



Instantaneous CO₂ emission modelling for a Euro 6 start-stop vehicle based on portable emission measurement system data and artificial intelligence methods

Maksymilian Mądziel¹

Received: 10 February 2023 / Accepted: 7 November 2023 / Published online: 29 December 2023
© The Author(s) 2023

Abstract

One of the increasingly common methods to counteract the increased fuel consumption of vehicles is start-stop technology. This paper introduces a methodology which presents the process of measuring and creating a computational model of CO₂ emissions using artificial intelligence techniques for a vehicle equipped with start-stop technology. The method requires only measurement data of velocity, acceleration of vehicle, and gradient of road to predict the emission of CO₂. In this paper, three methods of machine learning techniques were analyzed, while the best prediction results are shown by the gradient boosting method. For the developed models, the results were validated using the coefficient of determination, the mean squared error, and based on visual evaluation of residual and instantaneous emission plots and CO₂ emission maps. The developed models present a novel methodology and can be used for microscale environmental analysis.

Keywords Vehicle emission · Exhaust measurement · Air pollution · CO₂ · Portable emission measurement system · Machine learning

Introduction

Transportation is a major sector of the economy that causes environmental pollution and climate change. Emissions from transportation, mainly road transport, contribute significantly to the amount of greenhouse gases in the atmosphere (Wang et al. 2020c; Giannakis et al. 2020). Automotive is the second largest CO₂ producing sector with a global total share of approximately 1/4 of total emissions (Rubio et al. 2020). The health effects associated with road transport-related pollution have led to stricter environmental regulations for vehicle exhaust gases (Grigoratos et al. 2019; Giechaskiel et al. 2019). As a result, in 1992, the European Union introduced emission standards for the maximum permitted amounts of harmful components of exhaust gases (Varella et al. 2019). The permissible amounts and standards

are constantly evolving, and work is currently underway to publish Euro 7 standards, which, along with the CO₂ emission standards regulations, will set a new direction for future vehicle technology (Selleri et al. 2022). The new regulations, in line with the European Green Deal, aim to make Europe a climate-neutral continent by 2050 (Montanarella and Papanagos 2021; Sikora 2021). Moreover, several Asian and Latin American countries are currently applying, to some extent, the European emission requirement (Bharj et al. 2019).

The amount of emissions from exhaust components, including CO₂, is estimated based on vehicle test procedures — WLTP (Worldwide Harmonized Light-Duty Vehicles Test Procedure), supplemented by road tests (Pavlovic et al. 2018; DiPierro et al. 2019). Many researchers study real-world emissions by performing various driving cycles on selected routes, not necessarily based on homologation procedures (Weller et al. 2019; Gao et al. 2022). The estimated emissions are for a specific vehicle model, but the challenge is to estimate emissions for all vehicle traffic, which consists of many different types of vehicles that run on different fuels and have different emissions-reducing technologies. One such technology is the start-stop system.

The start-stop technologies used by vehicle manufacturers are low-cost solutions that rely on the internal combustion

Responsible Editor: V.V.S.S. Sarma

✉ Maksymilian Mądziel
mmadziel@prz.edu.pl

¹ Faculty of Mechanical Engineering and Aeronautics,
Rzeszow University of Technology, 35-959, Rzeszow,
Poland

engine to automatically shut down when the vehicle is stationary and restart at the driver's request (Santos et al. 2020). Technology is particularly useful in city center, where the vehicle performs many stop-and-go operations and spends a lot of time in traffic congestion (Wang et al. 2020a). This reduces the fuel consumption of the idle phase of the engine and, consequently, also reduces total exhaust emissions (Zhu et al. 2022).

Proper estimation of vehicle emissions is a major challenge. Measurement of emissions from vehicular traffic is even more problematic. A method that makes it possible to estimate emissions for vehicular traffic is to use, for example, models that allow emissions to be calculated (Perugu 2019; Kumar et al. 2021). A general division of emission models includes macro- and microscales (Hulagu and Celikoglu 2021). The macroscale allows emission estimates based on fewer data, mainly the technical data of the vehicle and its average velocity (Rodriguez-Rey et al. 2021). Models that allow for more accurate estimates are microscale models that need road data such as velocity, acceleration, and road gradient as a function of time (Lejri et al. 2018). Based on such calculations, it is possible to estimate emissions for the entire route, determine emission maps, and average emission factors, which, for example, can be related to emission standards, such as the actual applicable Euro 6 standard. Accurate measurement and estimation of vehicle emissions are important in the context of conducting, for example, environmental analyses of the impact of harmful transportation on the health of pedestrians traveling along major urban arteries (Quaassdorff et al. 2016; Borge et al. 2016). Creating emission maps based on emission models provides information on where and in what quantities harmful components of exhaust gases accumulate (Sanchez et al. 2017). The modeling approach is even more important because we can make models of future road investments and analyze them even before making a decision to build a particular type of road solution.

Accurate measurement of CO₂ emissions is also very important, as this component is considered to be the largest contributor to global warming (Campisi et al. 2021). Carbon dioxide has the largest share of all greenhouse gases (excluding water vapor) in their total amount and is the main source of global warming (Yoro and Daramola 2020; Franta 2018). Total carbon dioxide emissions in 2001 reached 34 million tons. Currently, the world's anthropogenic CO₂ emissions reach about 38 billion tons per year (Anenberg et al. 2019). Unless radical policies are implemented to reduce this emission, annual CO₂ emissions are estimated to increase to 41 billion tons by 2025 (Mikhaylov et al. 2020).

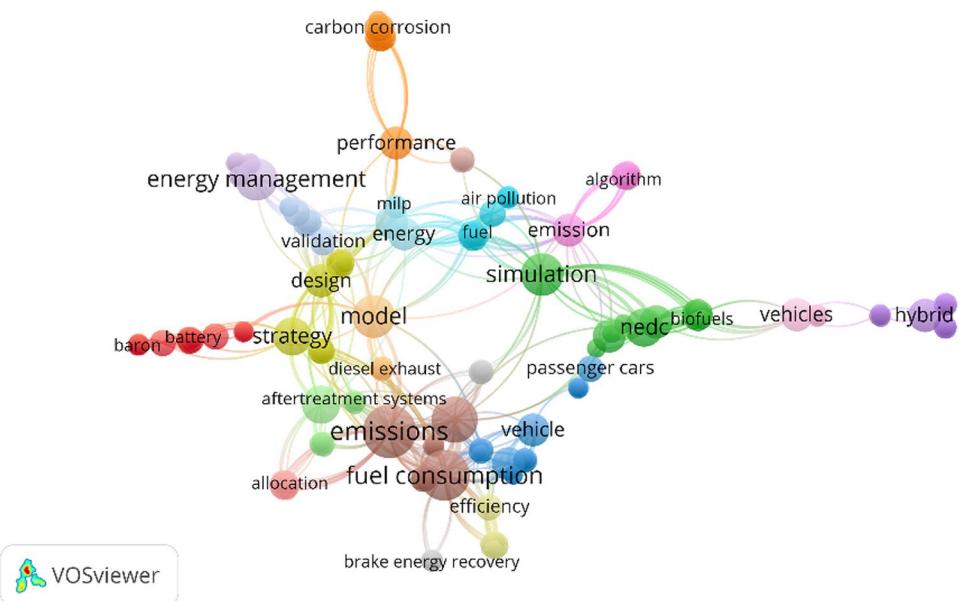
A review of articles in the field of using artificial intelligence for emission prediction shows that the base of such papers is small. Publication data from the Web of Science Core collection show that there are only 82 papers that

present the results of using AI in emissions modeling. More entries are found for the keywords emissions modeling and machine learning. One such paper is (Hoang et al. 2021) in which the authors use the artificial neural network (ANN) method to predict emissions from biodiesel-fueled diesel engines. Another example is the work (Le Cornec et al. 2020), in which the authors create emission models with a particular focus on NO_x emissions using machine learning techniques. In this work, models using neural network multilayer perceptron (MLP) techniques showed the best predictive capabilities for NO_x emission estimates. Another work using artificial intelligence techniques is (Azeez et al. 2019), where the authors created a model of CO emissions based on traffic volume data with the use of ANN. The resulting model achieved a prediction accuracy of about 80.6%. Computational models also apply to the topic of electric vehicles; an example is the work (Zhu et al. 2019), in which the authors made an accurate model of the charge load of EV using the gated recurrent unit (GRU) method. However, there is a lack of work dedicated to predicting emissions for vehicles with start-stop technology. Choosing the right machine learning method for CO₂ prediction is crucial, as not all of them will be able to restate zero emissions during most vehicle stops.

Publication data in the Web of Science Core collection show 68 results of papers for the keywords emission, model, and start-stop, while none of these papers after analysis of abstracts addresses the scope of emission modeling for start-stop technology. Therefore, this paper will be one of the first to address the scope of emissions modeling for vehicles with start-stop systems. For the above keywords, a graphic was generated using the VOSviewer software (Shah et al. 2019) showing the link of all keywords in these articles to each other (Fig. 1). From it, one can see a strong connection between the topics of emissions, fuel consumption, model, and vehicles, with simulation also present in close proximity. No start-stop keyword is found in this graphic, further confirming the lack of developed models of this type.

In view of items from the analysis of commonly available literature, the purpose of this work will be to perform a research methodology to obtain a correct model to estimate CO₂ emissions for a vehicle with start-stop technology using artificial intelligence methods. In the work, a microscale emission model is created. It is based on real data measured by the portable emission measurement system (PEMS). Furthermore, the data were measured via an OBD II (on-board diagnostic) interface. Data on speed, acceleration, vehicle coordinates, and actual CO₂ emissions were selected to create a CO₂ emissions model for start-stop technology. The most important and desirable feature of the obtained model is the CO₂ prediction capabilities for a zero value of these emissions under start-stop vehicle operations. The obtained

Fig. 1 VOSviewer network visualization based on Web of Science article review for the research topic keywords: emission, model, and start-stop



model, a modern and accurate model using artificial intelligence techniques, can be used for environmental analysis by, for example, road managers or city authorities as a component.

