Available online at www.sciencedirect.com

ScienceDirect

00 (2020) 1–28

Modelling of instantaneous emissions from diesel vehicles with a special focus on NO_x: Insights from machine learning techniques

Clémence M. A. Le Cornec[a], Nick Molden[b], Maarten van Reeuwijk[a], Marc E. J. Stettler[a,*]

[a] *Department of Civil and Environmental Engineering, Imperial College London, London, SW7 2AZ, United Kingdom*
[b] *Emissions Analytics, Kimball Smith Limited, Kings Worthy House, Court Road, Kings Worthy, Winchester, SO23 7QA, United Kingdom*

Abstract

Accurate instantaneous vehicle emissions models are vital for evaluating the impacts of road transport on air pollution at high temporal and spatial resolution. In this study, we apply machine learning techniques to a dataset of 70 diesel vehicles tested in real-world driving conditions to: (i) cluster vehicles with similar emissions performance, and (ii) model instantaneous emissions. The application of dynamic time warping and clustering analysis by NO_x emissions resulted in 17 clusters capturing 88% of trips in the dataset. We show that clustering effectively groups vehicles with similar emissions profiles, however no significant correlation between emissions and vehicle characteristics (i.e. engine size, vehicle weight) were found. For each cluster, we evaluate three instantaneous emissions models: a look-up table (LT) approach, a non-linear regression (NLR) model and a neural network multi-layer perceptron (MLP) model. The NLR model provides accurate instantaneous NO_x predictions, on par with the MLP: relative errors in prediction of emission factors are below 20% for both models, average fractional biases are -0.01 (s.d. 0.02) and -0.0003 (s.d. 0.04), and average normalised mean squared errors are 0.25 (s.d. 0.14) and 0.29 (s.d. 0.16), for the NLR and MLP models respectively. However, neural networks are better able to deal with vehicles not belonging to a specific cluster. The new models that we presented rely on simple inputs of vehicle speed and acceleration, which could be extracted from existing sources including traffic cameras and vehicle tracking devices, and can therefore be deployed immediately to enable fast and accurate prediction of vehicle NO_x emissions. The speed and the ease of use of these new models make them an ideal operational tool for policy makers aiming to build emission inventories or evaluate emissions mitigation strategies.

Keywords:
NO_x emissions, instantaneous models, artificial neural networks, machine learning

1. Introduction

Outdoor air pollution significantly affects public health and is responsible for about 428,000 premature deaths annually in Europe (EEA, 2017). Emissions of nitrogen oxides (NO_x = NO + NO_2) substantially deteriorate air quality ultimately leading to adverse short and long term environmental and health effects (EEA, 2017). The European regulations on ambient air quality limit NO_2 concentrations whereas the European regulations on exhaust emissions

*Corresponding author:
Email address: m.stettler@imperial.ac.uk (Marc E. J. Stettler)

limit the NO_x emission factor (grams of NO_x emitted per kilometre). Exposure to NO_2 is associated with an increased risk of cardiovascular diseases, decreased lung functionalities, aggravated asthma, low birth weights and premature deaths (EEA, 2017; Andersen et al., 2012; Vineis et al., 2006). NO_2 and O_3 have also been linked with the degradation of natural ecosystems and reduction in biodiversity (EEA, 2017). Despite evidence of adverse impacts and increasingly stringent regulations by the EU on both air quality and emissions, NO_2 concentrations in Europe have remained significantly higher than the limits set for the protection of human life (EEA, 2017).

Road transport accounted for 39% of the total NO_x emissions in Europe in 2015 (EEA, 2017). The contribution of road transport to ambient NO_2 concentration is even higher in urban areas where emissions typically happen close to the ground. In London, the Roads Task Force, Technical Note 21 (2012) estimated that road transport contributed to 47% of the total NO_x emissions in 2012. Diesel engines, which have historically had intrinsic higher distance-specific NO_x emissions than gasoline engines (Grinham and Parry, 2018), are responsible for approximately 80% of the NO_x emissions due to road transport. The market penetration of diesel vehicles, which represents 40% of the total fleet in the UK (Grinham and Parry, 2018), and the significantly higher emissions measured in real driving conditions (O'Driscoll et al., 2016, 2017) compared to type approval partially explain the failure to reach air quality standards. A various number of mitigation strategies exist to control air pollution in urban areas. The distinction between high and low emitting vehicles is typically based on the emission standard as attested by the manufacturer during type approval, despite the huge discrepancies in emissions between type approval and real-world driving conditions (O'Driscoll et al., 2016, 2017; Ligterink et al., 2013; Weiss et al., 2012; Franco et al., 2014). To mitigate this latter issue, Real-Driving Emissions (RDE) test procedure was introduced for new vehicles from September 2017 onwards (European Commission, 2018).

