the country and t represents the time period from 2013 to 2021. The coefficients β_1-β_7 measure the impact of each explanatory variable on GHG emissions, and ϵ_{it} is the error term. The analysis aims to uncover the relationships between transportation infrastructure, vehicle usage, and GHG emissions, providing insights into the effectiveness of various policy measures and technological advancements in reducing emissions. The selection of the model is grounded in a comprehensive review of the literature, which identifies significant relationships between GHG emissions and various explanatory variables (Aminzadegan et al., 2022).

The primary model utilized is a panel data model, which is advantageous for examining data across multiple countries over time, thus capturing both cross-sectional and temporal variations. The estimated models include Pooled Ordinary Least Squares (POLS), Arellano-Bond Dynamic Panel Data (DPD) models, Fixed Effects (FE) estimator, and Weighted Least Squares (WLS). These models are selected for their capability to address the complexities inherent in the data, such as heterogeneity across countries and potential endogeneity issues.

POLS provides a baseline estimation, assuming homogeneity across countries. This model estimates the relationship between GHG emissions and the explanatory variables by minimizing the sum of squared residuals. However, while POLS is straightforward and widely used, it may not sufficiently account for country-specific effects or time dynamics, which can lead to biased estimates if important factors are omitted or if there is significant variation across countries (Wooldridge, 2010).

Arellano-Bond DPD model incorporates lagged dependent variables to capture the dynamic nature of GHG emissions. This approach is particularly useful for addressing potential endogeneity and autocorrelation issues that may arise from the temporal dependence of emissions data (Arellano & Bond, 1991). The Arellano-Bond estimator is a common technique in this context, which helps control for unobserved heterogeneity and serial correlation in the error terms, leading to more reliable estimates. While panel FE models control for unobserved country-specific characteristics that may influence GHG emissions. By including country-specific intercepts, these models account for heterogeneity across countries and focus on within-country variations over time (Hsiao, 2014). This approach helps isolate the impact of the explanatory variables on GHG emissions while controlling for time-invariant factors unique to each country.

WLS addresses heteroscedasticity by assigning different weights to observations based on the variance of the error terms. This method, alongside standard panel data techniques (Baltagi, 2008), improves the efficiency of the estimates by accounting for the unequal variance in the data, leading to more reliable and robust results.

In addition to traditional econometric models, the analysis also explores the application of ML techniques to enhance predictive accuracy and model complex relationships between variables. Techniques such as Support Vector Machines (SVM), Bagging, and Artificial Neural Networks (ANN) are employed to complement the traditional econometric models (Bishop, 2006; Vapnik, 2000). SVM is valued for its ability to handle non-linear relationships and high-dimensional data, making it suitable for complex environmental datasets. Bagging, a form of ensemble learning, combines multiple models to reduce variance and improve prediction accuracy, thus enhancing the robustness of the results (Breiman, 1996). ANN, particularly those with multiple layers, are utilized to capture non-linear interactions and provides a flexible framework for modeling complex data patterns, which is crucial for understanding the multifaceted nature of GHG emissions.

The integration of traditional econometric models with ML techniques in this research is driven by the complementary strengths of these methodologies in addressing the complexities of GHG emissions. Econometric models, such as FE, Arellano-Bond DPD, and WLS, are grounded in theoretical frameworks and provide robust causal inferences while controlling for heterogeneity and endogeneity across countries. However, these models may fall short in capturing non-linear relationships or handling high-dimensional datasets, which are characteristic of environmental and socio-economic interactions. As discussed by Chan et al. (2022) ML methods represent, instead, a great approach for the estimation of nonlinear relationships. ML techniques, like SVM, Bagging, and ANN, excel in capturing intricate, non-linear patterns and leveraging large, complex datasets for predictive accuracy. By combining these approaches, the research aims to harness the theoretical rigor of econometric models while utilizing the flexibility and predictive power of ML methods. This synergy enables a deeper exploration of the drivers of GHG emissions, accommodating both theoretical hypotheses and data-driven insights, and ultimately supports the development of more targeted and effective environmental policies.

Across ML techniques, the selection of SVM, Bagging, and ANN is guided by the fact that they reflect three distinct paradigms in machine learning—SVM for robust classification, ensemble methods like Bagging for variance reduction, and neural networks for uncovering complex interdependencies. SVM was chosen because of its foundational role in machine learning for structured, non-linear datasets; while Bagging represents a core ensemble method, included because it generalizes the learning process across multiple subsamples of data to counteract overfitting, a common challenge in cross-national panel data (Breiman, 1996). Meanwhile, the ANN framework is specifically selected for its established efficacy in capturing highly complex, non-linear relationships (Bishop, 2006).

4. Main results and discussion

Table 2 presents the results of various econometric models used to analyze the determinants of GHG emissions across European countries, focusing on transportation-related variables. The estimated models include POLS, Arellano-Bond DPD, FE, and WLS models.

Table 2
Estimates results.

	POLS	Arellano & Bond DPD	Panel FE	WLS
MOPEDS	−0.005	−0.037***	−0.005	−0.002
	(0.014)	(0.009)	(0.013)	(0.015)
PC1000	0.196	0.171	0.196	−0.062
	(0.219)	(0.303)	(0.153)	(0.208)
OR	−0.091*	−0.142*	−0.091*	−0.138**
	(0.045)	(0.059)	(0.040)	(0.050)
OVER125	−0.038	−0.024**	−0.038***	−0.048*
	(0.020)	(0.009)	(0.010)	(0.020)
PCLESS2	0.126**	−0.054	0.126*	0.144**
	(0.049)	(0.037)	(0.060)	(0.050)
PC10TO20	−0.077	0.109*	−0.077	−0.103
	(0.058)	(0.052)	(0.060)	(0.057)
TRAM	0.014	−0.005	0.014	0.026
	(0.015)	(0.008)	(0.025)	(0.018)
Constant	1.741	38.828***	1.774*	2.940**
	(1.001)	(9.140)	(0.742)	(0.949)
N	252	223	252	247
AIC	−383.891		−441.891	−387.689
BIC	−228.596		−388.949	−229.767
Sargan test		0.8057		
R^2	0.924	0.44 (Pseudo)	0.353 (Within)	0.983

The table provides a comparative view of the estimated coefficients for each variable, along with standard errors and key model fit statistics.

Notes: Robust Standard Errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The POLS model provides a baseline estimation and assumes homogeneity across countries. The results show a statistically significant positive relationship between PCLESS2 and GHG emissions, with a coefficient of 0.126 ($p < 0.05$). Additionally, the negative coefficient for OR at −0.091 ($p < 0.10$) indicates that an increase in road length is associated with lower GHG emissions, possibly due to improved traffic flow and reduced congestion. Moving to the DPD model, which incorporates lagged dependent variables to capture the dynamic nature of GHG emissions, the results reveal a significant negative coefficient for MOPEDS at −0.037 ($p < 0.01$). This suggests that increasing the number of mopeds with non-petrol engines is associated with reduced GHG emissions, supporting the argument that alternative fuel vehicles play a crucial role in lowering urban emissions. Moreover, the positive coefficient for PC10TO20 at 0.109 ($p < 0.10$) indicates that older passenger cars, which are typically less efficient and emit more pollutants, contribute to higher GHG emissions.

The FE model, which controls for unobserved heterogeneity across countries and focuses on within-country variations, also shows a significant positive coefficient for PCLESS2 at 0.126 ($p < 0.10$), corroborating the findings from the POLS model with identical coefficients, highlighting the complex relationship between newer vehicles and emissions. Furthermore, the significant negative coefficient for OVER125 at −0.038 ($p < 0.01$) implies that motorcycles with larger engine displacements are associated with lower GHG emissions, possibly due to their higher fuel efficiency compared to cars for short-distance travel. In the WLS model, which accounts for heteroscedasticity by assigning different weights to observations, the results indicate a significant negative relationship between OR and GHG emissions at −0.138 ($p < 0.05$). This finding reinforces the importance of efficient road networks in reducing emissions. The positive coefficient for PCLESS2 at 0.144 ($p < 0.05$) again emphasizes the paradoxical impact of newer vehicles on emissions due to their production and increased usage.

In the classical econometric models shown in Table 2, all variables, except for "Country" and "Year," have been log-transformed to ensure linearity and normality. Robust Standard Errors were utilized across all models, except for the DPD Model (Arellano-Bond alternative), where a two-step estimation procedure was implemented to address heteroskedasticity and overfitting concerns. To complement the main analysis and address these issues, ML models were also employed. Following normalization in classical models, all variables, log-transformed in standard models, were standardized using the "scale" function in the following ML analysis to facilitate comparison and interpretation.