Methods

The general structure of the work is shown in Fig. 2. The study can be divided into three main stages. In the first part, it was important to properly prepare the vehicle and install the PEMS system. The PEMS system was properly calibrated against the reference gases, and an OBD II interface was connected to the vehicle to record vehicle parameters. In this part, it was also necessary to prepare a driving route that would take into account a wide range of velocity

and driving characteristics for different engine loads. In the second step, road tests were conducted on the selected parts of city roads to collect enough data. The focus was on measuring the following data, which were then used to construct computational models using artificial intelligence techniques, namely driving velocity, vehicle acceleration, road gradient, and CO₂ emissions. This data set, after initial filtering of the results, was then uploaded to a repository and transferred to Google Colab. Google Colab is a Jupyter environment provided and operated by Google with the ability to work with CPUs, GPUs, and even TPUs (Carneiro et al. 2018; Bisong 2019). The operation of coding the models and creating the associated visualizations was done in the Python environment. Python is a general-purpose high-level programming language with an extensive suite of standard libraries whose guiding idea is the readability

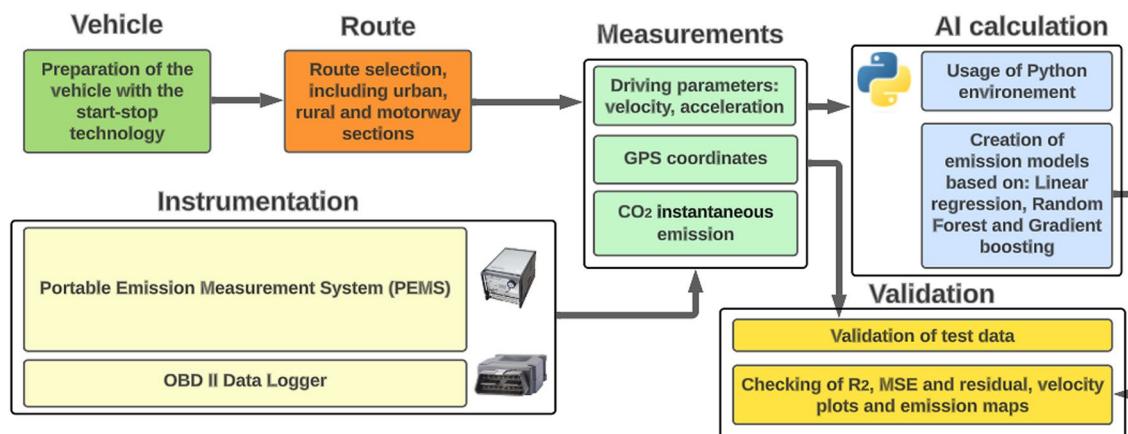


Fig. 2 Basic logic and scheme of the performed research

and clarity of its source code. Its syntax is characterized by clarity and conciseness (Raschka et al. 2020; Hao and Ho 2019). The article describes a data storage approach for model development that combines the capabilities of Google Colab and a GitHub data repository. This strategy provides several advantages, including version control to maintain a comprehensive history of dataset changes, ensuring transparency and traceability throughout the development process. It also facilitates collaborative work among researchers, offering a secure cloud-based environment for model development. Data security is upheld through GitHub's access controls, which restrict data access to authorized individuals. Critically, the methodology allows for ongoing model improvement through data updates. The approach enables automated data pipelines that fetch and preprocess data from the GitHub repository, enabling the real-time integration of new data points. This adaptability enables continual refinement and enhancement of the future model's accuracy. Linear regression, random forest, and gradient boosting methods were selected for calculating computational models. These are methods with different abilities to represent real data. The validation of the models obtained was carried out based on MSE and R^2 , as well as visual validation based on emission maps, instantaneous emission, and residual plots. The methodology developed, with special emphasis on start-stop technology, is one of the first of its kind. However, the main objective of the work was to create a CO₂ emissions model that gives the most accurate picture of emissions. Creating CO₂ models for stop-start technology is problematic because this vehicle, unlike a conventional one without this system, does not emit any harmful compounds during most stop-start operations. The use of appropriate computational techniques that can reflect the characteristics of such emissions is needed.

The tests were carried out on a Euro 6 compliant vehicle manufactured in 2018. The vehicle has a 1560-dm³ diesel engine; the maximum power of the vehicle is 88 kW at 3500 rpm. The vehicle has a manual transmission with a 6-stage gearbox. The vehicle's fuel injection system is CR (common rail), while the exhaust after-treatment system is DPF (diesel particulate filter), SCR (selective catalytic reduction), and DOC (diesel oxidation catalysts). The unloaded weight of the vehicle is 1429 kg. The vehicle is equipped with start-stop technology, which allows the engine to temporarily be turned off while at a standstill, such as while waiting for a green light, which is designed to save fuel (Lijewski et al. 2021). This system is extremely useful and has a significant impact on reducing emissions, including CO₂ under congested conditions (Hao et al. 2020; Deng et al. 2020). Before starting the vehicle road test, the vehicle was checked with the use of stationary exhaust gas analyzers, according to the periodic diagnostic test procedure. For road tests, the

route shown in Fig. 3 was selected. The route was driven twice, representing a total of 111 km, and more than 3000 data records were recorded for the model creation with a frequency of 1 s. Such values were sufficient to create an emission model for CO₂ for the case investigated. The selected parameters of the total distance with division to parts of the route measured using GPS and OBD II are presented in Table 1.

The rationale behind selecting this particular route was based on the need to include a comprehensive representation of the road conditions encountered during the road test. This comprehensive route was deliberately designed to incorporate segments reflecting urban, rural, and highway environments. It is important to note that the primary objective of the road test was not to determine emission factors or adhere strictly to the measurement procedure typically associated with real driving emissions (RDE) tests. Instead, the overarching objective was to accumulate a diverse and extensive dataset, capturing a wide spectrum of velocity and acceleration patterns. Equally vital was the effort to maintain a well-rounded cross section of road gradients throughout the route. This particular parameter, coupled with velocity and acceleration data, constituted the fundamental building blocks for the development of the CO₂ calculation model, emphasizing the multifaceted nature of the data collection approach.

The test vehicle was adapted to connect a PEMS instrumentation, Horiba OBS-2200 (Fig. 4). The PEMS system to measure CO₂ has a nondispersive infrared (NDIR) spectrometer (Jaworski et al. 2018). The measurement accuracy of the NDIR analyzer for this PEMS system is $\pm 2.5\%$ (Tipanluisa et al. 2021). A full specification of the accuracy of the measurement and the selected technical parameters of the PEMS is shown in Table 2.

The PEMS can be installed in the trunk of the vehicle under test, while the flow meter measuring sensors are connected to the exhaust pipe. The exhaust gas sampling tube must be heated to 190 °C to avoid condensation of hydrocarbons (Santos et al. 2021). In addition, ambient temperature and humidity sensors are connected to the system, as well as a GPS transmitter. Vehicle data was recorded in PEMS memory and connected to the OBD II interface, ELM327. Data that were measured included velocity (V), vehicle acceleration (a), fuel consumption, CO₂ emission, humidity, air temperature, latitude, longitude, and altitude above sea level.

Calculations of CO₂ emissions from the NDIR detector of the PEMS system were made based on the following Eq. (1):

$$\text{CO}_2\text{MASS}(t) = C_{\text{EXCO}_2}(t + DT_{\text{CO}_2}) \cdot M_{\text{CO}_2} \cdot Q_{\text{EX}}(t) \\ \cdot \frac{1}{60} \cdot \frac{1}{100} \cdot \frac{1}{22.415} \cdot \frac{273.15}{293.15} \quad (1)$$