Emissions models, combined with atmospheric dispersion models, are required to assess the impacts of road transport on the local air quality and implement successful mitigation strategies. While most of the current emissions models are based on physical, chemical, empirical or semi-empirical assumptions, Panis et al. (2006) report that it is extremely difficult to model instantaneous NO_x emissions using these classical approaches. Existing emissions models achieve limited performance to estimate NO_x emissions in real-world driving conditions (Smit et al., 2010). O'Driscoll et al. (2016) demonstrate that COPERT4.11 underestimates NO_x emissions in real-driving conditions by a factor 1.6 compared to real measurements. PHEM overestimates NO_x emissions in real-driving conditions by up to 25% (Matzer et al., 2017). Finally, Ntziachristos et al. (2016) show that VERSIT+ emission factors underestimate NO_x emissions by a factor up to 1.5 compared to emissions in real-world driving conditions. The lack of understanding of the highly non-linear processes behind real-world emissions partially explains the limited performance of the existing emission models. Improvements to real-time estimation of NO_x emissions is therefore vital to the assessment

of road transport responsibility in air quality issues and will only be possible by improving our understanding of the processes governing real-world emissions. With the continuous increase in computational power and data availability, unsupervised machine learning techniques, which are able to successfully find previously unknown patterns in datasets, could help to improve emission models. Similarly, Artificial Neural Networks (ANNs), which have shown their effectiveness in solving a wide range of non-linear problems, could help to improve emissions models.

To the knowledge of the authors, no studies have used unsupervised machine learning techniques in order to acquire a better understanding of the processes underlying real-world instantaneous NO_x emissions and ultimately improve emission models. However, ANNs have been used to predict fuel consumption (Parlak et al., 2006; Thompson et al., 2000) or particulate emissions (de Lucas et al., 2001). A drawback of these studies was that only a small amount of data, acquired on a chassis-dynamometer for standardised type-approval tests (US06, NEDC), was available to perform the network training and validation. In addition, the networks' architectures developed were simple, consisting of a multi-layer perceptron of one or two hidden layers. The models show reasonable performance but are specific to a small set of vehicles operated in specific conditions and are extremely unlikely to generalise to real-world driving conditions. Only a limited number of studies have explored the potential of ANNs to predict NO_x emissions from road vehicles (e.g. (Alonso et al., 2006; Kesgin, 2004; Obodeh and Ajuwa, 2009; Atkinson et al., 1998; Johri and Filipi, 2014; Traver et al., 1999). In all these studies, the data available to develop these models was limited to as small number of chassis-dynamometer tests and comprised only a small set of parameters. Subsequently, the network architectures consisted of one hidden layer multilayer perceptron, which is appropriate in the context of a small dataset where the variability is limited. However, a more complex architecture is likely to be required for a larger dataset comprising tests from several different vehicles in real-world driving.

In this study, we will use the large PEMS dataset reported in (O'Driscoll et al., 2016), which comprised 36 Euro 5 and 34 Euro 6 diesel passenger vehicles, allowing to overcome these traditional aforementioned limitations. The present study aims to (i) explore the dataset available by performing a clustering analysis, (ii) use the results of the clustering analysis to improve existing emission models, (iii) develop new models to estimate NO_x emissions from vehicles in real-world driving conditions using ANNs, (iv) compare these new models with existing instantaneous models, and (v) assess and compare the performance of the new models with COPERT5 on a real-case study. Details and data omitted from the main text are included in the Supporting Information (SI) as referenced.

2. Methods

As mentioned earlier, there is no single emission model that meets all criteria of accuracy and simplicity required for policy-making purposes. Current general instantaneous emission models such as look-up tables (Manzie et al.,

2007; Hatzopoulou and Miller, 2010) or non-linear regression models (Panis et al., 2006) are unable to reach the accuracy required (Hagena et al., 2006; Panis et al., 2006) while vehicle-specific models present the desired accuracy but would require every vehicle on the market to be tested in real-world driving conditions, which is currently unrealistic. This observation provided the motivation to: (i) perform a clustering analysis on the dataset in order to group together vehicles presenting similar emitting behaviours in similar operating conditions, which would ultimately help reduce the number of models needed to reach the desired accuracy; (ii) conduct an analysis on the results of the clustering to identify external variables able to explain the results of the clustering; (iii) update the existing instantaneous emission models based on the clusters previously defined; (iv) use Artificial Neural Networks to develop a new instantaneous emission model based on the clusters previously defined; (v) unify the clusters-based models into general models; (vi) quantify the performance and the uncertainties of each model.

This section firstly gives a description of the dataset used in this study in Section 2.1. The methodology for clustering is presented in Section 2.2. Section 2.3 then describes the methodology used to update existing instantaneous emission models and finally a description of the model developed in this study is provided in subsection 2.4.

2.1. Dataset description

The dataset used in this study was supplied by Emission Analytics and was previously described by O'Driscoll et al. (2017). A Portable Emissions Measurements System (PEMS) was used to measure real world driving emissions of 36 Euro 5 and 34 Euro 6 diesel passenger vehicles in the Greater London area between 2012 and 2016. The vehicles were made by 19 different manufacturers (Ford, Nissan, Mercedes, Citroen, Kia, Peugeot, Skoda, Volkswagen, Toyota, Audi, Land Rover, Hyundai, Mitsubitshi, Ssangyong, BMW, Fiat, Volvo, Mazda, Vauxhall). All these vehicles were fitted with a Diesel Oxidation Catalyst (DOC), a Diesel Particulate Filter (DPF) and an Exhaust Gas Recirculation (EGR) system. The Euro 6 vehicles had a mixture of NO_x aftertreatment systems, 7 used EGR only, 16 used EGR + Lean NO_x Trap (LNT) and 11 used EGR + Selective Catalytic Reduction (SCR). We refer the interested reader to O'Driscoll et al. (2017) for a more detailed description of the data collection process.