In Fig. 4, we present the results of ML models applied to assess the importance of various features in predicting GHG emissions across European countries.[1] The two panels in the figure illustrate the feature importance rankings derived from SVM and Bagging algorithms, which highlight key transportation-related variables that significantly impact GHG emissions.

The SVM model in the first panel identifies PC1000 as the most critical feature influencing GHG emissions. This confirms our econometric findings, which indicate that higher car density correlates with increased emissions due to heightened vehicle use, fuel consumption, and traffic congestion. Notably, the SVM model also highlights MOPEDS as a significant feature. This finding aligns with Clairotte et al. (2020), who emphasize the environmental effects of two-stroke mopeds in urban areas. The model further emphasizes

[1] To maintain consistency with the models adopted in the classical econometric framework, we control for "Country" and "Year" in our ML regressions. However, we exclude these variables in the feature importance graphs.

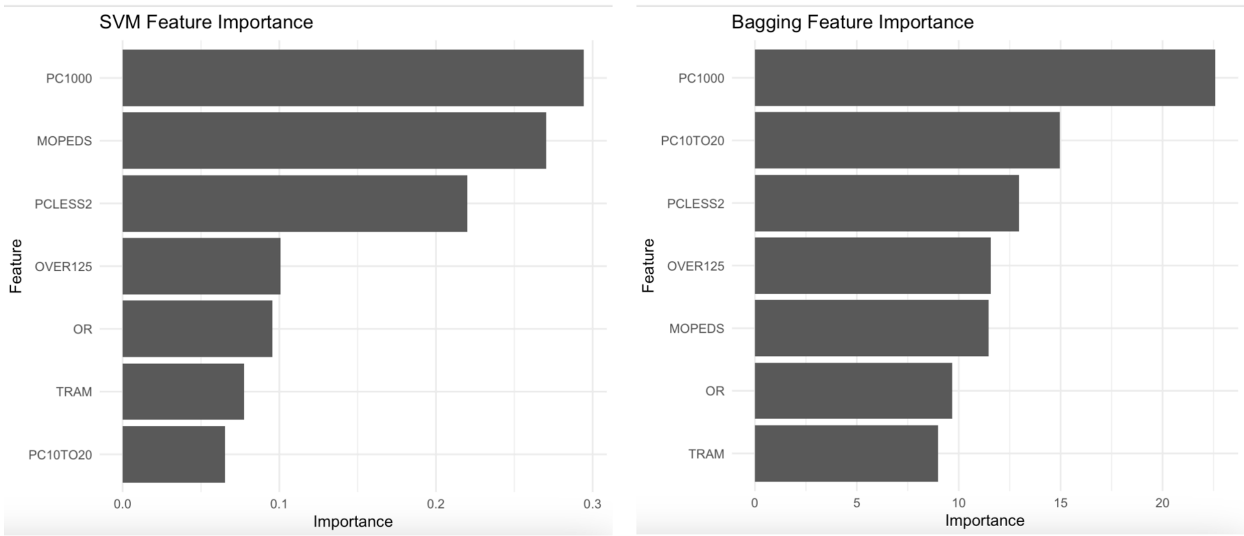

Fig. 4. SVM and Bagging models' Importance Features.

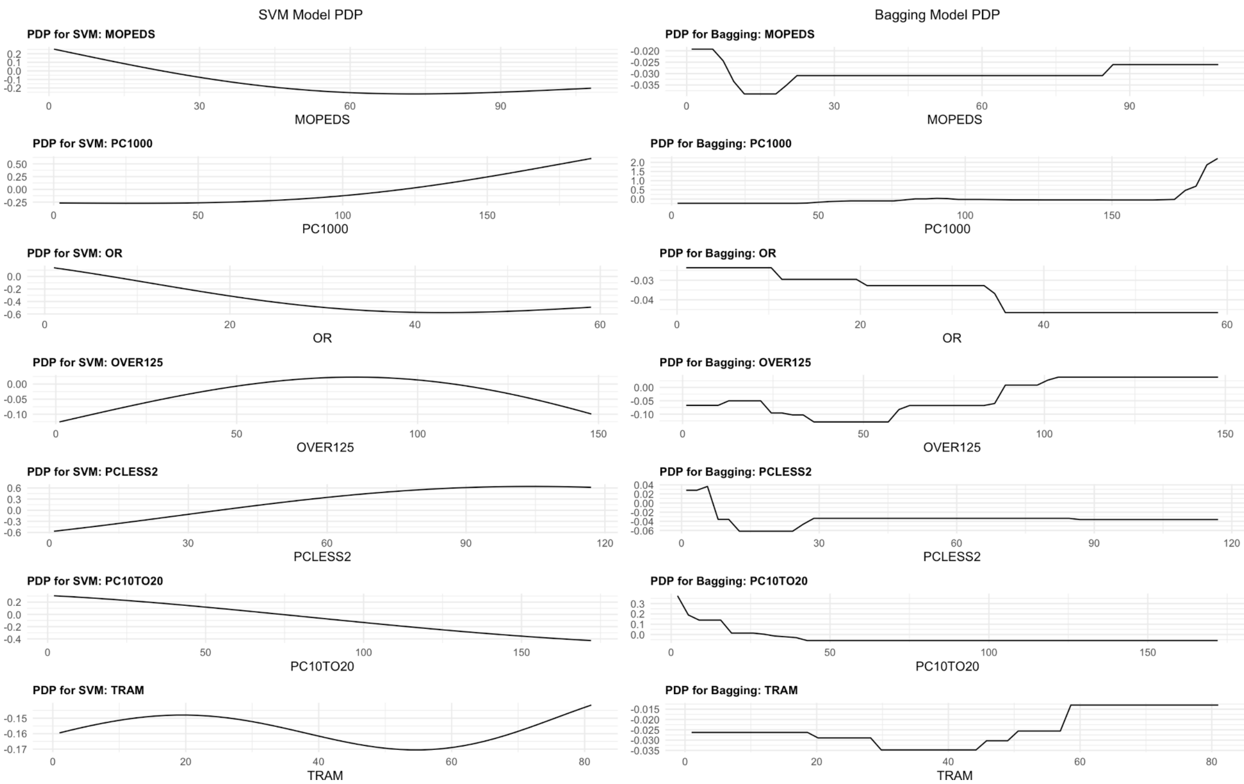

Fig. 5. Partial Dependence Plots.

the importance of PCLESS2, indicating their complex relationship with GHG emissions, influenced by both production processes and increased vehicle usage. Other notable features include OR, OVER125, and TRAM, each showing varying degrees of importance in affecting GHG emissions.

The second panel of the figure presents the feature importance of the Bagging model, which corroborates some findings from the SVM model while also providing additional insights. Here, PC1000 remains the most crucial feature, highlighting its pivotal role in emissions. The model also ranks PC10TO20 as the second most important feature, suggesting that older vehicles, with their lower fuel efficiency and higher emission rates, significantly contribute to GHG emissions. This is consistent with literature indicating that older vehicles are less likely to meet stringent environmental standards with GHG savings initiated by diesel cars clearly overestimated (Cames & Helmers, 2013). The prominence of MOPEDS and OVER125 in the Bagging model further highlights the impact of vehicle type and fuel on emissions, with MOPEDS showing a slightly weaker relationship due to their lower emissions profile. The presence of TRAM as an important feature underscores the role of public transport systems in reducing urban emissions. OR and PCLESS2 also emerge as significant, reflecting their contributions to GHG emissions through infrastructure and vehicle lifecycle effects.

Therefore, the feature importance rankings in Fig. 4 provide clear insights into the primary drivers of GHG emissions. The consistent prominence of PC1000 as the top predictor across both models underscores the critical role of car density in shaping emissions. Meanwhile, the notable importance of MOPEDS in both models highlights the potential of alternative fuel vehicles in urban emission reduction strategies.

Fig. 5 showcases the Partial Dependence Plots (PDPs) for the SVM and Bagging models, illustrating the impact of various features on GHG emissions.