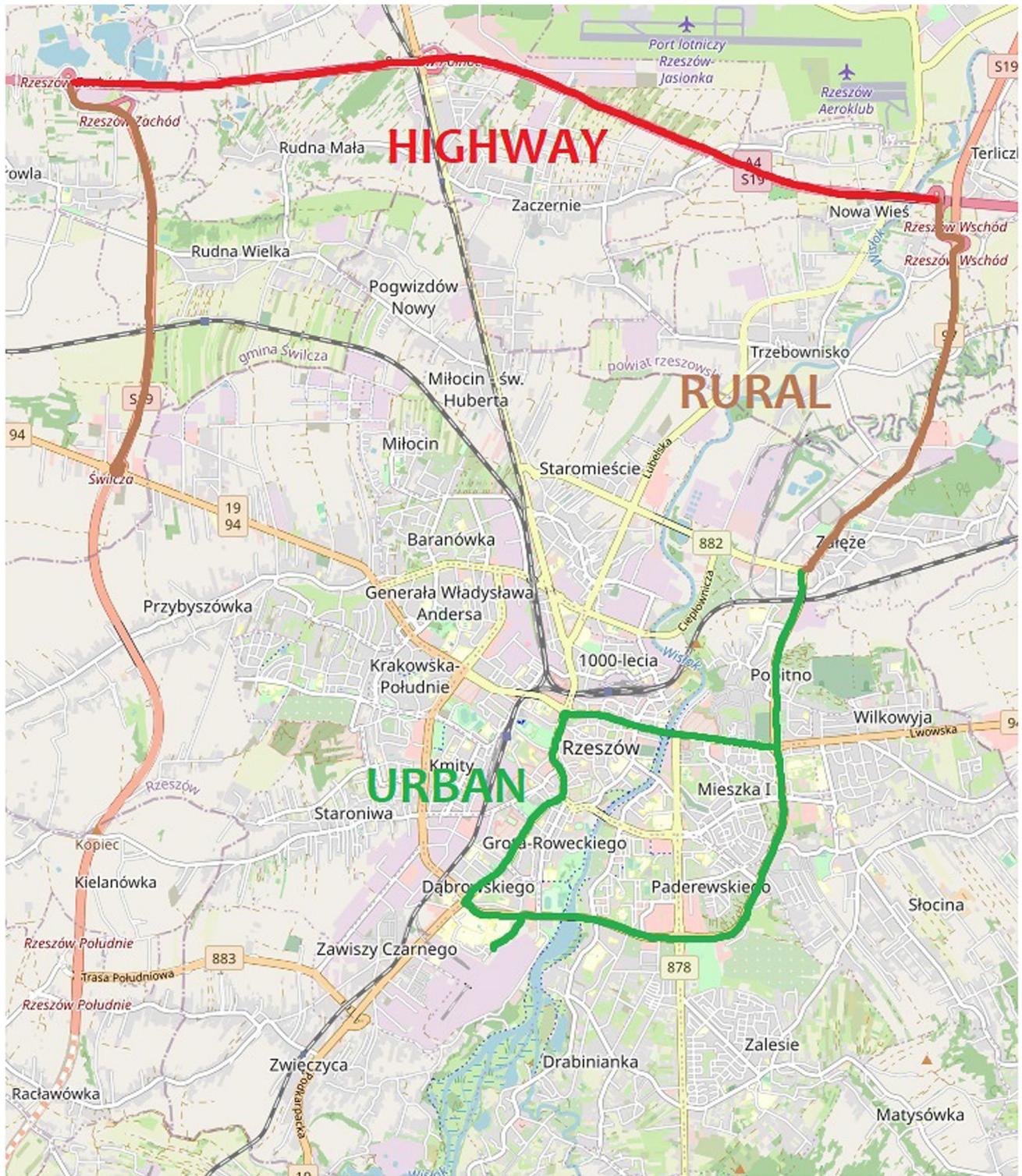


Fig. 3 View of the test road with division to the researched part: urban, rural, and highway

Table 1 Selected measurement parameters of the research route

| Parameter | Measurements |
|-----------------------------------|--------------|
| Total distance (km) | 111.4 |
| Distance of the urban part (km) | 32.24 |
| Distance of the rural part (km) | 39.66 |
| Distance of the highway part (km) | 39.5 |
| Average speed (km/h) | 44.01 |
| Urban average speed (km/h) | 29.77 |
| Rural average speed (km/h) | 90.59 |
| Highway average speed (km/h) | 110.26 |

where:

$$\begin{aligned} \text{CO}_2\text{MASS}(t) & \quad \text{real-time instantaneous CO}_2 \text{ emissions} \\ & \quad (\text{g/s}), \\ \text{DT}_{\text{CO}2} & \quad \text{analyzer measurement delays relative} \\ & \quad \text{to flow meter (s)}, \\ \text{C}_{\text{EXCO}2}(t + \text{DT}_{\text{CO}2}) & \quad \text{concentration of CO}_2 (\%), \\ \text{M}_{\text{CO}2} & \quad \text{molecular weight CO}_2 (\text{g/mol}), \\ \text{Q}_{\text{EX}}(t) & \quad \text{flue gas flow rate under standard con-} \\ & \quad \text{ditions (293.15 K, 101.3 kPa) at time} \\ & \quad t (\text{l/min}). \end{aligned}$$

The data collected were used to create computational models of CO₂ emissions. The input data were the velocity and acceleration of the vehicle, as well as the road gradient;

this set of data can be defined as explanatory variables. The dependent variable, on the other hand, was CO₂ emissions. The models were developed in Python programming language for computational techniques: linear regression, random forest, and gradient boosting. Such a range of chosen techniques will allow comparison and selection of the best one, which will best reflect CO₂ emissions for the start-stop technology of the vehicle. The MSE and R² indices were used to validate the prepared models. It was also important to perform a visual verification based on actual and modelled emissions graphs, as well as an analysis of emissions maps and residual plots for the models obtained.

Results

It was important for the development of a CO₂ emission model for start-stop technology for the vehicle tested that the model prepared could reflect the moments when the vehicle's engine is turned off, e.g., during momentary stops at junctions. Current microscale models for emission maps and for calculating instantaneous emission values did not allow such results (Eijk et al. 2018; Mądziel and Campisi 2022). Therefore, it is necessary to analyze instantaneous emission plots for CO₂, emission maps, residual graphs, and validate the models using R² and MSE indices.

The input data for the development of the emission models were the parameters: velocity, acceleration, and

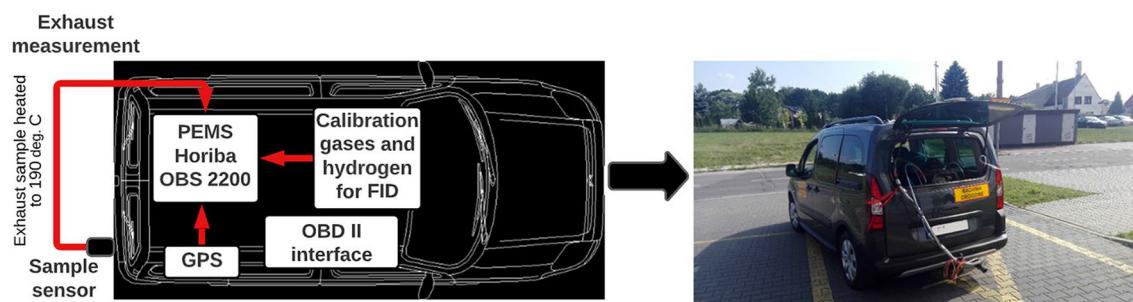


Fig. 4 PEMS platform installed in the vehicle in the study, scheme and the real photo

Table 2 Selected technical parameters of the PEMS Horiba OBS-2200 system

| Parameter | Measurement method | Accuracy |
|---|--|---|
| The concentration of exhaust gas components: CO ₂ | NDIR — nondispersive (infrared), range 0–10% 0.1–1 Hz | ± 2.5% |
| Sampling frequency | Up to 1 h | - |
| The heating time of the analyzers | Mass flow rate | In the range of the ± 1.5% of full scale or within ± 2.5% of readings |
| Gas flow | | |

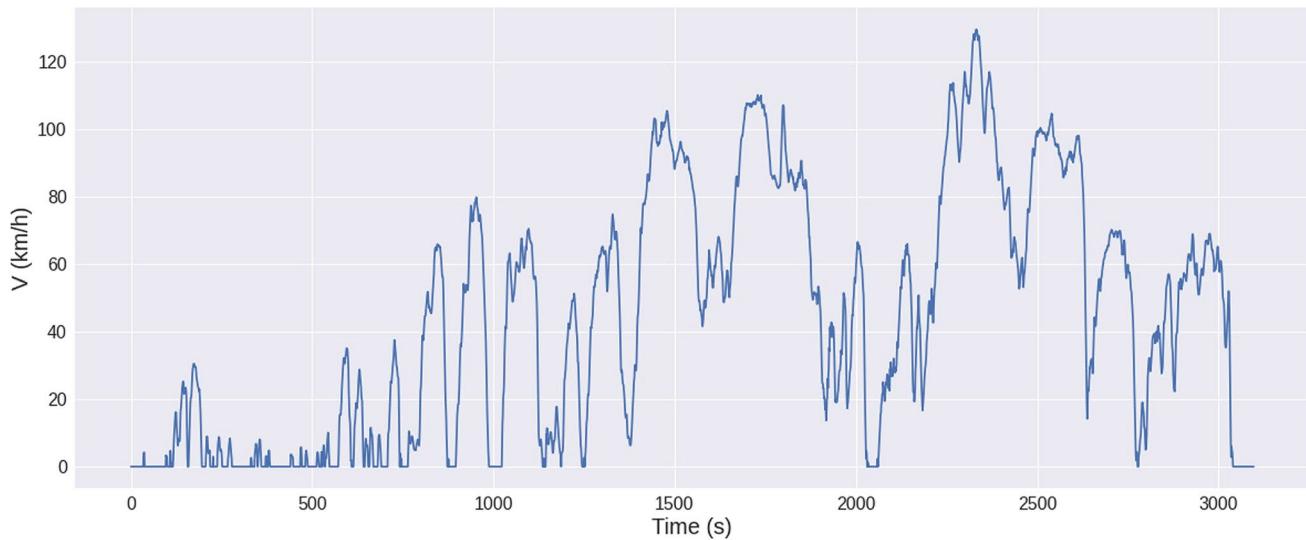


Fig. 5 Velocity of the vehicle as an input for the machine learning modeling

road gradient. These data were collected from the PEMS recording and are additionally recorded via OBDII for additional verification of the values obtained. The velocity data for the route tested for the driving cycle are shown in Fig. 5.

The purpose of collecting velocity data was to collect them for values that characterize urban, rural, and highway driving. The appropriate range covered speeds between 0 and 130 km/h. It is important to collect enough data, as the actual portion will be split into 80% as a training set and 20% as a test set. With more data collected, a more accurate model of CO₂ emissions can be created. However, it should be noted that testing with the PEMS system is relatively expensive, as a large number of calibration gases are required to zero the instrument and prepare it for performing and collecting the correct road

test data. Collecting a sufficient amount of data is crucial to the correct implementation of future predictions, as evidenced, for example, by the work of Kan et al. (2018a, b) and Zhang et al. (2022).

Equally important as data collection for velocity is data for vehicle acceleration. Data for an acceleration during the driving cycle are shown in Fig. 6.