All the vehicles were tested on the same route and for our analysis we have extracted a geographically consistent segment of the route using the instantaneous coordinates as measured by the embedded GPS. This sub-section was in an urban area and had a length of 10.3 km. All the vehicles in the dataset travelled this journey one to three times and instantaneous speed, coordinates, external temperature and NO_x emissions were recorded. A total of 110 segments were extracted: 54 for Euro 5 vehicles and 56 segments for Euro 6 vehicles. The average dynamic characteristics of the segments of are given in Table 1.

Table 1: Dynamic characteristics of the test segment

Parameter	Value
Duration [min]	27.55 ± 6.56
Distance [km]	10.30 ± 0.095
Average speed [km/h]	23.47 ± 2.50
Kinetic intensity[1] [km^{-1}]	2.12 ± 0.32
% of idle [%]	20.75 ± 5.75
Relative Positive Acceleration (RPA) [m/s^2]	0.19 ± 0.03
95th percentile of the product of RPA and speed	12.49 ± 3.99

[1] Defined in Equation 19 in O'Keefe et al. (2007) as:

$$KI = \frac{\sum_{j=1}^{N-1} positive(1/2(v_{j+1}^2 - v_j^2) + g(h_{j+1} - h_j))}{\sum_{j=1}^{N-1} \overline{v_{j,j+1}^3} \cdot \Delta t_{j,j+1}} \quad (1)$$

where v_j is the instantaneous speed at time j, g is the gravitational constant, h_j is the instantaneous altitude at time j, $\Delta t_{j,j+1}$ is the time step between time t and t+1 and $\overline{v_{j,j+1}^3}$ is defined as:

$$\overline{v_{j,j+1}^3} = \frac{v_{j+1}^3 + v_{j+1}^2 v_j + v_{j+1} v_j^2 + v_j^3}{4} \quad (2)$$

2.2. Clustering

Clustering is the process of grouping a set of objects into clusters so that objects within clusters are similar to each other and dissimilar to objects in other clusters (Park and Jun, 2009). An important step in most clustering algorithms is the selection of a distance metric, which determines how the similarity between two elements is calculated (Velmurugan, 2012). This step is particularly challenging when working with time-series because a standard data point to data point distance metric is not appropriate due to potential variations in speed and length between two time-series that are compared. Dynamic Time Warping (DTW) is an algorithm designed to solve the aforementioned issue and estimate the distance between two time-series that may vary in speed and length. For a detailed description of the DTW algorithm, the interested reader is referred to Section S1 of SI.

In this study, DTW was applied to the instantaneous measured NO_x emissions of the 110 test segments previously extracted. The distance d between two tests was quantified by the integral of the time-warped absolute NO_x emissions over the test route. By applying this procedure to each test pair, an emission distance matrix d_{ij} was constructed; this matrix is square, hollow and symmetric, each row containing the computed distances from one specific test to all the others. Two tests presenting similar emitting behaviours (namely there are similarities in the structure of the time series of NO_x emissions across the route) will consequently be close to each other and similarly far from other tests. The distance matrix can therefore be used to cluster the dataset. The k-means++ algorithm, a popular unsupervised clustering algorithm, was identified as a suitable technique for this task. For a more detailed description

of the k-means++ algorithm, the interested reader is referred to Section S2 of SI. The k-means++ algorithm, using a Euclidean distance evaluation, was applied to the previously defined distance matrix for 2 to 20 clusters. The optimal number of clusters was finally estimated using the Davies-Bouldin and the Calinsky-Harabatz indices. For more details about the the Davies-Bouldin and the Calinsky-Harabatz indices, the interested reader is referred to Section S3 of SI.

The resulting clusters were subsequently analysed with respect to emission factors, vehicle aftertreatment, vehicle engine size and vehicle weight in order to identify a physical variable able to explain the clusters. In addition, the defined clusters were used in the next steps of this study to update existing instantaneous emission models and develop a new emission model. For the sake of completeness, the clustering analysis was also performed for CO and CO_2 for which results are discussed in the main text and figures are available in SI.

2.3. Calibration of existing emission models

Clustering based on the NO_x enables the identification of groups of vehicles presenting similar emitting behaviours. These clusters can then be used to adjust existing emission models whose previous general formulation limited the accuracy of the instantaneous NO_x prediction. In this study, we consider an instantaneous speed and acceleration look-up table (hereafter referred to as the LT model) and the non-linear regression developed by Panis et al. (2006) (hereafter referred to as the NLR model, described in the SI). As the aim of this study is to demonstrate the potential of machine learning techniques applied to vehicle emissions modelling, the analysis has not been conducted on other instantaneous emission models. One look-up table with bins of 2 m/s for speed and 0.5 m/s^2 for acceleration was developed per cluster. The instantaneous mean NO_x emission was computed for each cell of the look-up table. Similarly, the twelve parameters of the NLR model were refitted for each cluster.

In order to evaluate the robustness of the methodology, bootstrapping was conducted by randomly withdrawing 20 tests out of the 110 composing the original dataset before the determination of the coefficients of the emission models (i.e. before computing the look-up table and before refitting the twelve parameters of the non-linear regression). This 20 tests belonged to the clusters containing more than one test (later referred to as main clusters), selected in a way that 20% of the tests in each cluster is used as validation set. For clusters containing only one test, 20% of the test was randomly selected and used as validation set. This step was performed 400 times to ensure that the results were repeatable. Three of the runs (3 x 20 tests) were aggregated together and used to assess the performance of the models with respect to real data. The whole 400 bootstraps were used to study the convergence properties of the models. We make the reader attentive to the fact that the clustering was performed only once, before any further analysis, on all the tests.