The PDPs for the SVM model reveal distinct trends for each feature. For MOPEDS, the SVM model shows a negative trend, indicating that as MOPEDS' value increases, the predicted outcome decreases. This trend is relatively smooth and linear. In contrast, for the feature PC1000, there is a positive relationship with a gradual increase in GHG emissions as PC1000 values rise, suggesting that higher PC1000 values lead to higher predicted outcomes. The feature OR displays a consistently decreasing trend, implying that higher OR values negatively impact the predicted outcome. For OVER125, the relationship seems to be overall slightly positive, with a modest increase in the predicted outcome as OVER125 increases, although the change is not steep. The feature PCLESS2 seems to show a complex relationship where lower values initially have a strong impact, which lessens as the feature value increases, leading to a slight increase. The trend for PC10TO20 appears negative. Similarly, for TRAM, the relationship is generally negative, indicating that an increase in TRAM values results in a decrease in the predicted outcome.

The PDPs for the Bagging model present similar trends compared to the SVM model. For MOPEDS, the plot is almost flat with a slight decrease, indicating that the MOPEDS feature has a negligible effect on the predicted outcome in the Bagging model. Conversely, the relationship for PC1000 is significantly positive, with a steep increase in the predicted outcome as PC1000 values increase, suggesting strong sensitivity to this feature. The feature OR shows small fluctuations, but overall, the impact of OR on the predicted outcome appears minimal, with slight decreases at various points. For OVER125, the plot shows some fluctuation with a modest positive trend overall, indicating that the Bagging model has a varying but generally positive response to OVER125. The feature PCLESS2 exhibits a mostly flat trend with minor decreases, suggesting that PCLESS2 might not significantly affect the Bagging model's predictions. The plot for PC10TO20 indicates a decreasing trend, like the SVM model, but with more pronounced steps, showing the model's sensitivity to changes in PC10TO20 values. The relationship for TRAM is flat with minor negative steps, similar to the SVM model but less pronounced.

Hence, the PDPs in Fig. 5 reveal complex, non-linear relationships between transportation variables and GHG emissions, providing actionable insights for policy design. For instance, the gradual increase in emissions with PC1000 suggests the necessity of reducing car dependency through expanded public transport systems. The plots also highlight the nuanced impact of variables like PCLESS2 and PC10TO20, suggesting the need for interventions targeting vehicle lifecycle emissions and the retirement of older, less efficient fleets.

Fig. A.9 in the Appendix presents the performance comparison of the Bagging and SVM models in predicting GHG emissions, assessed using three metrics: Mean Squared Error (MSE), R^2, and Root Mean Squared Error (RMSE). The Bagging model exhibits lower MSE and RMSE values, suggesting it is better at minimizing errors and providing more accurate predictions. Additionally, the Bagging model has a relatively high R^2, indicating it effectively explains the variance in the data. Conversely, the SVM model, despite its slightly higher MSE and RMSE values, still demonstrates strong predictive power with a high R^2, reflecting its capability to capture the complex relationships within the data. This comparison highlights the strengths of Bagging in error reduction and the robustness of SVM in handling intricate data patterns.

The combined insights from both classical econometric and ML models underscore the complex nature of transportation's impact on GHG emissions. The detailed analyses reveal the significant influence of variables like passenger car density, newer vehicles, and public transportation systems on emission levels. These findings highlight the necessity for targeted policies that promote sustainable vehicle use and enhance public transportation infrastructure. Such policies should integrate technological advancements with infrastructure development to effectively reduce emissions. Therefore, it is crucial for technology and policy to work in tandem, driving the transition towards a more sustainable transportation system and contributing to broader environmental sustainability goals. The

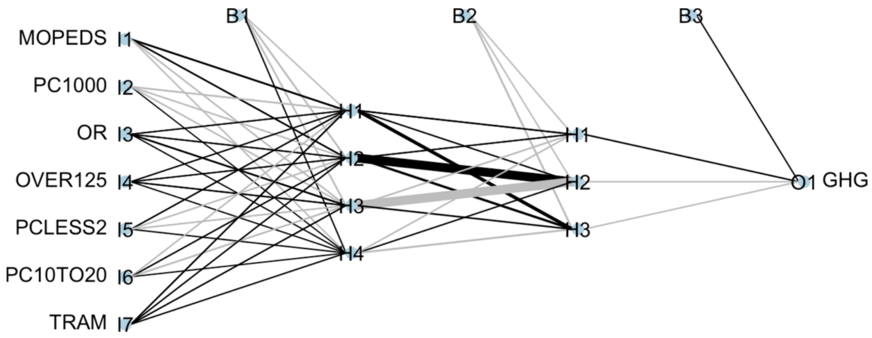

Fig. 6. ANN architecture.

evidence emphasizes that a holistic approach, involving both advanced analytical tools and comprehensive policy frameworks, is essential to achieve meaningful reductions in GHG emissions.

5. Robustness checks

In order to ensure the robustness of our findings and to validate the consistency of the results, we employed ANN for the analysis. Fig. 6 presents our robustness checks by the application of ANN with deep learning to model intricate relationships between various transportation-related variables and GHG emissions.[2] The structure of the network in the ANN model is presented in this figure, showing the complex interconnections between input variables and their impact on GHG emissions.

Fig. 6 shows a neural network structure that illustrates the intricate relationship between transportation variables and GHG emissions. The flow of influence moves from the input nodes (I1 to I7) to the hidden nodes (H1 to H4 in the first layer and H1 to H3 in the second layer), passing through the bias nodes (B1 to B3) and ultimately reaching the output node, O1. The varying strengths of influence that each variable has on the output are captured by the thickness of the connections between the nodes.

MOPEDS (I1) exhibits moderate connections to the hidden nodes, emphasizing its role in mitigating GHG emissions, particularly through the adoption of alternative fuel technologies. PC1000 (I2) maintains some few but strong connections to hidden nodes; while OR (I3) displays balanced connections, indicating its potential to indirectly influence emissions by improving traffic flow and reducing congestion. OVER125 (I4) demonstrates significant connections, highlighting the role of high-efficiency vehicles in reducing emissions compared to conventional vehicles for short-distance travel. Moreover, PCLESS2 (I5) and PC10TO20 (I6) continue to exhibit a complex pattern of influence, likely reflecting the lifecycle emissions associated with newer vehicle manufacturing and usage, as well as, the challenges posed by outdated vehicle technologies. Finally, TRAM (I7) presents balanced strong connections, reinforcing the importance of electrified public transportation systems in reducing emissions.

The hidden nodes (H1–H3) act as integrators, synthesizing the effects of the input variables. Bias nodes (B1–B3) contribute to adjusting the relationships between variables, ensuring that non-linear effects are adequately captured. The output node (O1), representing GHG emissions, is strongly influenced by H1, reflecting the network's ability to model the non-linear and interdependent nature of the variables.

Fig. A.10 in the Appendix presents the evaluation metrics for the ANN model used. The model's performance is assessed using Mean Absolute Error (MAE), MSE, R2, and RMSE. The high R^2 value demonstrates that the ANN model effectively captures a substantial proportion of the variance in GHG emissions, indicating a robust explanatory power. The relatively low values of MAE reflect the model's precision in predicting emissions. This confirms the ANN's capability to accurately model the complex relationships between various transportation-related variables and GHG emissions.

Overall, the ANN model with deep learning confirms the robustness of our findings and highlights the importance of employing advanced analytical techniques to fully understand the drivers of GHG emissions in the transportation sector. The model's ability to reveal deep, underlying patterns in the data supports the argument for targeted, multifaceted policy interventions to reduce emissions and promote sustainable transportation practices.

6. Conclusions and policy implications

This study provides a comprehensive evaluation of the factors influencing GHG emissions in European countries, particularly focusing on transportation-related variables. Our empirical analysis employed various classic econometric models and ML techniques to uncover the multifaceted dynamics between transportation infrastructure, vehicle use, and GHG emissions. The research draws from an extensive body of literature that examines the environmental impacts of transportation and underscores the complex interactions between various factors and GHG emissions.

[2] In this specific case, we exclude "Year" and "Country" as controls from the regression, as ANN models are able to perform effectively even without these specifications.

The study reveals a significant reduction in GHG emissions across Europe from 2013 to 2021, although regional disparities persist, with southern regions lagging behind their northern counterparts in emission reductions. The analysis confirms a direct correlation between PC1000 and increased emissions, highlighting that higher car density contributes to greater fuel consumption and traffic congestion, which in turn leads to higher GHG emissions. On the other hand, the study identifies a reduction in emissions associated with greater use of MOPEDS and motorcycles with engine displacements OVER125. The use of TRAM also emerged as a significant factor in reducing urban emissions, which is in line with the work of Nanaki et al. (2017) on the positive impact of public transportation systems on emission reductions.