Similarly to velocity, the acceleration parameter must be related to different driving characteristics. In urban conditions, acceleration often has to be dynamic, but sometimes in congested conditions, it is somewhat reduced. In rural conditions, acceleration refers to a velocity greater than 60–100 km/h. In general, acceleration is somewhat less dynamic in these ranges due to traffic conditions as well as vehicle capabilities, while driving for different driving styles is relevant here. Acceleration is equally relevant for

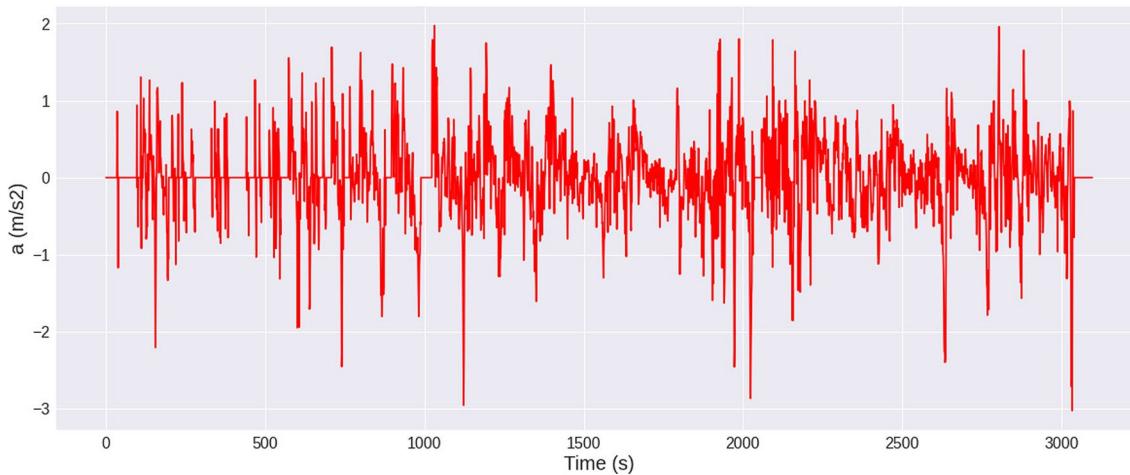


Fig. 6 Acceleration (a) of the vehicle as an input for the machine learning modeling

highway driving, where speeds range from 100 to 130 km/h. The importance of this set for the development of a good model is described, among others, in Zhang et al. (2021a) and Peng et al. (2022).

Another important parameter that provided an input value for the development of CO₂ emission models using artificial intelligence techniques was the road gradient. The road gradient is a factor that greatly influences the results obtained for CO₂ emissions. It also forms the input base for the emission models already developed. The importance of this parameter in relation to fuel consumption and emissions is indicated, among others, in the works Rosero et al. (2021) and Liu et al. (2019).

The data for velocity, acceleration, and gradient constituted the so-called independent variable, which served as input data to create a CO₂ emission model for a vehicle with start-stop technology. Therefore, the CO₂ data could be treated as a dependent variable. For the creation of a model for artificial intelligence techniques, in particular for machine learning techniques as a branch of artificial intelligence, it is important that the created models perform well and calculate with good accuracy for a new series of data that have been unused so far. Therefore, the input data for the creation of the model is divided into two portions, which are often referred to as the train-test split (Salazar et al. 2022; Rashidi et al. 2019).

The first model learning technique used was linear regression. Linear regression is used to find a linear relationship between the target and one or more predictors. Currently, vehicle emission models are being developed using this method; examples include the works of Wang et al. (2022) and Madrazo and Clappier (2018).

The second method used was the random forest machine learning method. It is based on classification algorithms (Speiser et al. 2019). Random forest, as the name suggests, is a method based on so-called decision trees, which in combination form a random forest. The number of decision trees is very large, and they operate among themselves as an ensemble. Each individual tree in the random forest produces a class prediction, and the class with the model most votes becomes the prediction (Sheykhmousa et al. 2020; Schonlau and Zou 2020). In fact, a large number of relative tree models acting as an ensemble will outperform each of the individual component models in terms of prediction accuracy (Balyan et al. 2022). The reason for the good accuracy of this model is that trees protect themselves from making individual errors.

The third method used is gradient boosting. Boosting is a method to convert weak learners into strong learners (Bentéjac et al. 2021). In boosting, each new tree is a fit to a modified version of the original data set. The gradient boosting algorithm begins by training a decision tree in which each observation is assigned an equal weight. After evaluating the first tree, we increase the weights of the observations that

are difficult to classify and decrease the weights of those that are easy to classify (Zhang et al. 2021b; Li et al. 2020). A second tree is then created on these weighted data. In this case, the idea is to improve the predictions of the first tree. Gradient boosting trains many models in an additive, sequential, and gradual manner. One of the main motivations for using gradient boosting is that it allows a user-specified cost function to be optimized, rather than a loss function, which typically offers less control and generally does not correspond to real-world applications (Punmiya and Choe 2019; Wang et al. 2020b).

Figure 7 presents the results of the scatter plot of predicted versus observed CO₂ emission and residual plots for all investigated machine learning methods. Statistical analysis of the results was carried out based on the work (Piñeiro et al. 2008). On the basis of the predicted versus observed graphs, it can be seen that the strongest correlation between the prediction data and the actual data is for the case of the gradient-boosting machine learning technique. The opposite situation applies to linear regression and the random forest technique. For the residuals plot, the distance that separates the prediction from the 0-value line and the symmetries of the created point cloud are evaluated. Positive residual values say that the average prediction was too low, while negative values mean that the prediction was too high, a value of 0 means that the prediction was perfect. Therefore, a closer localization near the 0 axis means that the created model is better able to reflect and calculate values close to the real ones (Kozak and Piepho 2018). In the case of the residuals graph, similarly to the predicted versus observed graphs, the best results are presented by the gradient boosting method. This state of affairs is justified by the most symmetrical distribution of results, which tends to form clusters near the center of the graph.

Figure 8 shows a plot of instantaneous CO₂ emissions for real and model data for two selected methods of machine learning techniques: linear regression and gradient boosting. From this graph, we can also see how many times the vehicle has gone into start-stop procedure. The engine-off locations are periods where the CO₂ emissions are 0 g/s. For the entire test route, 753 data records were found where the emission is 0 g/s. This means that the start-stop technology for the engine-off period was 12.55 min. This is a comparison of the methods and instantaneous emission results for the dataset studied. On the basis of this, it can be observed which method gives the best reflection of emission values, paying special attention to the zero emission values. These are, respectively, places where the vehicle is stopped and the engine is shut down, while not every machine learning technique is able to reflect such an engine operating condition. It can be seen from Fig. 8 that the linear regression method does not estimate CO₂ emissions correctly, erroneously indicating even values below 0. Compared to this

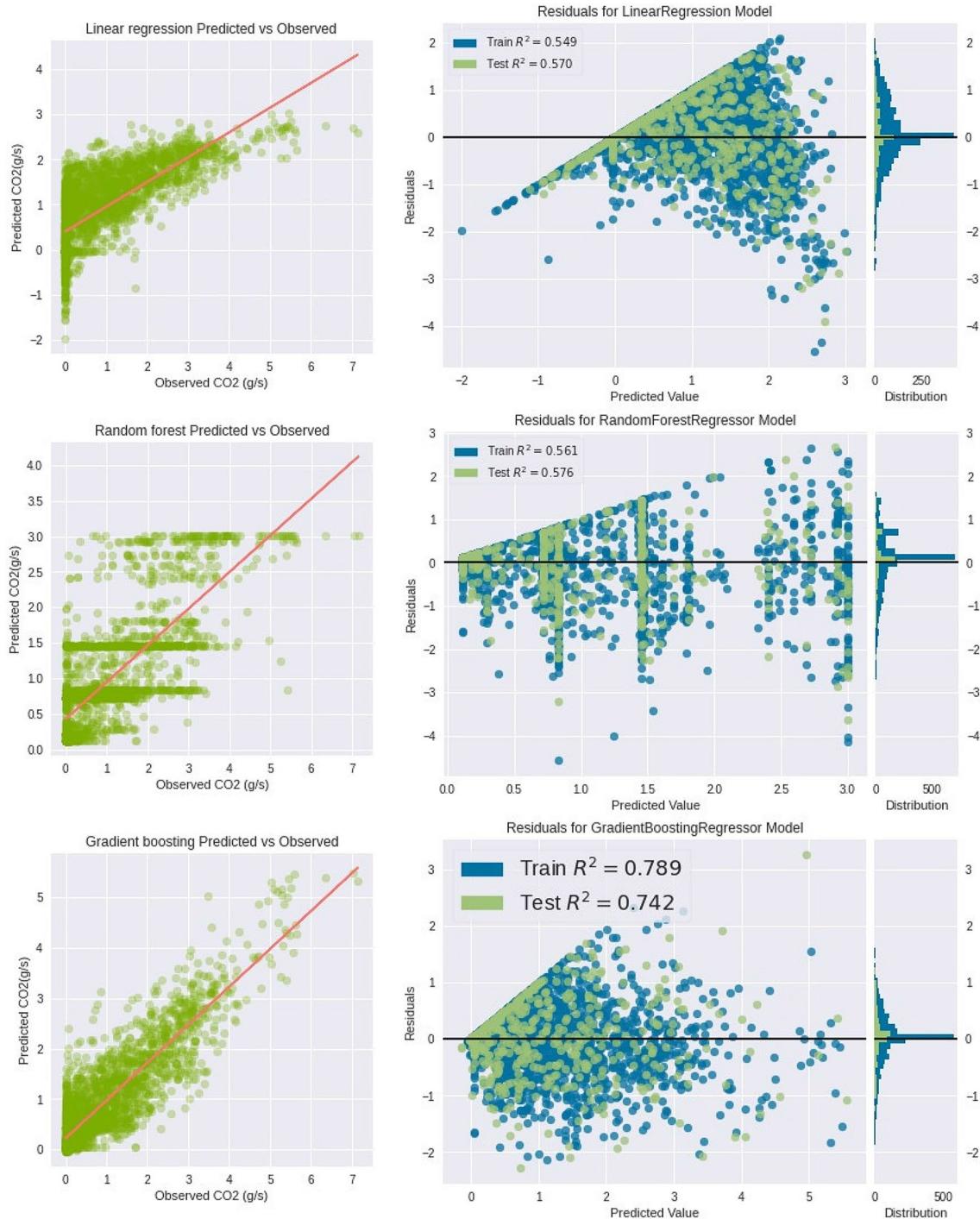


Fig. 7 Scatter plot of observed versus predicted CO₂ emission and residual plots for all machine learning methods

method, the gradient boosting method satisfactorily reflects the estimate of CO₂ emissions, even for states where the engine is turned off. An area with greater differences in CO₂ emissions is the state for a cold engine, where there was increased fuel consumption at the start of the road test, with consequent increased CO₂ emissions. For practical use

of the model and its simplicity, a separate model was not made for the engine state when it is characterized by the so-called cold start state.