2.4. Neural network model description

Artificial Neural Networks were identified as a potentially effective method to create new emission models with improved performance compared to existing emission models. The interested reader is referred to Section S4 of SI for a brief non-exhaustive introduction to Artificial Neural Networks. One important feature that should be highlighted is the absence of a methodology to determine the optimal architecture and optimization parameters for specific problems. Architectures and parameters must therefore be tested in an iterative search to improve the results, until a trade-off between the time spent and the desired accuracy is reached.

The model presented in this study is a Multi-Layer Perceptron (MLP) consisting of five layers of 100, 75, 50, 25 and 1 neurons respectively, with ELUs activation functions and a dropout rate of 0.1. The Root Mean Squared Error (RMSE) was chosen as loss function as the only aim is to match the real emission values. The smaller the RMSE, the better the agreement between the prediction and the real value. The initial learning rate λ was set to 0.0001, which is a standard value. The training algorithm is Adam (Kingma and Ba, 2014), which was demonstrated in a previous study to present improved performance and convergence compared to Adagrad, Stochastic Gradient Descent (SGD), RMSProp and Adadelta (Kingma and Ba, 2014). The model was developed using TensorFlow 1.6 (Abadi et al., 2015) and Python 3.6 (Python Software Foundation).

In order to allow a fair comparison between the MLP model and the LT and NLR models, only the instantaneous speed and acceleration at each time step were used as inputs to the MLP model. The same randomly selected 3 x 20 tests described in Section 2.3 were withdrawn from the training set. These three runs were used to estimate the accuracy and the robustness of the methodology, with respect to the unseen data on the three runs. In order to study the convergence properties of the MLP model, the bootstrapping procedure was performed 400 times, with 20% of the tests in each cluster randomly removed at each run. The resulting mean emission map was compared with the LT and the NLR models emission maps. Since instantaneous emissions at time t are a complex non-linear process that might depend not only on the operating conditions at time t but also on the operating conditions a few seconds before, the impact of including some time history data was explored. Both models were therefore provided with not only the instantaneous speed and instantaneous acceleration at time t but also the instantaneous speed and instantaneous acceleration from time $t - X$ to t, where X belong to [0; 10], resulting in 10 different models. However, and contrary to what was expected, including time history data did not help improving the accuracy of the MLP model (Figure S44 in the SI).

2.5. Dealing with trips that do not belong to a cluster

For tests that do not belong to one of the main clusters, two options were compared for each model (i.e. LT, NLR and MLP). The first option (hereafter referred to as LT_1, NLR_1 and MLP_1) consisted of developing a specific model

for the test. The second option diverges slightly for the MLP model compared to the LT and NLR models. For the LT and NLR models (hereafter referred to as LT_2 and NLR_2), the final predictions were interpolated from the predictions of the clusters using an inverse distance weighting function. The distance between the tests and the centroids of the clusters was computed using the DTW algorithm.

The instantaneous emissions for the LT_2 and NLR_2 models are therefore computed as:

$$E_{\mathrm{N}} = \begin{cases} E_{\mathrm{i}} & \text{if} \quad N \in S_i \\ \dfrac{\sum_{\mathrm{i}} E_{\mathrm{i}} / \overline{d_{\mathrm{iN}}}}{\sum_{\mathrm{i}} 1 / \overline{d_{\mathrm{iN}}}} & \text{if} \quad N \notin \bigcup_{i=1}^{M} S_i \end{cases} \tag{3}$$

where N is the cluster to which the evaluated test belongs to, S_i represents the set vehicles belonging to cluster i, M is the total number of clusters, $\overline{d_{iN}}$ represents the distance between cluster N to the mean of cluster i, and E_i denotes the predicted instantaneous emissions of the cluster i for LT_1 or NLR_1 model.

For the MLP model (hereafter referred to as MLP_2), one unique model for all the main clusters and individual tests not belonging to one of the main cluster was defined. The MLP_2 model was provided with not only the instantaneous speed and acceleration but also the distances of the segment to all the main clusters' centroids. A complete description of the six models and their inputs is available in Table 2.

The accuracy and robustness of the MLP_1 model was compared with the LT_1 and the NLR_1 models. The same approach was conducted to compare the accuracy and robustness of the MLP_2 model with the LT_2 and the NLR_2 models. A cross-comparison of the two types of models was also performed.

2.6. Code availability

The Python implementation of the fast DTW of Salvador and Chan (2007) was used. Instructions to install the package can be found in (Salvator, Stan and Chan, Philip, 2019). The python implementation of the k-means++ algorithm of the scikit-learn package, developed by Pedregosa et al. (2011), was used.

Finally, the MLP model implemented in this work is freely available (Le Cornec, Clemence M. A, 2019). The dataset described in this study is the property of Emissions Analytics and cannot be disclosed. Therefore, to test the code, the interested reader can use the freely available dataset from the UK Department for Transport, (Department for Transport, 2016). We verified our code using the euro6-renault-megane-1461.xlsx file, available in the Euro 6 vehicle data zip file. This file was used both as training and testing dataset to ensure that the code was running properly.