The econometric models reveal that OR is negatively correlated with GHG emissions, suggesting that extensive road networks outside urban areas may facilitate efficient traffic flow and reduce congestion, as posited by Demirel et al. (2008). This highlights the importance of infrastructure improvements in mitigating emissions. The findings from the ML models further support the traditional econometric results, with PC1000, MOPEDS, and PCLESS2 being the most significant features influencing GHG emissions. The ML models emphasize the need for targeted policies that promote sustainable vehicle use and efficient public transportation systems.

The policy implications of this study are grounded in the empirical results of both econometric and machine learning models, which highlight key transportation-related determinants of GHG emissions. For instance, the Arellano DPD model highlights the role of alternative fuel vehicles, particularly mopeds, in reducing urban emissions, suggesting the need for policy incentives to accelerate their adoption. The significant association between road infrastructure and emissions, as demonstrated in the WLS model, underscores the necessity of prioritizing investments in efficient road networks to improve traffic flow and minimize congestion, which can significantly enhance fuel efficiency. Additionally, the findings from machine learning models, particularly the importance rankings in SVM and Bagging, reveal the critical role of passenger car density in emissions. This underscores the need to reduce car dependency through the expansion of modern public transport systems, such as trams, and the development of car-sharing services and electric vehicle charging networks (Jung & Koo, 2018).

Economic policies should provide tax incentives for the adoption of clean technologies and subsidies for public transport development. Additionally, stricter emission standards and support for research and development in clean technology and alternative transportation choices are necessary to drive significant reductions in GHG emissions as well as to mitigate car ownership (Sun et al., 2024). Public awareness campaigns and educational programs are vital to fostering sustainable mobility behaviors among the population. International cooperation and knowledge sharing can help standardize policies and disseminate successful strategies, leveraging ML insights for more targeted and effective emission reduction strategies. Creating a supportive economic environment and facilitating the exchange of best practices between regions will advance efforts to reduce GHG emissions and promote sustainable mobility across Europe.

Despite the comprehensive approach of this study, several limitations should be noted. The analysis does not fully explore the underlying causes of regional disparities in GHG emission reductions, which hinder the development of effective, tailored interventions. While the correlation between car fleet size and emissions is evident, causation is not established, and the study does not consider emerging vehicle technologies such as hydrogen fuel cells and biofuels, which are critical for future emission reductions (Keeley et al., 2024; Akinpelu et al., 2023; Velandia Vargas et al., 2022). The suggested interventions lack detailed feasibility and cost analyses, making it difficult to assess their practical implementation through tax rebates and other incentives (Brown et al., 2020). While promising, the ML models require further validation to ensure robustness and address potential overfitting. Additionally, the study's temporal scope from 2013 to 2021 may not capture long-term trends and recent developments, limiting its comprehensiveness.

Future research should delve deeper into the socio-economic, political, and infrastructural factors underlying regional disparities in GHG emission reductions. Establishing causation between car fleet characteristics and emissions is crucial, and the analysis should be expanded to include emerging technologies like hydrogen fuel cells and biofuels. Comprehensive feasibility and cost-benefit analyses of infrastructure improvements, focusing the attention to metropolitan-scale planning and deploying programs that enhance regional accessibility, are necessary (Tao & Cao, 2023). Enhancing and validating ML models to predict emission trends with greater accuracy is essential, along with analyzing the economic impacts and political feasibility of proposed policies. Research should also address the challenges of international cooperation and extend the temporal scope beyond 2021 to capture long-term trends.

CRediT authorship contribution statement

Cosimo Magazzino: Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition, Formal analysis, Conceptualization. **Alberto Costantiello:** Writing – original draft, Project administration, Conceptualization. **Lucio Laureti:** Writing – original draft, Project administration. **Angelo Leogrande:** Writing – original draft, Software, Investigation, Data curation. **Tulia Gattone:** Writing – review & editing, Visualization, Validation, Formal analysis.

Declaration of Competing Interest

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

Appendix

.

Fig. A1. Relationship between GHG and new passenger cars in Europe.

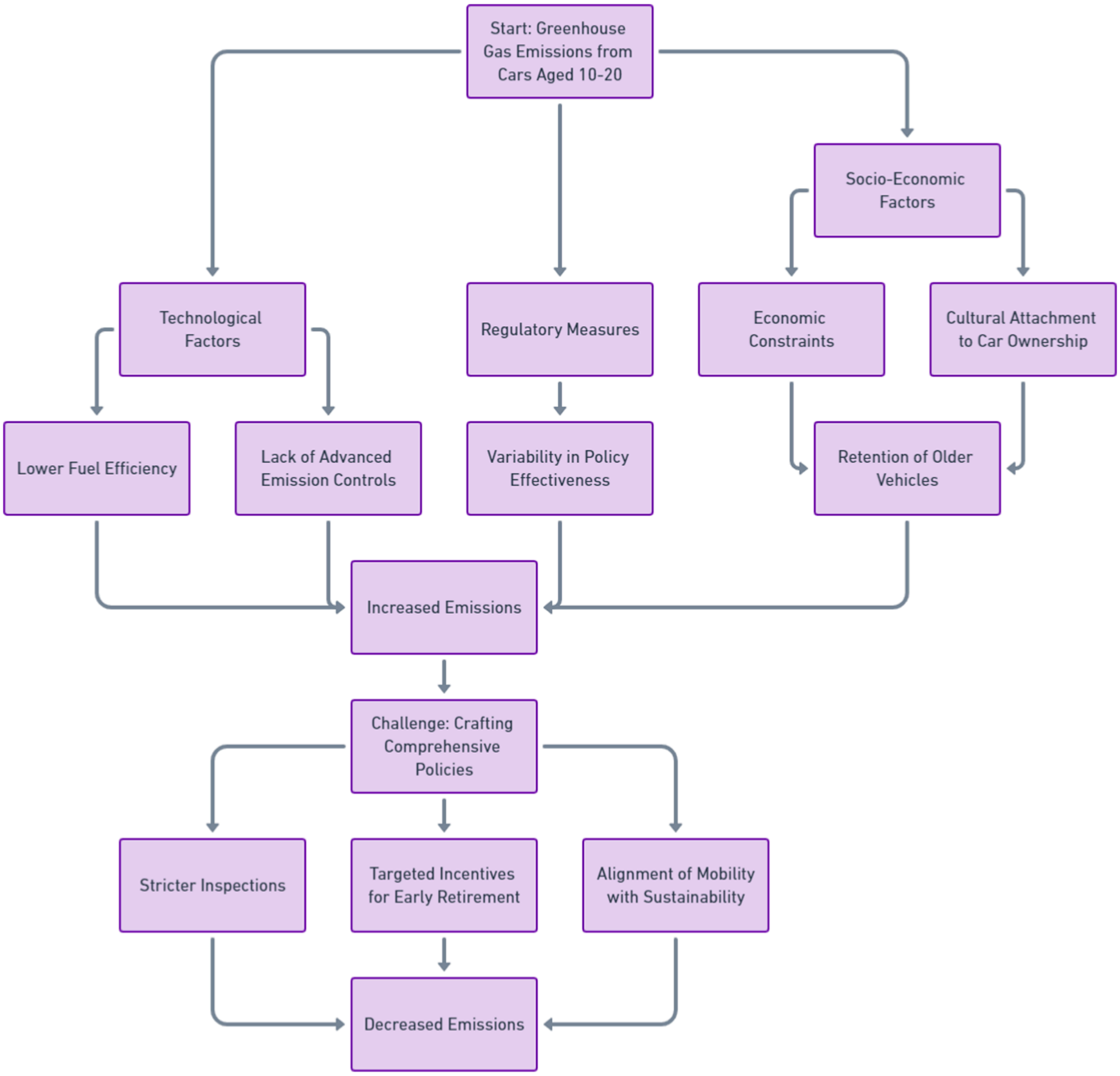

Fig. A2. Relationship between GHG and vehicles aged between 10 and 20 years.