Figure 9 shows the emission maps for the real-world data and the predictive data for all the machine learning techniques analyzed.

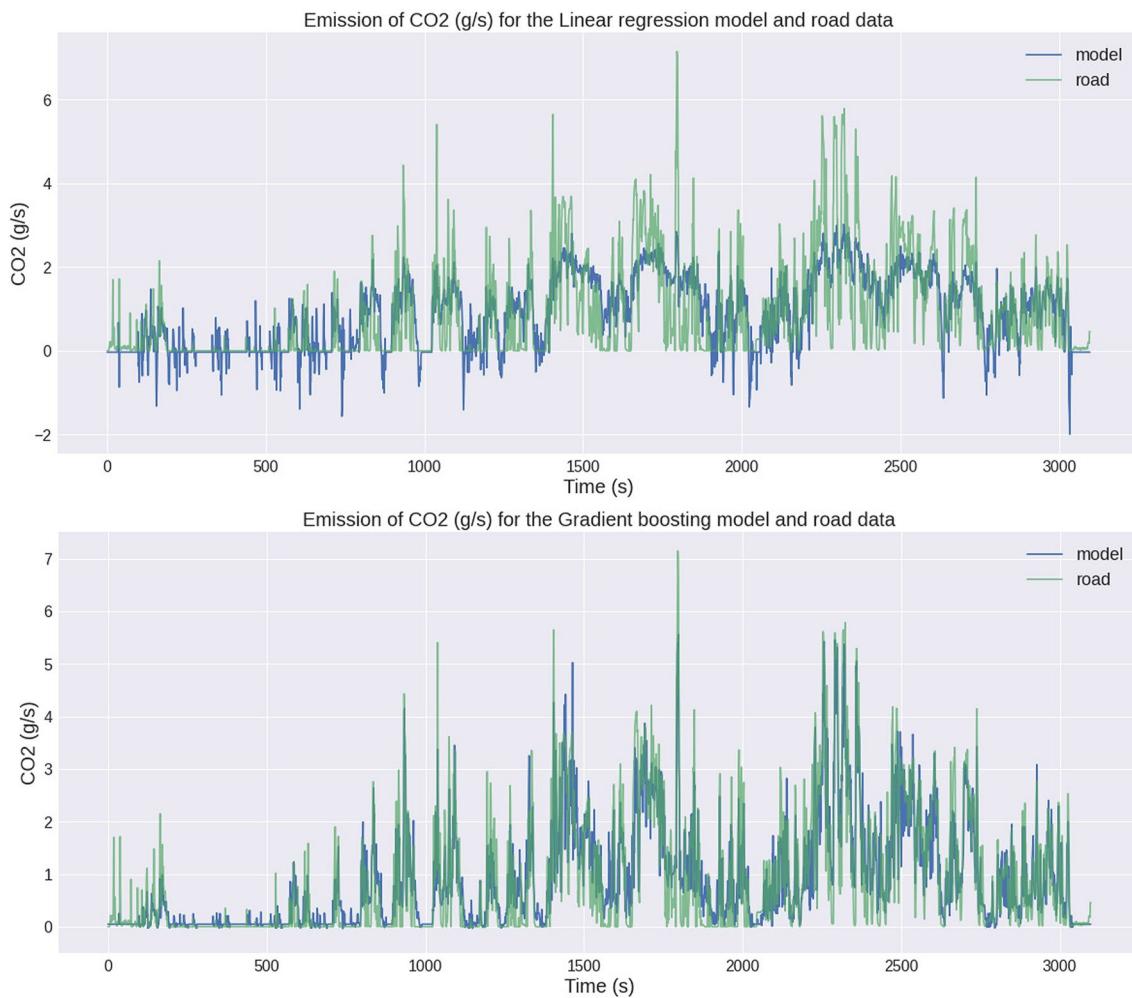


Fig. 8 Instantaneous CO₂ emission for linear regression and gradient boosting calculation method with comparison to road data

Based on the visualization of CO₂ emission maps for start-stop technology, it is possible to see which model has the best predictive ability. For all the CO₂ emission map data sets analyzed, the scale of values was unified to a range from 0 to 7 g/s. For example, the biggest differences for CO₂ prediction indications are found for the linear regression method, which largely underestimates CO₂ emissions, e.g., for a highway section, the emissions are relatively high over the course of the entire length of the highway compared to real data, where increased CO₂ emissions occur only for selected areas of highway driving. A similar underestimation characterizes the random forest method. The map for the gradient boosting technique shows the closest to real values, which actually indicates very close values of CO₂ emissions. This method has been shown to be very accurate both in terms of instantaneous emissions and CO₂ emission maps for start-stop technology. It is also worth noting that, for the urban part, where there were many stopping points, this method performs a

very accurate CO₂ calculation, also indicating zero CO₂ emission locations.

The validation of the models was also carried out using R^2 and MSE. The coefficient of determination was calculated according to formula (2) (Ueki and Alzheimer’s Disease Neuroimaging Initiative 2021):

$$R^2 = \frac{SS_M}{SS_T} = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad (2)$$

where:

- R^2 coefficient of determination,
- SS_M sum of squares for the model,
- SS_T total sum of squares,
- y_t the actual value of the dependent variable,
- \hat{y}_t predicted values of the dependent variable,
- \bar{y} The average value of the actual dependent variable,
- $\sum_{t=1}^n$ the sum of $k = 1$ to n .

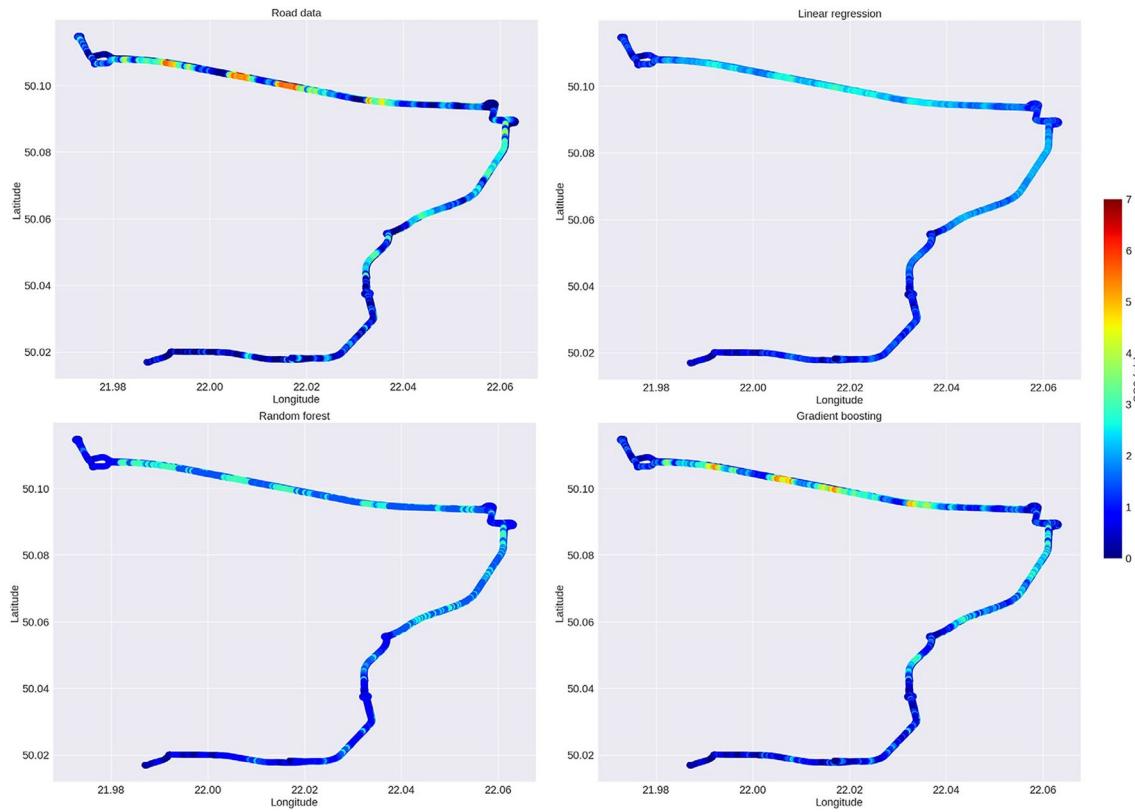


Fig. 9 CO₂ emission maps for the road data, linear regression, random forest, and gradient boosting prediction

The coefficient of determination describes how much variation in the explanatory variable was defined by the model (Chicco et al. 2021). It takes values in the range 0–1. Some sources state that the model fit is better the closer the R^2 value is to unity, while this is incorrect, because the sheer number of observation data makes the model R^2 value decrease; then, the evaluation should be carried out on the basis of other model validation parameters (Mohammad 2020). For the case under study, this involves analyzing instantaneous emission graphs and CO₂ emission maps.

The evaluation of the model's correctness is the most important part in the model development process because on this basis; it can be determined to what extent the prepared model fulfills its intended purpose. The validation of the obtained exhaust emission models was carried out on the basis of instantaneous and emission map results, using data that were not used for the calibration of earlier models. To validate the models obtained, a widely used coefficient was used to evaluate the prediction error, MSE. The mean squared error measures the amount of error in statistical models (Wang and Lu 2018; Gao et al. 2020) and assesses the average square difference between the observed and predicted values (Chani-Cahuana et al. 2018).