Table 2: Summary of the models used in the study

Model	Description	Inputs	Number of models
LT_1	Updated look-up table model. One look-up table per cluster is developed. If a test does not belong to one of the main cluster, an individual look-up table is developed.	$v(t)$, $a(t)$	Number of main clusters + Number of tests outside the main clusters
LT_2	Updated look-up table model. One look-up table per cluster is developed. If a test does not belong to one of the main cluster, the prediction is interpolated from the main clusters on a distance basis.	$v(t)$, $a(t)$, distances to the main clusters	Number of main clusters
NLR_1	Updated Int Panis regression model. One regression model per cluster is developed. If a test does not belong to one of the main cluster, an individual regression model is developed.	$v(t)$, $a(t)$	Number of main clusters + Number of tests outside the main clusters
NLR_2	Updated Int Panis regression model. One regression model per cluster is developed. If a test does not belong to one of the main cluster, the coefficients of the regression are interpolated from the main clusters on a distance basis.	$v(t)$, $a(t)$, distances to the main clusters	Number of main clusters
MLP_1	Multi-layer perceptron model. One MLP per cluster is developed. If a test does not belong to one of the main cluster, an individual MLP is developed.	$v(t)$, $a(t)$	Number of main clusters + Number of tests outside the main clusters
MLP_2	General multi-layer perceptron model.	$v(t)$, $a(t)$, distances to the main clusters	1

3. Results

The present study was conducted with two main points of focus. The results of the clustering are analysed: (i) analysis of the main clusters (clusters containing more than 3 tests) is provided in terms of driving dynamics and physical characteristics (vehicle weight, engine size, NO_x aftertreatment system) of the vehicles in each cluster; then (ii) special attention is given to the analysis of tests of the same vehicle classified in different clusters. The performance of the six models (LT_1, LT_2, NLR_1, NLR_2, MLP_1 and MLP_2) are analysed, with a focus on the analysis of the importance of spatial variability in emissions.

Quantitative analysis of the driving dynamics was performed using the average speed, the Relative Positive Acceleration (RPA, minimum metric of the dynamic boundary conditions in (EU) 2016/646), the 95^{th} percentile of $v*a_{pos}$ ($v*a_{pos}[95]$, the maximum metric of the dynamic boundary conditions in (EU) 2016/646) and the percentage of idling (percentage of the test where the engine was running while the vehicle was not in motion). Quantitative analysis of the performance of the three studied models is performed using the Normalised Mean Square Error (NMSE, a metric quantifying the overall deviation or scatter between the two sets of results) and Fractional Bias (FB, a metric quantifying mean bias and therefore identifying potential sources of systematic error) (Chang and Hanna, 2004).

3.1. Clustering analysis

3.1.1. Analysis of the main clusters

The best partition is achieved with 17 clusters. Six clusters (hereafter referred to as the main clusters) regroup 97 tests, which represent 88% of the dataset. The eleven other clusters regroup tests of vehicles presenting an emitting behaviour that differs significantly from those of the main clusters. Figure 1 presents the main clusters on the x-axis with respect to (a) the average NO_x emission factor over the test segment, (b) the engine size, (c) the vehicle weight. Each manufacturer is depicted with a different marker, while the aftertreatment technology is displayed in red when the vehicle is equipped with an EGR system only, in blue when the vehicle is equipped with an LNT system in addition to the EGR system and in green when the vehicle is equipped with an SCR system in addition to the EGR system. Finally, the distribution of the NO_x emission factors within each cluster is shown in (d). Figure 1a) shows that the NO_x emission factors inside each cluster vary much less than between the clusters, which is expected since the clustering was carried out and based on NO_x emissions. This is further demonstrated in Figure 1d). Moreover, it can be observed that neither the manufacturer, the aftertreatement system, the engine size nor the vehicle weight fully explain these clusters even if some partial conclusions can be drawn. Clusters 4 and 10 contain the tests with the lowest emission factors, presenting cluster average emissions factors close to the Euro 6 and Euro 5 limits, respectively. With one exception, cluster 4 contains only vehicles equipped with LNT or SCR systems, which is in agreement with

the assertion from Weiss et al. (2012) that vehicles equipped with EGR systems are not able to reach Euro 6 type approval limits in real-driving conditions. Cluster 10 contains four vehicles each equipped with an EGR system only, but these four vehicles have a significantly smaller engine compared to the rest of the vehicles in this cluster. At the other side of the spectrum, vehicles in cluster 8, which present the highest emission factors distribution are vehicles only equipped with an EGR system. A closer inspection of Figure 1 indicates that in general, vehicles equipped with the same aftertreatment system from the same manufacturer (which uses the same family of engines) tend to be grouped together (with one exception, vehicles from Volkswagen equipped with an LNT system are grouped in cluster 4 while most of the vehicles from Volkswagen equipped with an EGR system only are grouped in cluster 5 and 11). The engine optimisation strategies of the different manufacturers are probably one of the main reasons for the discrepancies between the vehicles. Moreover, the overlap of clusters 4 and 10 in terms of NO_x emission factors can be explained by their different emission patterns, as evidenced by Figure 2, which presents the mean measured instantaneous NO_x emissions as a function of the distance travelled. This is in agreement with the grouping performed during the clustering and emphasizes that two vehicles presenting a similar emission factor do not necessarily present the same emitting behaviour.