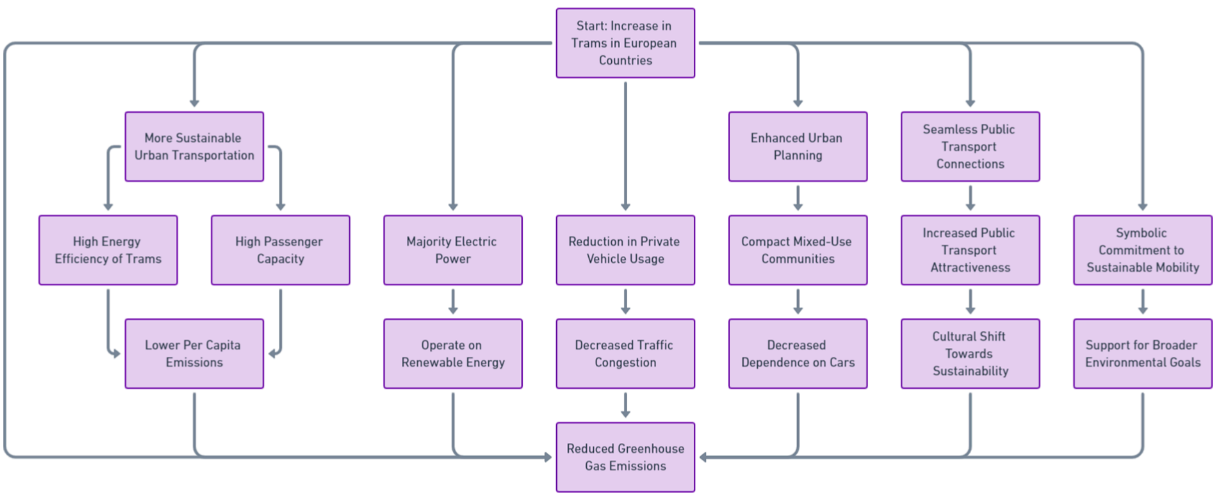

Fig. A3. Relationship between GHG and tram in European countries.

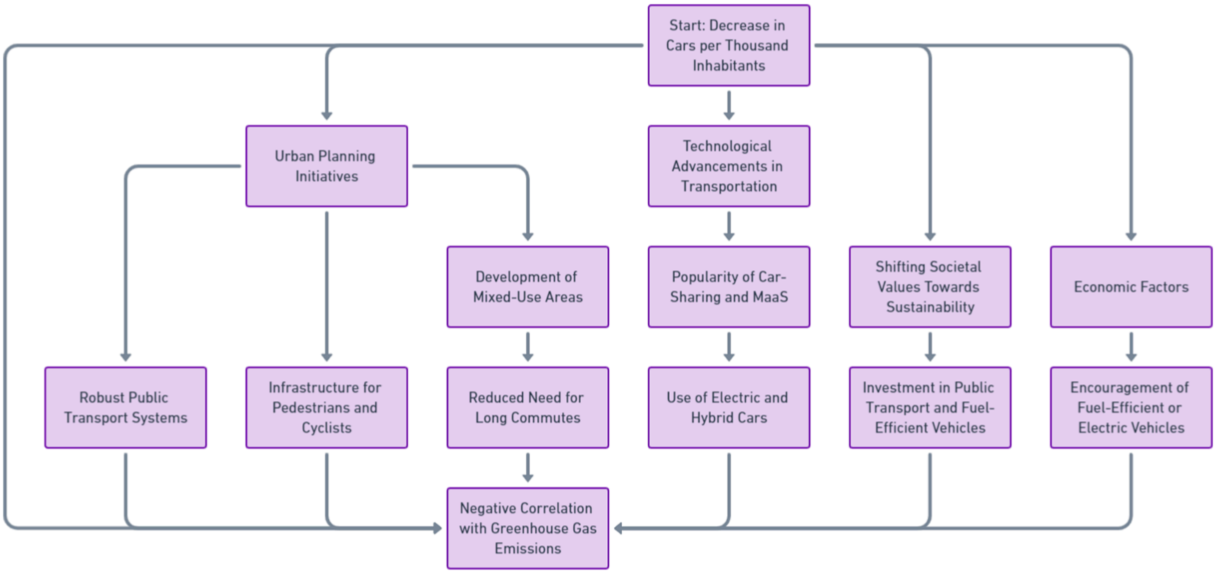

Fig. A4. Relationship between the GHG value and passenger cars per thousand inhabitants.

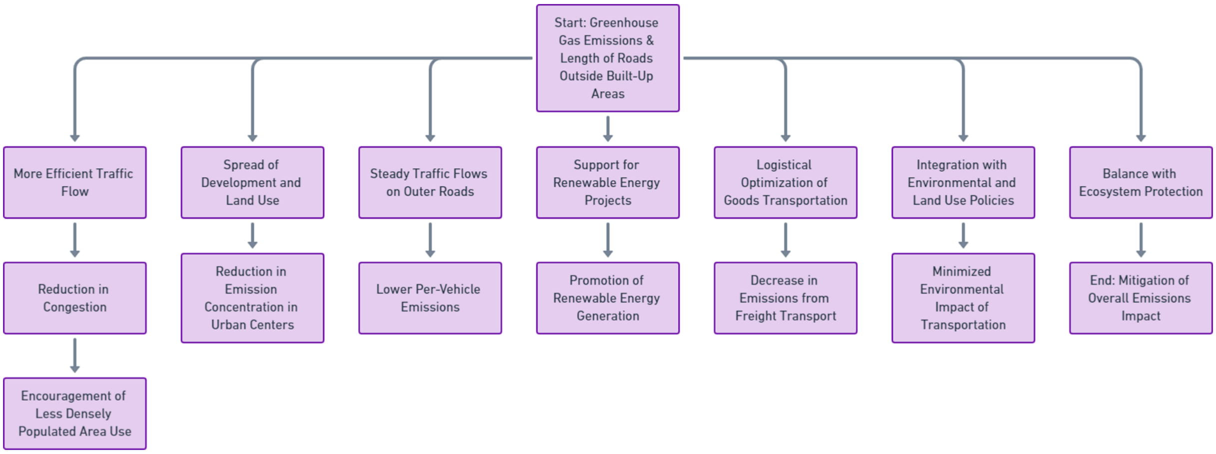

Fig. A5. The relationship between GHG and length of roads outside built up areas.

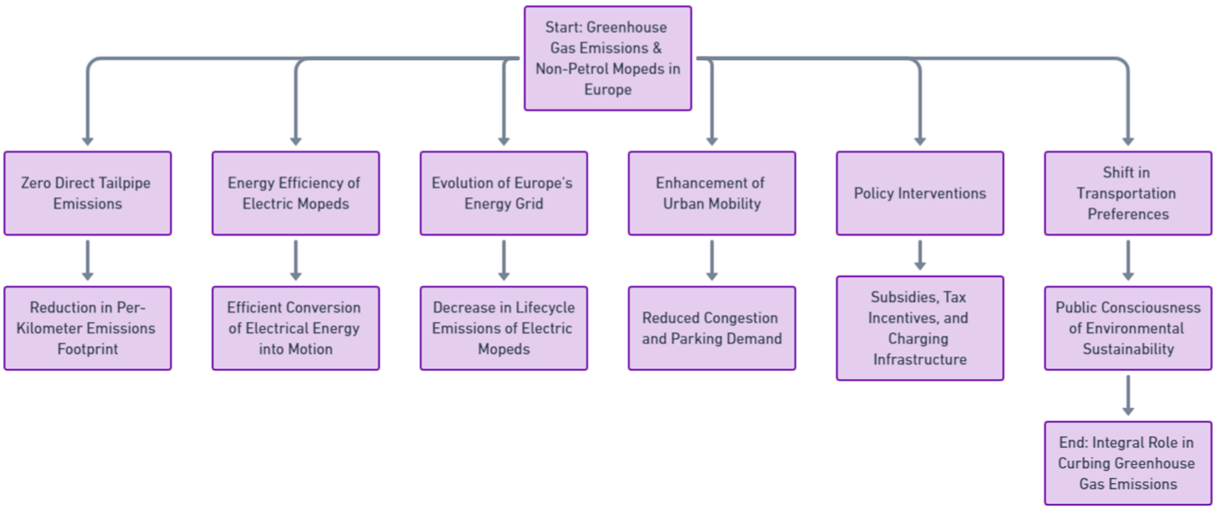

Fig. A6. The relationship between GHG and non-petrol Mopeds in Europe.