The mean square error was calculated based on the formula (Chachlakis et al. 2021) (3):

$$MSE = \frac{\sum (y - y_t^P)^2}{n} \quad (3)$$

where:

y the actual value of the dependent variable,
 y_t^P predicted values of the dependent variable,
 n number of observations.

The squaring of the differences in the MSE method serves several purposes. Squaring the differences eliminates negative values for the differences and ensures that the mean squared error is always greater than or equal to zero (Liu et al. 2018; Sun and Huang 2020). It is almost always a positive value. Only a perfect model without error produces an MSE of zero. Obtaining an MSE equal to zero is impossible in practice. The results shown in Table 3 confirm the previously inference methods for the analyzed to create a CO₂ model using artificial intelligence methods. For both the training set and the test set, the gradient boosting method had the smallest prediction error rates. The same is true for R^2 , which for the gradient boosting method for the training

Table 3 MSE and R^2 results for the training and test set

| Method | Training MSE | Training R^2 | Test MSE | Test R^2 |
|-------------------|--------------|----------------|----------|------------|
| Linear regression | 0.611291 | 0.549273 | 0.592783 | 0.56976 |
| Random forest | 0.596003 | 0.560545 | 0.584536 | 0.575746 |
| Gradient boosting | 0.286721 | 0.78859 | 0.355065 | 0.742295 |

set was 0.78, while for the test set, it was 0.74. These results indicate a very good representation of the real data for the model data. The worst method of those analyzed turned out to be the linear regression method, which achieved the weakest results for both the MSE and R^2 .

Discussion

The developed CO₂ emissions model for start-stop technology allows a more accurate representation of emissions, especially in urban traffic conditions, than commonly available emissions models. For example, the Versit + Enviver microscale model, developed by the Dutch research institute TNO, does not allow the analysis of this type of technology in vehicles (Quaassdorff et al. 2022; Severino et al. 2022). The VERSIT + speed profile-based emissions model, which is used in Enviver, is a multivariate regression model in which the driving cycle of a given vehicle is a variable (Anagnostopoulos and Kehagia 2018). It requires that speed profiles be obtained in advance, from which emission factors (g/km) can be estimated for different classes of vehicles (Chauhan et al. 2019). Another example of a model that also does not have an emission model specification for start-stop technology is the CMEM model. Based on power measurement, the CMEM model was developed in 2006. Emission processes are divided into different categories that correspond to physical phenomena associated with vehicle operation (Dong 2022). Each component of the exhaust gas is modeled analytically and includes parameters specific to its generation process. When developing these models, emissions were measured both directly in the engine and at the exhaust outlet of the vehicle under study (Kan et al. 2018a, b). The model requires detailed data, such as air resistance. Sometimes, however, parameters such as the air resistance value for the vehicle are not available. Therefore, instead of instantaneous power, this model uses vehicle-specific power (VSP), which is defined as the engine output per unit of vehicle mass and is expressed as a function of vehicle speed, road gradient, and acceleration (Perugu 2019; Acuto et al. 2022).

Other models of CO₂ emissions are also available that are based on artificial intelligence techniques. An example is the model described in Mądziel et al. (2021). The authors of this work developed a CO₂ emission model for a hybrid vehicle using the Gaussian process regression (GPR) machine learning technique. The model yielded R^2 values of 0.69, while the MSE was 1.08. The velocity, acceleration, and road gradient parameters were also used for the input data to create the model. Another example of the use of machine learning techniques can be found in work (Jaworski et al. 2019), in which the authors, for vehicles that meet Euro 2-Euro 6 standards, created a special emissions model dedicated to the traffic characteristics of driving in roundabouts. The best method of machine learning techniques for this work turned out to be the boosted regression trees method, while validation of the model results was based on statistical analysis of coefficient of regression, root mean square error, and mean absolute error, and based on the visual assessment of the results, it was shown that the obtained models are well represented by real data. Another example is a novel approach to modeling machine learning emission to reduce vehicle brake emissions (Wei et al. 2022). The authors used a dataset of 600 real-world braking events. On the basis of the five algorithms, the importance of the different features of the fragments was discussed. Prediction performance was measured using R^2 and RMSE. An interesting work in terms of emission modeling is work (Xu et al. 2020) where the authors proposed the deep learning reinforcement emission control strategy, which automatically learns the optimal traffic flow. Another work that presents a similar methodology for creating a CO₂ emissions model is the work (Mądziel 2023a). This work shows the development of a novel methodology to create microscale CO₂ emission models tailored specifically designed for LPG vehicles. The model is meticulously constructed, drawing on data collected from road tests utilizing the portable emission measurement system (PEMS) and the onboard diagnostic (OBDII) interface. Using advanced gradient-boosting machine learning techniques, this model was derived following an in-depth exploratory data analysis. Using vehicle velocity and engine RPM as key explanatory variables for a precise CO₂ prediction. Rigorous validation of the model shows its precision, rendering it a valuable tool for continuous CO₂ emission analysis and the creation of emission maps essential for environmental assessments in urban areas. The validation coefficients underscore its effectiveness, with an R^2 test score of 0.61 and an MSE test score of 0.77. Another example would be a paper (Estrada et al. 2023) in which the authors use different input variables to create the model which limits its use for purposes such as simulating vehicle movement. This article introduces a novel approach to developing hybrid models designed to

facilitate the testing of EMS and the assessment of fuel consumption, as well as CO₂ and pollutant emissions (CO, NO_x, and THC). In the context of this research, the authors highlight the limitations of static models like the widely-used map-based approach, which struggles to accurately quantify pollutant emissions due to the significant influence of transient effects. The primary innovation in this paper lies in the utilization of computational neural networks (CNN) to characterize pollutant emissions, thereby achieving high precision for both instantaneous and cumulative values. The input parameters used in this methodology are classical measurements associated with ICEs, which encompass engine speed, air mass flow, torque, and exhaust temperature. Another work that is a good and current example of the use of artificial intelligence methods for modeling CO₂ emissions from vehicles is that of Mądziel (2023c), where this process is described for low emission zones in cities. In this research, a two-dimensional emission model for hybrid vehicles is introduced. This model utilizes artificial neural networks specifically designed for low-emission zones. The primary achievement of this study is the development of a tailored CO₂ emission model for hybrid vehicles, enabling the simulation of different road solutions. Notably, the CO₂ emission model exhibited a high R^2 coefficient of 0.73 and a low MSE of 0.91. There are many methods for developing data for emission models, while it is important to choose the right one relative to the assumed expectations. This problem poses a major challenge for researchers, while a growing body of work contributes to better matching the appropriate methods with the expected predictive results. When it comes to selecting the right method, it is also useful to make use of compiled reviews of research works in the field of vehicle emissions modeling, an example is the work of (Mądziel 2023b).

From a review of the sources of other works, a gap emerged in the absence of a model specifically dedicated to vehicle equipped with start-stop technology. The model developed using artificial intelligence techniques addresses this need. Undoubtedly, this model can be used for a more accurate analysis of emission sources, especially in the urban part, where the highest traffic density actually occurs. A better representation of the emissions for such areas can undoubtedly contribute to a more accurate analysis of vehicle emissions and the impact of these emissions on the health of pedestrians or cyclists who travel along urban arteries or cross them at pedestrian crossings. Ecological analyses of vehicle emission sources are growing in importance every year. Microscale models can also be used to study the feasibility of investments in terms of ecological analysis. Emission maps obtained from the model can be helpful to decision-makers in city managements, in order to make better decisions for regional development of urban agglomerations taking into account environmental aspects.

Conclusions

In this study, a methodology for creating a machine learning model for predicting CO₂ emissions is presented for a test vehicle equipped with start-stop technology. Several possible machine learning techniques for CO₂ prediction are presented, including linear regression, random forest, and gradient boosting methods, and the validation indices of these methods are indicated. The gradient boosting calculation method has the best predictive capability for CO₂ emissions. The model validation indices for this method were determined separately for the training set (which represented 80% of the data used to build the model) and the test set (which consisted of 20% of previously unseen data). For the training set, the mean squared error (MSE) was 0.28, and the coefficient of determination (R^2) was 0.78. In the case of the test set, the MSE was 0.35, and the R^2 was 0.74.

The main conclusions of the work are:

- The computational technique of machine learning gradient boosting for the case studied had the best predictive capabilities for start-stop technology.
- The developed models could be used in the development of environmental analysis by, for example, local governments.
- The developed methodology may be scalable for the analysis of emissions from other vehicles with start-stop technology.
- The developed model is based on input parameters: velocity, vehicle acceleration, and road gradient. Therefore, it can also be used in simulation post-programming such as Vissim.

The limitations of the work are the development of the model based only on the data of one vehicle with start-stop technology. Although this is a preliminary model that will be expanded, the idea of the work was to select a machine learning technique that can reflect zero emissions during vehicle stops.

Future work includes continuous improvement of the model by measuring CO₂ emissions for more vehicles. The work will specifically address the inclusion of hybrid vehicles in the test fleet. It is desirable to learn and test another pool of different artificial intelligence computing techniques that can be used to estimate CO₂ emissions even more accurately under traffic conditions.

Nomenclature *a*: Vehicle acceleration (m/s²); ANN: Artificial neural network; CLD: Chemiluminescence detector; CO: Carbon monoxide; CO₂: Carbon dioxide; CPU: Central processing unit; CR: Common rail; DOC: Diesel oxidation catalysts; DT: Analyzer measurement delays; DPF: Diesel particulate filter; FID: Flame ionization detector; GPU: Graphics processing unit; GPR: Gaussian process regression; GRU: Gated recurrent unit; M: Molecular weight; MLP: Multilayer

perceptron; MSE: Mean squared error; NDIR: Nondispersive infrared; NO_x: Nitrogen oxides; OBD: On-board diagnostic; PEMS: Portable emission measurement system; R^2 : Coefficient of determination; SCR: Selective catalytic reduction; TPU: Tensor processing unit; Q : Gas flow rate; WLTP: Worldwide Harmonized Light-Duty Vehicles Test Procedure; V : Vehicle velocity (km/h)

Author contribution Data curation: Maksymilian Madziel; formal analysis: Maksymilian Madziel; investigation: Maksymilian Madziel; methodology: Maksymilian Madziel; project administration: Maksymilian Madziel; writing original draft: Maksymilian Madziel writing, review, and editing: Maksymilian Madziel. All authors read and approved the final manuscript.