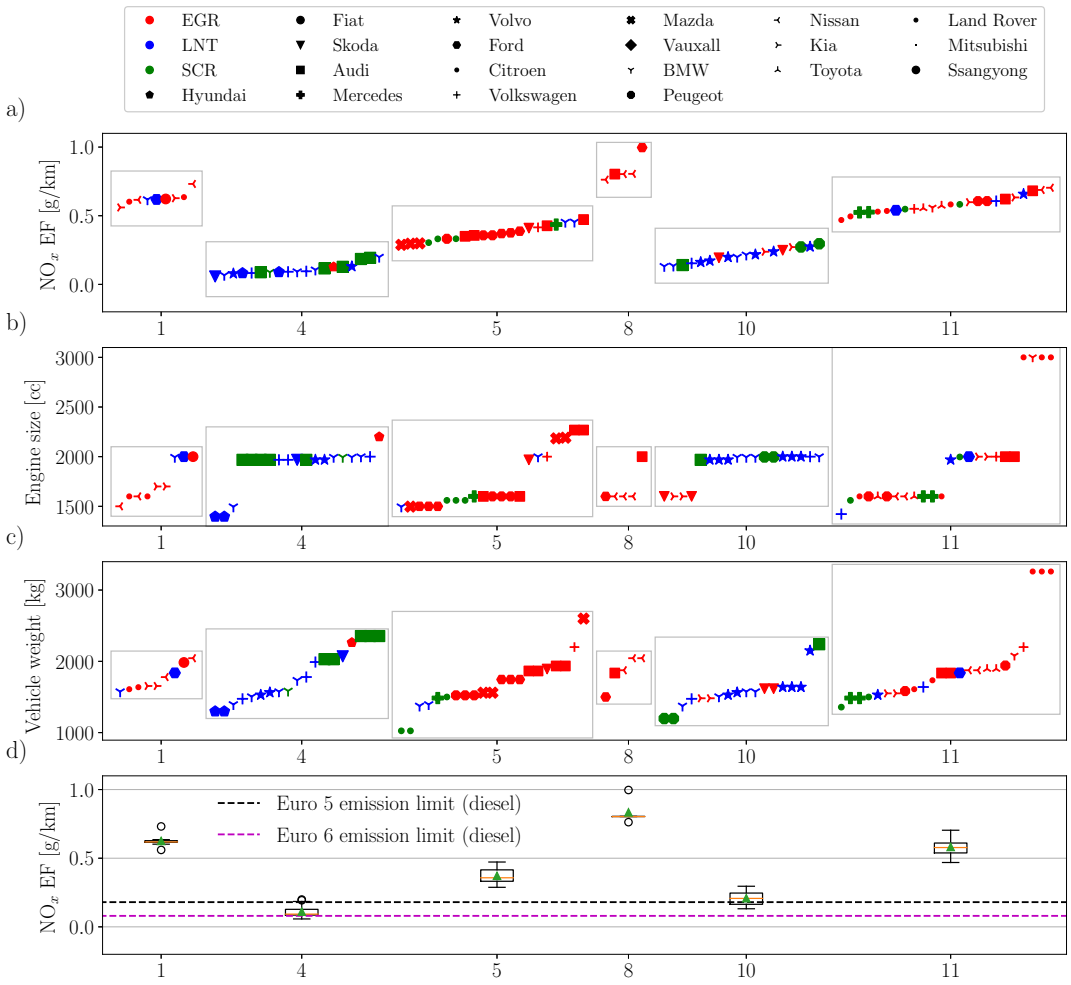


Figure 1: (a) Vehicle manufacturer, NO_x aftertreatment technology and emission factor of each vehicle in the cluster, (b) Vehicle manufacturer, NO_x aftertreatment technology and engine size of each vehicle in the cluster, (c) Vehicle manufacturer, NO_x aftertreatment technology and vehicle weight of each vehicle in the cluster and (d) Boxplots of emission factors per cluster. The whiskers of the boxplots represent the lower and upper quartile respectively. The median of the data is represented by an orange line, while the mean is depicted with a green triangle. Clusters containing less than three tests are not depicted here.

For an analysis of the main clusters in terms of CO, CO_2 and f_{NO_2}, see Section S6.3 of the SI. CO and CO_2 emissions are uncorrelated with vehicular NO_x emissions, which indicates that specific knowledge on CO and CO_2 emissions is unlikely to give any insights on NO_x emissions.

3.1.2. CO and CO_2 clustering

A clustering analysis was also performed on CO and CO_2. For a description of the approach and a detailed analysis of the results, see Section S6.3 of the SI. It is demonstrated that the variability in CO_2 emissions is significantly less marked that the variability in NO_x emissions. From a modelling perspective, performing a clustering analysis in