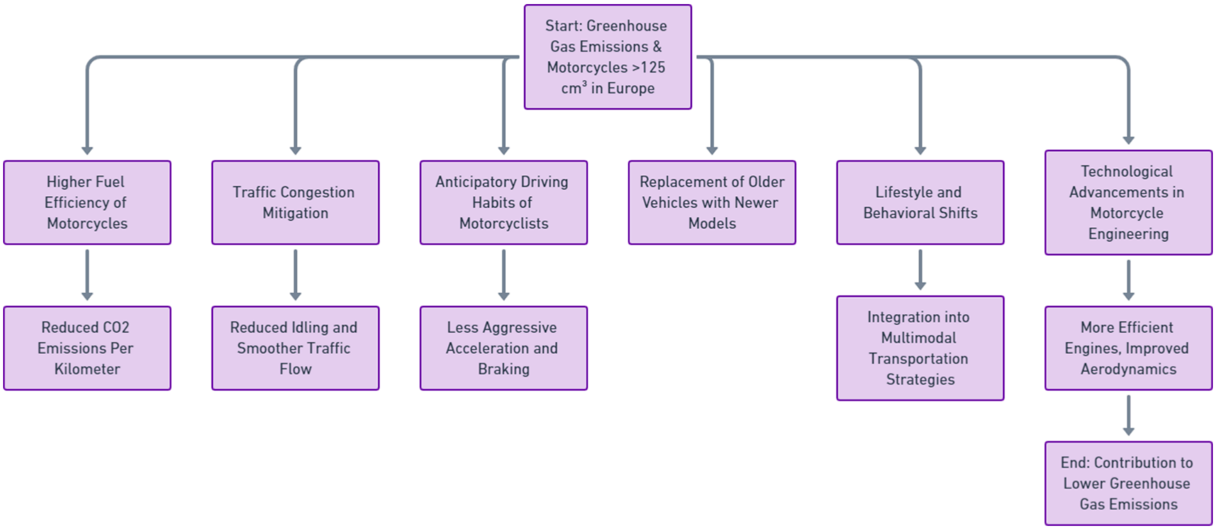

Fig. A7. The relationship between greenhouse gas emissions and motorcycles over 125 cm^3 in Europe.

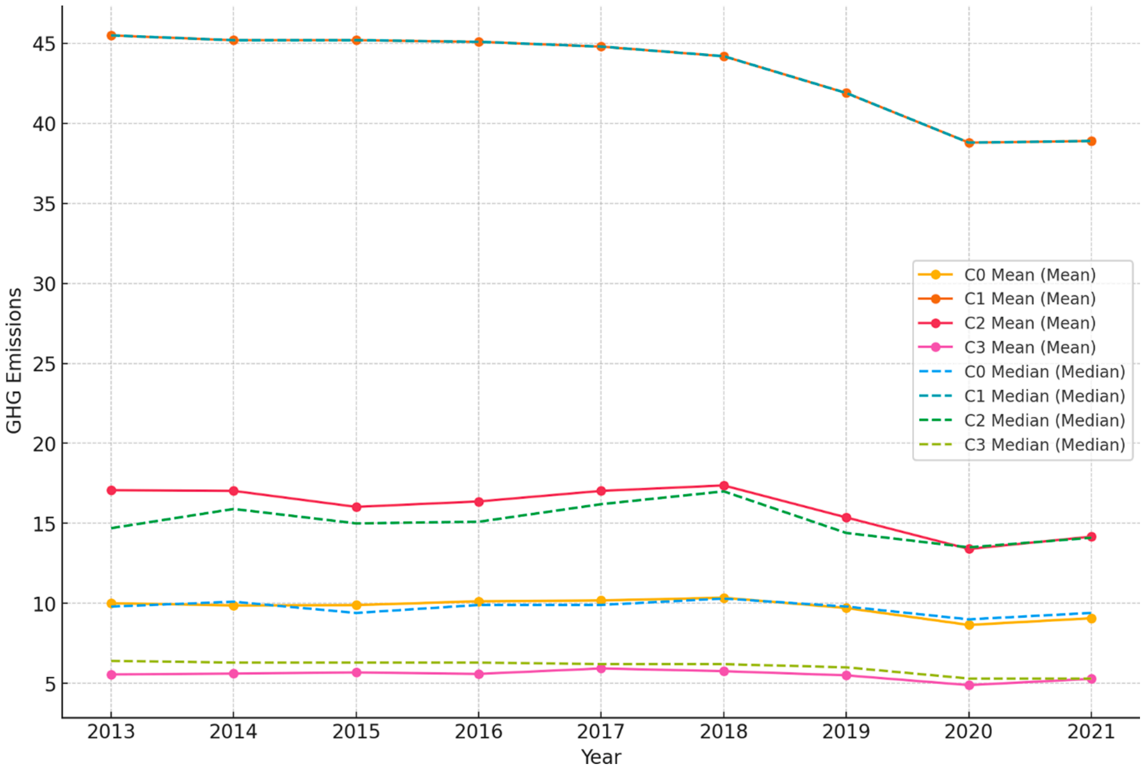

Fig. A8. Mean and median of clusters.

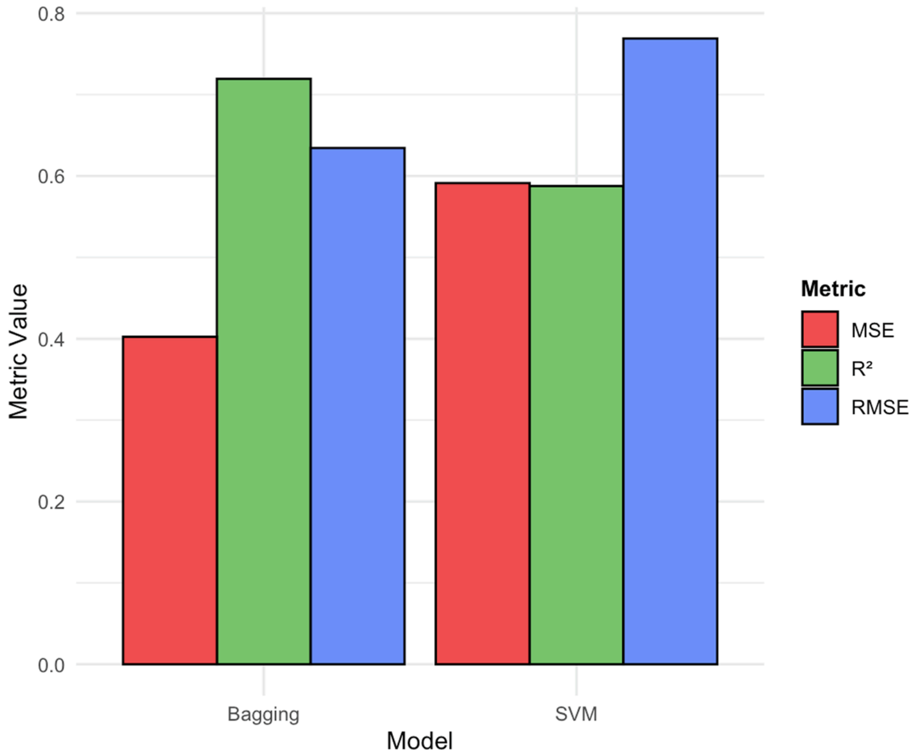

Fig. A9. Models' performance comparison.

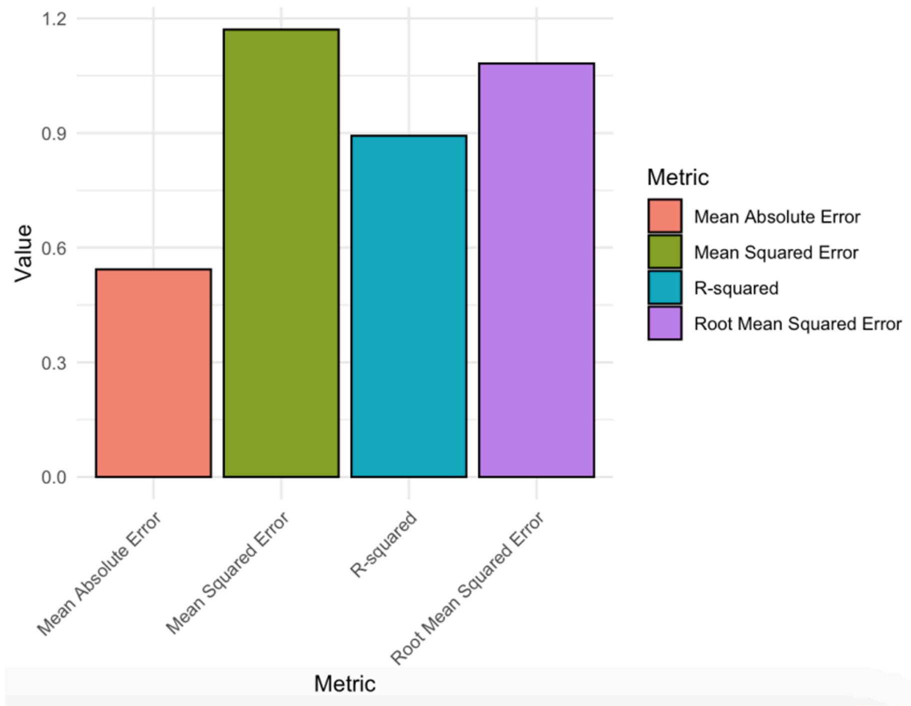

Fig. A10. ANN performance metrics.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.trd.2025.104602.