Funding This work was supported by the Ministry of Infrastructure and Development under the Eastern Poland Development Operational Program including European Regional Development Fund, which financed the research instruments.

Data availability The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Acuto F, Coelho MC, Fernandes P, Giuffrè T, Macioszek E, Granà A (2022) Assessing the environmental performances of urban roundabouts using the VSP methodology and AIMSUM. Energies 15(4):1371
- Anagnostopoulos A, Kehagia F (2018) Turbo-roundabouts as an alternative to roundabouts in terms of traffic safety, capacity and pollutant emissions. In Proceedings of the 7th Pan-Hellenic Road Safety Conference, Larissa, Greece (pp. 11–12)
- Anenberg SC, Achakulwisut P, Brauer M, Moran D, Apte JS, Henze DK (2019) Particulate matter-attributable mortality and relationships with carbon dioxide in 250 urban areas worldwide. Sci Rep 9(1):1–6
- Azeez OS, Pradhan B, Shafri HZ, Shukla N, Lee CW, Rizzei HM (2019) Modeling of CO emissions from traffic vehicles using artificial neural networks. Appl Sci 9(2):313
- Balyan AK, Ahuja S, Lilhore UK, Sharma SK, Manoharan P, Algarni AD, ... & Raahemifar K (2022) A hybrid intrusion detection model using EGA-PSO and improved random forest method. Sensors 22(16):5986
- Bentéjac C, Csörgő A, Martínez-Muñoz G (2021) A comparative analysis of gradient boosting algorithms. Artif Intell Rev 54(3):1937–1967
- Bharj RS, Kumar R, Singh GN (2019) On-board post-combustion emission control strategies for diesel engine in India to meet Bharat stage VI norms. In Advanced Engine Diagnostics (pp. 105–125). Springer, Singapore
- Bisong E (2019) Building machine learning and deep learning models on Google cloud platform. Apress, Berkeley, CA, pp 59–64
- Borge R, Narros A, Artíñano B, Yagüe C, Gómez-Moreno FJ, de la Paz D, ... & Vardoulakis S (2016) Assessment of microscale spatio-temporal variation of air pollution at an urban hotspot in Madrid (Spain) through an extensive field campaign. Atmos Environ 140:432–445
- Campisi T, Mądzioł M, Nikiforidis A, Basbas S, Tesoriere G (2021) An estimation of emission patterns from vehicle traffic highlighting decarbonization effects from increased e-fleet in areas surrounding the City of Rzeszow (Poland). In International Conference on Computational Science and Its Applications (pp. 683–698). Springer, Cham
- Carneiro T, Da Nóbrega RVM, Nepomuceno T, Bian GB, De Albuquerque VHC, Reboucas Filho PP (2018) Performance analysis of Google Colaboratory as a tool for accelerating deep learning applications. IEEE Access 6:61677–61685
- Chachlakis DG, Zhou T, Ahmad F, Markopoulos PP (2021) Minimum mean-squared-error auto-correlation processing in coprime arrays. Digit Signal Process 114:103034
- Chani-Cahuana J, Fager C, Eriksson T (2018) Lower bound for the normalized mean square error in power amplifier linearization. IEEE Microwave Wirel Compon Lett 28(5):425–427
- Chauhan BP, Joshi GJ, Parida P (2019) Car following model for urban signalised intersection to estimate speed based vehicle exhaust emissions. Urban Climate 29:100480
- Chicco D, Warrens MJ, Jurman G (2021) The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci 7:e623
- Deng Z, Yang Q, Yang X (2020) Optimal design and experimental evaluation of magneto-rheological mount applied to start/stop mode of vehicle powertrain. J Intell Mater Syst Struct 31(8):1126–1137
- DiPierro G, Millo F, Cubito C, Ciuffo B, Fontaras G (2019) Analysis of the impact of the WLTP procedure on CO₂ emissions of passenger cars (no. 2019–24–0240). SAE Technical Paper
- Dong Z (2022) Quantifying fuel consumption and emission saving of hybrid vehicles over gasoline in real traffic (Doctoral dissertation, The Ohio State University)
- Eijk ARA, Stelwagen U, Overheid OV (2018) EnViVer voor Vlaanderen Ontwikkeling van Vlaamse EnViVer emissiemodellen. Den Haag: TNO
- Estrada PM, de Lima D, Bauer PH, Mammetti M, Bruno JC (2023) Deep learning in the development of energy management strategies of hybrid electric vehicles: a hybrid modeling approach. Appl Energy 329:120231
- Franta B (2018) Early oil industry knowledge of CO₂ and global warming. Nat Clim Chang 8(12):1024–1025
- Gao C, Shi D, Maydeu-Olivares A (2020) Estimating the maximum likelihood root mean square error of approximation (RMSEA) with non-normal data: a Monte-Carlo study. Struct Equ Model: A Multi-Disciplinary Journal 27(2):192–201
- Gao J, Wang Y, Chen H, Laurikko J, Liu Y, Pellikka AP, Li Y (2022) Variations of significant contribution regions of NO_x and PN

- emissions for passenger cars in the real-world driving. *J Hazard Mater* 424:127590
- Giannakis E, Serghides D, Dimitriou S, Zittis G (2020) Land transport CO₂ emissions and climate change: evidence from Cyprus. *Int J Sustain Energ* 39(7):634–647
- Giechaskiel B, Joshi A, Ntziachristos L, Dilara P (2019) European regulatory framework and particulate matter emissions of gasoline light-duty vehicles: a review. *Catalysts* 9(7):586
- Grigoratos T, Fontaras G, Giechaskiel B, Zacharof N (2019) Real world emissions performance of heavy-duty Euro VI diesel vehicles. *Atmos Environ* 201:348–359
- Hao J, Ho TK (2019) Machine learning made easy: a review of Scikit-learn package in Python programming language. *J Educ Behav Stat* 44(3):348–361
- Hao L, Namuduri CS, Gopalakrishnan S, Lee CJ, Shidore NS, Pandi M, Vandermeir T (2020) Brushless fast starter for automotive engine start/stop application. *IEEE Trans Ind Appl* 56(6):6041–6052
- Hoang AT, Nižetić S, Ong HC, Tarelko W, Le TH, Chau MQ, Nguyen XP (2021) A review on application of artificial neural network (ANN) for performance and emission characteristics of diesel engine fueled with biodiesel-based fuels. *Sustain Energy Technol Assess* 47:101416
- Hulagu S, Celikoglu HB (2021) Comparative evaluation of macro and micro approaches to emission modeling using GPS data: a case study. *Trans Res Procedia* 52:629–636
- Jaworski A, Mądziel M, Lejda K (2019) Creating an emission model based on portable emission measurement system for the purpose of a roundabout. *Environ Sci Pollut Res* 26(21):21641–21654
- Jaworski A, Lejda K, Mądziel M, Ustrzycki A (2018) Assessment of the emission of harmful car exhaust components in real traffic conditions. In IOP Conference Series: Materials Science and Engineering (Vol. 421, No. 4, p. 042031). IOP Publishing
- Kan Z, Tang L, Kwan MP, Zhang X (2018a) Estimating vehicle fuel consumption and emissions using GPS big data. *Int J Environ Res Public Health* 15(4):566
- Kan Z, Tang L, Kwan MP, Ren C, Liu D, Pei T, Li Q (2018b) Fine-grained analysis on fuel-consumption and emission from vehicles trace. *J Clean Prod* 203:340–352
- Kozak M, Piepho HP (2018) What's normal anyway? Residual plots are more telling than significance tests when checking ANOVA assumptions. *J Agron Crop Sci* 204(1):86–98
- Kumar PG, Lekhana P, Tejaswi M, Chandrakala S (2021) Effects of vehicular emissions on the urban environment-a state of the art. *Mater Today: Proceedings* 45:6314–6320
- Le Corne CM, Molden N, van Reeuwijk M, Stettler ME (2020) Modelling of instantaneous emissions from diesel vehicles with a special focus on NO_x: insights from machine learning techniques. *Sci Total Environ* 737:139625
- Lejri D, Can A, Schipper N, Leclercq L (2018) Accounting for traffic speed dynamics when calculating COPERT and PHEM pollutant emissions at the urban scale. *Transp Res Part d: Transp Environ* 63:588–603
- Li Q, Wen Z, He B (2020) Practical federated gradient boosting decision trees. *Proc AAAI Conf Artif Intell* 34(04):4642–4649
- Lijewski P, Szymlet N, Fuć P, Domowicz A, Rymaniak Ł (2021) The effect of start-stop systems on scooter exhaust emissions. *Transp Res Part d: Transp Environ* 91:102684
- Liu H, Rodgers MO, Guensler R (2019) The impact of road grade on vehicle accelerations behavior, PM_{2.5} emissions, and dispersion modeling. *Transp Res Part d: Transp Environ* 75:297–319
- Liu G, Chen L, Wang W, Chen X (2018) Hybrid beamforming based on minimum mean square error for multi-user multi-data stream system. In 2018 IEEE 4th International Conference on Computer and Communications (ICCC) (pp. 124–128). IEEE
- Madrazo J, Clappier A (2018) Low-cost methodology to estimate vehicle emission factors. *Atmos Pollut Res* 9(2):322–332. <https://doi.org/10.1016/j.apr.2017.10.006>
- Mądziel M (2023a) Liquified petroleum gas-fuelled vehicle CO₂ emission modelling based on portable emission measurement system, on-board diagnostics data, and gradient-boosting machine learning. *Energies* 16(6):2754. <https://doi.org/10.3390/en16062754>
- Mądziel M (2023b) Vehicle emission models and traffic simulators: a review. *Energies* 16:3941. <https://doi.org/10.3390/en16093941>
- Mądziel M (2023c) Future cities carbon emission models: hybrid vehicle emission modelling for low-emission zones. *Energies* 16:6928. <https://doi.org/10.3390/en16196928>
- Mądziel M, Campisi T (2022) Assessment of vehicle emissions at roundabouts: a comparative study of PEMS data and microscale emission model. *Arch Transport* 63(3):35–51
- Mądziel M, Jaworski A, Kuszewski H, Woś P, Campisi T, Lew K (2021) The development of CO₂ instantaneous emission model of full hybrid vehicle with the use of machine learning techniques. *Energies* 15(1):142
- Mikhaylov A, Moiseev N, Alešin K, Burkhardt T (2020) Global climate change and greenhouse effect. *Entrep Sustain Issues* 7(4):2897
- Mohammad GS (2020) Estimate accuracy of the sample determination coefficient R² and R² for the multiple regression models. *J Southwest Jiaotong Univ* 55(1)
- Montanarella L, Panagos P (2021) The relevance of sustainable soil management within the European Green Deal. *Land Use Policy* 100:104950
- Pavlovic J, Ciuffo B, Fontaras G, Valverde V, Marotta A (2018) How much difference in type-approval CO₂ emissions from passenger cars in Europe can be expected from changing to the new test procedure (NEDC vs. WLTP)? *Trans Res Part a: Policy and Practice* 111:136–147
- Peng C, Wang Y, Xu T, Chen Y (2022) Transient fuel consumption prediction for heavy-duty trucks using on-road measurements. *Int J Sustain Trans* 1–12
- Perugu H (2019) Emission modelling of light-duty vehicles in India using the revamped VSP-based MOVES model: the case study of Hyderabad. *Transp Res Part d: Transp Environ* 68:150–163
- Piñeiro G, Perelman S, Guerschman JP, Paruelo JM (2008) How to evaluate models: observed vs. predicted or predicted vs. observed? *Ecol Model* 216(3–4):316–322
- Punmiya R, Choe S (2019) Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing. *IEEE Trans Smart Grid* 10(2):2326–2329
- Quaassdorff C, Borge R, Pérez J, Lumbreras J, de la Paz D, de Andrés JM (2016) Microscale traffic simulation and emission estimation in a heavily trafficked roundabout in Madrid (Spain). *Sci Total Environ* 566:416–427
- Quaassdorff C, Smit R, Borge R, Hausberger S (2022) Comparison of microscale traffic emission models for urban networks. *Environ Res Lett* 17(9):094030
- Raschka S, Patterson J, Nolet C (2020) Machine learning in python: main developments and technology trends in data science, machine learning, and artificial intelligence. *Information* 11(4):193
- Rashidi HH, Tran NK, Betts EV, Howell LP, Green R (2019) Artificial intelligence and machine learning in pathology: the present landscape of supervised methods. *Acad Pathol* 6:2374289519873088
- Rodriguez-Rey D, Guevara M, Linares MP, Casanovas J, Salmerón J, Soret A, ... & García-Pando CP (2021) A coupled macroscopic traffic and pollutant emission modelling system for Barcelona. *Transp Res Part D: Transp Environ* 92:102725
- Rosero F, Fonseca N, López JM, Casanova J (2021) Effects of passenger load, road grade, and congestion level on real-world fuel