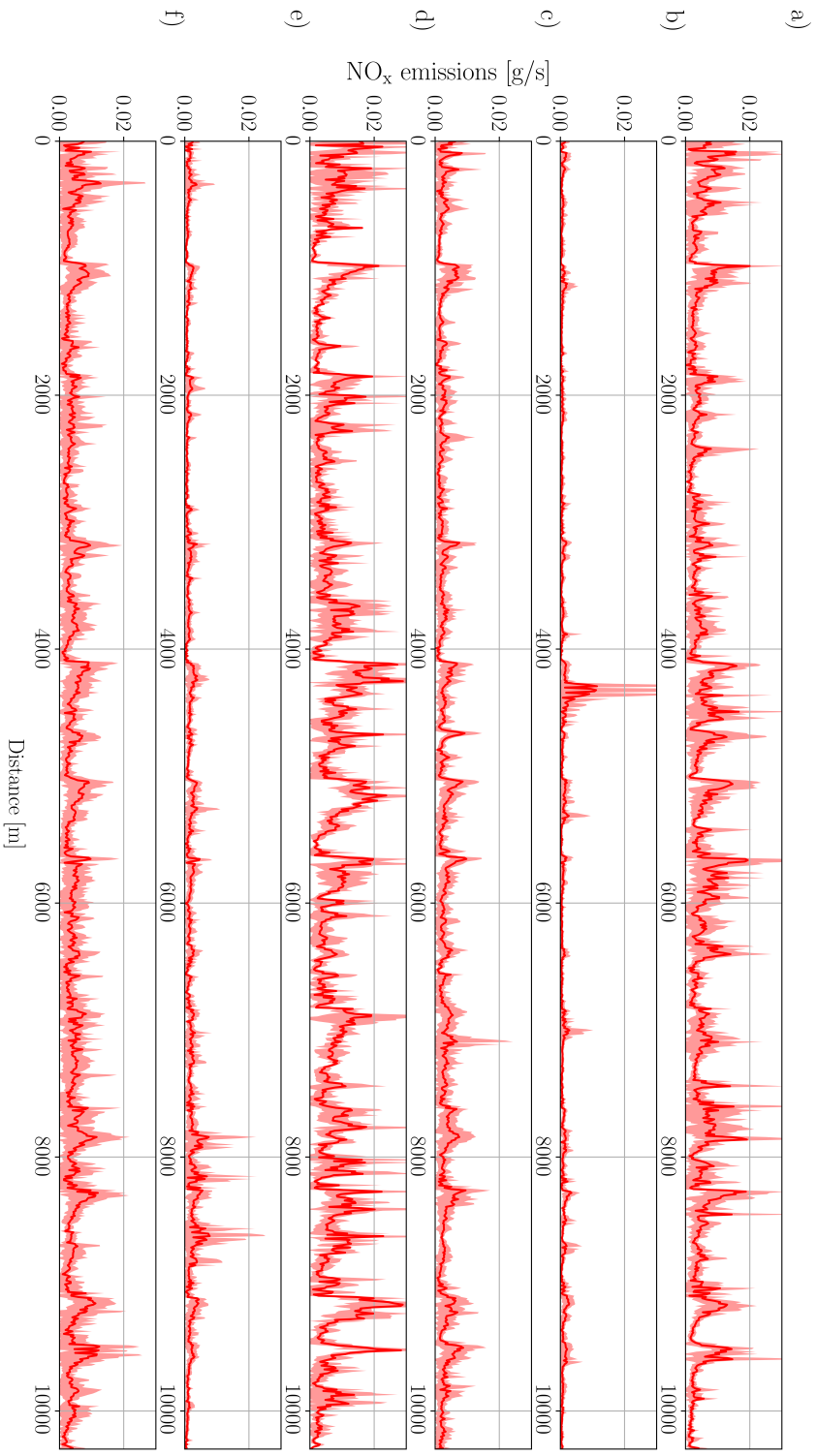


Figure 2: Mean measured instantaneous NO_x emissions as a function of the distance travelled for the main clusters (a) Cluster 1, (b) Cluster 4, (c) Cluster 5, (d) Cluster 8, (e) Cluster 10 and (f) Cluster 11. The standard deviation is depicted by the light red envelope.