Data availability

Data will be made available on request.

References

Adam, T., Farfaletti, A., Montero, L., Martini, G., Manfredi, U., Larsen, B., Santi, G., Krasenbrink, A., Astorga, C., 2010. Chemical characterization of emissions from modern two-stroke mopeds complying with legislative regulation in Europe (EURO-2). *Environ. Sci. Tech.* 44 (1), 505–512.

Akinpelu, A., Alam, M.S., Shafiullah, M., Rahman, S.M., Al-Ismail, F.S., 2023. Greenhouse gas emission dynamics of Saudi Arabia: Potential of hydrogen fuel for emission footprint reduction. *Sustainability* 15 (7), 5639.

Albuquerque, F.D., Maraqa, M.A., Chowdhury, R., Mauga, T., Alzard, M., 2020. Greenhouse gas emissions associated with road transport projects: Current status, benchmarking, and assessment tools. *Transp. Res. Procedia* 48, 2018–2030.

Al-Foraih, R., Sreekanth, K.J., 2021. Effects and critical analysis for emission reduction assessment in the road transportation sector. *International Journal of Global Warming* 25 (3–4), 271–287.

Alhindawi, R., Abu Nahleh, Y., Kumar, A., Shiwakoti, N., 2020. Projection of greenhouse gas emissions for the road transport sector based on multivariate regression and the double exponential smoothing model. *Sustainability* 12 (21), 9152.

Al-Osaimi, S., Sreekanth, K.J., Al-Foraih, R., Al-Kandari, S., 2020. Trends in road transportation fuel consumption and carbon emissions: A scenario analysis using system dynamic modeling. *Int. J. Sustain. Energ.* 39 (4), 349–361.

Aminzadegan, S., Shahriari, M., Mehranfar, F., Abramović, B., 2022. Factors affecting the emission of pollutants in different types of transportation: A literature review. *Energy Rep.* 8, 2508–2529.

Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. Econ. Stud.* 58 (2), 277–297.

Atabay, O., Djilali, N., & Crawford, C. (2022). Shared automated electric vehicle prospects for low carbon road transportation in British Columbia, Canada.*Vehicles, 4* (1), 102-123.

Axsen, J., Plötz, P., Wolinetz, M., 2020. Crafting strong, integrated policy mixes for deep CO_2 mitigation in road transport. *Nat. Clim. Chang.* 10 (9), 809–818.

Baltagi, B.H., 2008. Econometric Analysis of Panel Data, (4th ed.). Wiley.

Barrington-Leigh, C., Millard-Ball, A., 2017. More connected urban roads reduce US GHG emissions. *Environ. Res. Lett.* 12 (4).

Baur, A., Lauf, S., Förster, M., Kleinschmit, B., 2015. Estimating greenhouse gas emissions of European cities: Modeling emissions with only one spatial and one socioeconomic variable. *Sci. Total Environ.* 520, 49–58.

Bishop, C.M., 2006. *Pattern recognition and machine learning.* Springer.

Breiman, L., 1996. Bagging predictors. *Mach. Learn.* 24 (2), 123–140.

Brown, A.L., Fleming, K.L., Safford, H.R., 2020. Prospects for a highly electric road transportation sector in the USA. *Current Sustainable/renewable Energy Reports* 7, 84–93.

Camargo-Caicedo, Y., Mantilla-Romo, L.C., Bolaño-Ortiz, T.R., 2021. Emissions reduction of greenhouse gases, ozone precursors, aerosols and acidifying gases from road transportation during the COVID-19 lockdown in Colombia. *Appl. Sci.* 11 (4), 1458.

Cames, M., Helmers, E., 2013. Critical evaluation of the European diesel car boom - global comparison, environmental effects and various national strategies. *Environ. Sci. Eur.* 25 (15).

Casals, L., Martinez-Laserna, E., Garcia, B., Nieto, N., 2016. Sustainability analysis of the electric vehicle use in Europe for CO_2 emissions reduction. *J. Clean. Prod.* 127, 425–437.

Chan, F., Harris, M.N., Singh, R.B., Yeo, W.E., 2022. Nonlinear econometric models with machine learning. In: Chan, F., Mátyás, L. (Eds.), *Econometrics with Machine Learning*, 53. Springer, Advanced Studies in Theoretical and Applied Econometrics.

Clairotte, M., Suarez-Bertoa, R., Zardini, A.A., Giechaskiel, B., Pavlovic, J., Valverde, V., Astorga, C., 2020. Exhaust emission factors of greenhouse gases (GHGs) from European road vehicles. *Environ. Sci. Eur.* 32, 125.

Costagliola, M., Murena, F., Prati, M., 2014. Exhaust emissions of volatile organic compounds of powered two-wheelers: Effect of cold start and vehicle speed. Contribution to greenhouse effect and tropospheric ozone formation. *Sci. Total Environ.* 468–469, 1043–1049.

Creutzig, F., Mühlhoff, R., Römer, J., 2012. Decarbonizing urban transport in European cities: Four cases show possibly high co-benefits. *Environ. Res. Lett.* 7 (4), 044042.

Demirel, H., Sertel, E., Kaya, S., Seker, D., 2008. Exploring impacts of road transportation on environment: A spatial approach. *Desalination* 226 (1–3), 279–288.

Din, A.U., Ming, J., Rahman, I.U., Han, H., Yoo, S., Alhrahsheh, R.R., 2023. Green road transportation management and environmental sustainability: The impact of population density. *Heliyon* 9 (9), e0288493.

Domagała, J., Kadłubek, M., 2022. Economic, energy and environmental efficiency of road freight transportation sector in the EU. *Energies* 16 (1), 461.

Gan, Y., Wang, M., Lu, Z., Kelly, J., 2021. Taking into account greenhouse gas emissions of electric vehicles for transportation decarbonization. *Energy Policy* 155, 112353.

Gialos, A., Zeimpekis, V., Madas, M., Papageorgiou, K., 2022. Calculation and assessment of CO_2 emissions in road freight transportation: A Greek case study. *Sustainability* 14 (17), 10724.

Giechaskiel, B., 2020. Gaseous and particulate emissions of a Euro 4 motorcycle and effect of driving style and open or closed sampling configuration. *Sustainability* 12 (21), 9122.

Gogeri, I., Gouda, K.C., 2024. Assessment of country-specific GHG emissions from road transport sector in India. *Mapan* 39, 585–600.

Habib, Y., Xia, E., Hashmi, S.H., Ahmed, Z., 2021. The nexus between road transport intensity and road-related CO_2 emissions in G20 countries: An advanced panel estimation. *Environ. Sci. Pollut. Res.* 28 (41), 58405–58425.

Hadi, S.P., Al Hasibi, R.A., 2022. An integrated renewable energy system for the supply of electricity and hydrogen energy for road transportation which minimizes greenhouse gas emissions. *International Journal of Sustainable Energy Planning and Management* 35, 65–82.

Handayani, F., Pramesti, F., Wibowo, M., Setyawan, A., 2019. Estimating and reducing the release of greenhouse gases in local road pavement constructions. *International Journal on Advanced Science, Engineering and Information Technology* 9 (5), 1709–1715.

Hankey, S., Marshall, J., 2010. Impacts of urban form on future US passenger-vehicle greenhouse gas emissions. *Energy Policy* 38, 4880–4887.

He, X., Shen, W., Wallington, T.J., Zhang, S., Wu, X., Bao, Z., Wu, Y., 2021. Asia Pacific road transportation emissions, 1900–2050. *Faraday Discuss.* 226, 53–73.

Ho, B.Q., Vu, K., Nguyen, T., Le, Q., Rakholia, R., Carbajo, R.S., 2022. The emissions and measures in greenhouse gases management for climate change mitigation in developing countries: A case of Ho Chi Minh City, Vietnam. International Journal of Environmental Science and Development 13 (6), 257–263.