- consumption and emissions from compressed natural gas and diesel urban buses. *Appl Energy* 282:116195
- Rubio F, Llopis-Albert C, Valero F, Besa AJ (2020) Sustainability and optimization in the automotive sector for adaptation to government vehicle pollutant emission regulations. *J Bus Re-Search* 112:561–566
- Salazar JJ, Garland L, Ochoa J, Pyrcz MJ (2022) Fair train-test split in machine learning: mitigating spatial autocorrelation for improved prediction accuracy. *J Petrol Sci Eng* 209:109885
- Sanchez B, Santiago JL, Martilli A, Martin F, Borge R, Quaassdorff C, de la Paz D (2017) Modelling NO_x concentrations through CFD-RANS in an urban hot-spot using high resolution traffic emissions and meteorology from a mesoscale model. *Atmos Environ* 163:155–165
- Santos NDSA, Roso VR, Faria MTC (2020) Review of engine journal bearing tribology in start-stop applications. *Eng Fail Anal* 108:104344
- Santos NDSA, Alvarez CEC, Roso VR, Baeta JGC, Valle RM (2021) Lambda load control in spark ignition engines, a new application of prechamber ignition systems. *Energy Convers Manage* 236:114018
- Schonlau M, Zou RY (2020) The random forest algorithm for statistical learning. *Stand Genomic Sci* 20(1):3–29
- Selleri T, Gioria R, Melas AD, Giechaskiel B, Forloni F, Mendoza Villafuerte P, ... & Suarez-Bertoa R (2022) Measuring emissions from a demonstrator heavy-duty diesel vehicle under real-world conditions—moving forward to Euro VII. *Catalysts* 12(2):184
- Severino A, Pappalardo G, Olayode IO, Canale A, Campisi T (2022) Evaluation of the environmental impacts of bus rapid transit system on turbo roundabout. *Trans Eng* 9:100130
- Shah SHH, Lei S, Ali M, Doronin D, Hussain ST (2019) Prosumption: bibliometric analysis using HistCite and VOSviewer. *Kybernetes*
- Sheykhouza M, Mahdianpari M, Ghanbari H, Mohammadimanesh F, Ghamisi P, Homayouni S (2020) Support vector machine versus random forest for remote sensing image classification: a meta-analysis and systematic review. *IEEE J Sel Top Appl Earth Observ Re-Mote Sens* 13:6308–6325
- Sikora A (2021) European Green Deal—legal and financial challenges of the climate change. *Era Forum* 21(4):681–697 Springer Berlin Heidelberg
- Speiser JL, Miller ME, Tooze J, Ip E (2019) A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst Appl* 134:93–101
- Sun W, Huang C (2020) A carbon price prediction model based on secondary decomposition algorithm and optimized back propagation neural network. *J Clean Prod* 243:118671
- Tipanluisa L, Fonseca N, Casanova J, López JM (2021) Effect of n-butanol/diesel blends on performance and emissions of a heavy-duty diesel engine tested under the World Harmonised Steady-State cycle. *Fuel* 302:121204
- Ueki M, Alzheimer's Disease Neuroimaging Initiative (2021) Testing conditional mean through regression model sequence using Yanai's generalized coefficient of determination. *Comput Stat Data Anal* 158:107168
- Varella RA, Ribau JP, Baptista PC, Sousa L, Duarte GO (2019) Novel approach for connecting real driving emissions to the European vehicle laboratorial certification test procedure. *Environ Sci Pollut Res* 26(34):35163–35182
- Wang W, Lu Y (2018) Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model. *IOP Conf Ser: materials science and engineering* 324(1):012049 IOP Publishing
- Wang D, Jiang M, He K, Li X, Li F (2020a) Study on vibration suppression method of vehicle with engine start-stop and automatic start-stop. *Mech Syst Signal Process* 142:106783
- Wang L, Wu C, Tang L, Zhang W, Lacasse S, Liu H, Gao L (2020b) Efficient reliability analysis of earth dam slope stability using extreme gradient boosting method. *Acta Geotech* 15(11):3135–3150
- Wang T, Qz Z, Yang Z, Nichol T, Clarke G, Ge YE (2020c) Climate change research on transportation systems: climate risks, adaptation and planning. *Transp Res Part d: Transp Environ* 88:102553
- Wang X, Chen LWA, Lu M, Ho KF, Lee SC, Ho SSH, ... & Watson JG (2022) Apportionment of vehicle fleet emissions by linear regression, positive matrix factorization, and emission modeling. *Atmosphere* 13(7):1066
- Wei N, Men Z, Ren C, Jia Z, Zhang Y, Jin J, ... & Mao H (2022) Applying machine learning to construct braking emission model for real-world road driving. *Environ Int* 166:107386
- Weller K, Lipp S, Röck M, Matzer C, Bittermann A, Hausberger S (2019) Real world fuel consumption and emissions from LDVs and HDVs. *Front Mech Eng* 5:45
- Xu Z, Cao Y, Kang Y, Zhao Z (2020) Vehicle emission control on road with temporal traffic information using deep reinforcement learning. *IFAC-PapersOnLine* 53(2):14960–14965
- Yoro KO, Daramola MO (2020) CO₂ emission sources, greenhouse gases, and the global warming effect. In *Advances in carbon capture* (pp. 3–28). Woodhead Publishing
- Zhang W, Wu C, Zhong H, Li Y, Wang L (2021b) Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. *Geosci Front* 12(1):469–477
- Zhang L, Peng K, Zhao X, Khattak AJ (2021a) New fuel consumption model considering vehicular speed, acceleration, and jerk. *J Intell Trans Syst* 1–13
- Zhang C, Gong J, He J, Bai C, Yan X, Wang C, ... & Wang H (2022) Modeling the unobserved heterogeneity effects of the factors influencing the fuel consumption of heavy-duty diesel trucks under real road conditions: a preliminary investigation in China. *Energy Rep* 8:9586–9597
- Zhu J, Yang Z, Guo Y, Zhang J, Yang H (2019) Short-term load forecasting for electric vehicle charging stations based on deep learning approaches. *Appl Sci* 9(9):1723. <https://doi.org/10.3390/app9091723>
- Zhu R, Fu Y, Wang L, Hu J, He L, Wang M, ... & Su S (2022) Effects of a start-stop system for gasoline direct injection vehicles on fuel consumption and particulate emissions in hot and cold environments. *Environ Pollut* 308:119689

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