Hsiao, C. (2014). *Analysis of panel data* (3rd ed.). Cambridge University Press.

Jain, A.K., 2010. Data clustering: 50 years beyond K-means. *Pattern Recogn. Lett.* 31 (8), 651–666.

Jung, J., Koo, Y., 2018. Analyzing the effects of car sharing services on the reduction of greenhouse gas (GHG) emissions. *Sustainability* 10 (2), 539.

Kaufman, L., Rousseeuw, P.J., 1990. *Finding groups in data: An introduction to cluster analysis*. Wiley.

Keeley, A.R., Coulibaly, T.Y., Managi, S., 2024. Analysis of prospective demand for hydrogen in the road transportation sector: Evidence from 14 countries. *Int. J. Hydrogen Energy* 56, 853–863.

Kéry, I., Mullins, O., & Wigle, R. (2021). Background and policy directions for decarbonizing Canadian road transportation. LCERPA Working Paper, 2021-2.

Kousoulidou, M., Ntziachristos, L., Mellios, G., Samaras, Z., 2008. Road-transport emission projections to 2020 in European urban environments. *Atmos. Environ.* 42 (32), 7465–7475.

Lamb, W.F., Wiedmann, T., Pongratz, J., Andrew, R., Crippa, M., Olivier, J.G., Minx, J., 2021. A review of trends and drivers of greenhouse gas emissions by sector from 1990 to 2018. *Environ. Res. Lett.* 16, 073005.
Lee, H.J., Yoo, S.H., Lim, S., Huh, S.Y., 2023. External benefits of a road transportation system with vehicle-to-everything communications. *Transp. Policy* 134, 128–138.
Lejda, K., Jaworski, A., Mądziel, M., Balawender, K., Ustrzycki, A., Savostin-Kosiak, D., 2021. Assessment of petrol and natural gas vehicle carbon oxides emissions in the laboratory and on-road tests. *Energies* 14 (6), 1631.
Li, K., 2021. Footprint of export-related GHG emissions from Latin America and the Caribbean. IDB Working Paper Series, IDB-WP-01265.
Lim, H., Kim, J., Jo, H.H., 2020. Population age structure and greenhouse gas emissions from road transportation: A panel cointegration analysis of 21 OECD countries. *Int. J. Environ. Res. Public Health* 17 (21), 7734.
Lin, Y., Qin, Y., Wu, J., Xu, M., 2021. Impact of high-speed rail on road traffic and greenhouse gas emissions. *Nat. Clim. Chang.* 11 (11), 952–957.
Luo, Z., Wan, G., Wang, C., Zhang, X., 2017. Urban pollution and road infrastructure: A case study of China. *China Economic Review* 49 (C), 171–183.
Martin, E., Shaheen, S., 2011. Greenhouse gas emission impacts of carsharing in North America. *IEEE Trans. Intell. Transp. Syst.* 12 (4), 1074–1086.
Matsuhashi, K., Ariga, T., 2016. Estimation of passenger car CO_2 emissions with urban population density scenarios for low carbon transportation in Japan. *IATSS Res.* 39, 117–120.
Mebarki, B., Draoui, B., Allaoua, B., Draoui, A., 2020. Road transportation industry facing the energy and climate challenges. IntechOpen, In Renewable Energy-Resources, Challenges and Applications.
Minet, L., Wang, A., Hatzopoulou, M., 2021. Health and climate incentives for the deployment of cleaner on-road vehicle technologies. *Environ. Sci. Tech.* 55 (10).
Miravete, E., Moral, M., Thurk, J., 2018. Fuel taxation, emissions policy, and competitive advantage in the diffusion of European diesel automobiles. *Rand J. Econ.* 48 (3), 504–540.
Mukherjee, D., Bari, Q.H., Chowdhury, T., 2023. Curtailment of GHG emission due to implementation of metrorail service: A case study in Dhaka city. *AIP Conf. Proc.* 2713 (1).
Mukherjee, R., Rollend, D., Christie, G., Hughes, M., 2021. Towards indirect top-down road transport emissions estimation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 1092–1101.
Nanaki, E., Koroneos, C., Roset, J., López-Jiménez, P., 2017. Environmental assessment of 9 European public bus transportation systems. *Sustain. Cities Soc.* 28, 42–52.
Ntziachristos, L., Mamakos, A., Samaras, Z., Xanthopoulos, A., Iakovou, E., 2006. Emission control options for power two wheelers in Europe. *Atmos. Environ.* 40 (24), 4547–4561.
Nusa, K., Kodak, G., 2023. Comparison of Road and Maritime Transportations in Emissions Perspective. International Journal of Environment and Geoinformatics (IJEGEO) 10 (2), 048–060.
Omranian, A.R., Dabirinejad, S., Khorsandi, B., Habibian, M., 2023. Contribution of anthropogenic pollutant sources to greenhouse gas emissions: A case study from a developing country. *Environ. Sci. Pollut. Res.* 30 (27), 70159–70169.
Palander, T., Borz, S.A., Kärhä, K., 2021. Impacts of road infrastructure on the environmental efficiency of high capacity transportation in harvesting of renewable wood energy. *Energies* 14 (2), 453.
Pietrzak, K., Pietrzak, O., 2020. Environmental effects of electromobility in a sustainable urban public transport. *Sustainability* 12 (3), 1052.
Rauf, H., Umer, M., 2023. Development of AI-Augmented optimization technique for analysis & prediction of modal mix in road transportation. *PLoS One* 18 (11).
Roukounakis, N., Valkouma, E., Giama, E., Gerasopoulos, E., 2020. The development of a carbon footprint model for the calculation of GHG emissions from highways: The case of Egnatia Odos in Greece. *Int. J. Sustain. Transp.* 14 (1), 74–83.
Santopietro, L., Scorza, F., Murgante, B., 2022. Multiple components in GHG stock of transport sector: Technical improvements for SECAP baseline emissions inventory assessment. *TEMA* 15 (1), 5–24.
Sarigiannis, D., Kontoroupis, P., Nikolaki, S., Gotti, A., Chapizanis, D., Karakitsios, S., 2017. Benefits on public health from transport-related greenhouse gas mitigation policies in Southeastern European cities. *Sci. Total Environ.* 579, 1427–1438.
Shao, Q., Zhang, W., Cao, X. (J.), & Yang, J. (2023). Built environment interventions for emission mitigation: A machine learning analysis of travel-related CO_2 in a developing city. Journal of Transport Geography, 110, 103632.
Sun, S., Guo, L., Yang, S., Cao, J., 2024. Exploring the contributions of Ebike ownership, transit access, and the built environment to car ownership in a developing city. J. Transp. Geogr. 116, 103834.
Tao, T., Cao, J., 2023. Exploring nonlinear and collective influences of regional and local built environment characteristics on travel distances by mode. J. Transp. Geogr. 109, 103599.
Tao, T., Cao, J., 2024. Ineffective built environment interventions: How to reduce driving in American suburbs? Transp. Res. A Policy Pract. 179, 103924.
Vallarta-Serrano, S.I., Galindo-Muro, A.B., Cespi, R., Bustamante-Bello, R., 2023. Analysis of GHG emission from cargo vehicles in megacities: The case of the metropolitan zone of the Valley of Mexico. *Energies* 16 (13), 4992.
Vapnik, V.N., 2000. *The nature of statistical learning theory*. Springer.
Velandia Vargas, Capaz, R.S., Souza, S.P., Cavalett, O., Seabra, J.E.A., 2022. Evaluating decarbonisation pathways in road transportation via life cycle assessment. In: *Liquid Biofuels: Bioethanol*. Springer International Publishing, pp. 333–362.
Wang, A., Tu, R., Gai, Y., Pereira, L.G., Vaughan, J., Posen, I.D., Hatzopoulou, M., 2020. Capturing uncertainty in emission estimates related to vehicle electrification and implications for metropolitan greenhouse gas emission inventories. *Appl. Energy* 265 (C), 114798.
Wooldridge, J.M., 2010. *Econometric analysis of cross section and panel data*, (2nd ed.). MIT Press.
Zervas, E., 2010. Analysis of the CO_2 emissions and of the other characteristics of the European market of new passenger cars. 1. Analysis of general data and analysis per country. *Energy Policy* 38, 5413–5425.
Zhang, C., Zhao, X., Sacchi, R., You, F., 2023. Trade-off between critical metal requirement and transportation decarbonization in automotive electrification. *Nat. Commun.* 14, 1616.