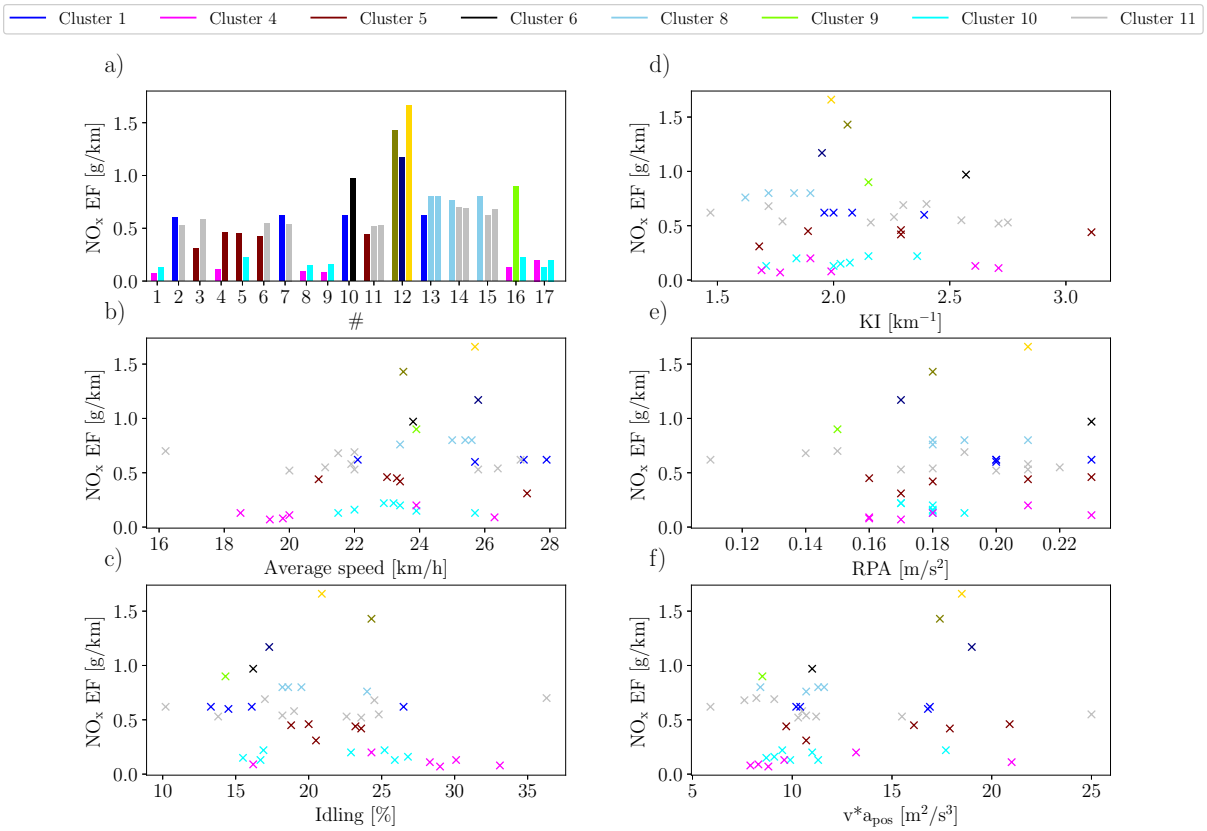


Figure 3: Real-world variation of the NO_x emission factor for the same vehicle when the tests of these vehicles fall in different clusters, as a function of a) the vehicle number, b) the segment average speed, c) the percentage of idling, d) the kinetic intensity, e) the relative positive acceleration and f) the 95th percentile of v*a_{pos}. The cluster number is represented by the colour of the bar or the colour of the point.

order to improve the CO_2 prediction accuracy is likely to be unnecessary. It is also demonstrated that the real-world variability in CO emissions among the vehicles is high, which indicates that, from a modelling perspective, an *a priori* clustering analysis would likely improve the accuracy of the CO prediction.

3.1.3. Intra-vehicle real-world driving variation

Further analysis reveals that for half of the vehicles that were tested more than once (17 out of 33), the tests are classified in different clusters. Table S1 in the SI summarises all the vehicles concerned and gives, in addition to the emission factor and ambient temperature for each of the tests, driving dynamics metrics such as the percentage of idling, the kinetic intensity, the RPA and the 95th percentile of v*a_{pos}. All the tests are compliant with the EU 2016/646 directive for RPA (which means that none of them were driven "too passively") while eight of the tests are not compliant with the EU 2016/646 directive for v*a_{pos}[95] (which means that these tests were driven "too aggressively" and would not be RDE compliant). Most of the tests present somewhat lower RPA compared to the average real world

RPA presented by Pelkmans and Debal (2006) ($\approx$ 0.3 m/s^2 at 25 km/h) and therefore the driving style can globally be described as "smooth". Figure 3.a) (exact numbers available in Table S1 in the SI) demonstrates that the emission factors for the same vehicle can vary by up to a factor of 7 (vehicle 16) while variations up to a factor of 2 are not unusual (vehicles 1 and 9). Ten vehicles out of the 17 present variations of the NO_x emission factor by more than a factor 1.4, while the remaining 7 vehicles present variations of the NO_x emission factor smaller than 30%, which could be explained by the normal real-world variability among the tests (traffic variations in particular). Therefore, the overall variations in emission factors can be significantly large. Figures S3 to S36 in the SI show that for the ten vehicles presenting significant variations in the emission factor, the successive tests present significantly different time-series, while for the seven other vehicles, significant differences are not noticeable at first glance. However, even for those seven vehicles, it can be observed from Table S2 in the SI that the tests classified in the same cluster present more similar mean, standard deviation and coefficient of variation than tests of the same vehicle classified in different clusters.

Table S1 shows that the ambient temperature does not seem to be the explanation for such discrepancies between successive tests of the same vehicle as it does not vary among tests of the same vehicle. Figure 3 shows that variations in percentage of idling (all the other driving dynamic metrics remaining constant) do not present a systematic effect, even for vehicles from the same manufacturer (vehicle 4 presents a lower emission factor for a greater percentage of idling while vehicle 10 presents the exact opposite effect). With the exception of one vehicle (vehicle 14), tests of the same vehicle presenting a lower RPA also present lower emission factors (vehicles 3, 6 and 15). However, a similar trend with $v*a_{pos}[95]$ cannot be found (higher $v*a_{pos}[95]$ does not always result in higher emission factors). Previous studies found contradictory outcomes on the effect of driving dynamics: while Gallus et al. (2017) presented results indicating that aggressive driving leads to an increase in NO_x emissions, Franco et al. (2014) showed that it is difficult to link an increase in NO_x emissions to driving dynamicity. The data presented in Figure 3 shows clearly that no single dynamic metrics is able to explain the variation in the measured NO_x emissions factor. The current dataset does not allow us to isolate with certainty a parameter able to explain these intra-vehicle real-world driving emitting variations and further studies should be conducted in order to understand the underlying phenomena.

In conclusion, the performed clustering analysis enables the grouping together of vehicles presenting similar emitting behaviours. However, it is not possible to find a simple relationship that is able to explain the clusters based only on physical parameters (vehicle weight, engine size, etc.), even though some patterns can be identified. Manufacturer calibration strategies are likely to be able to explain a significant part of the observed variations. Interestingly, a vehicle does not always present the same emitting behaviour, even on the same road. Driving dynamics seem to influence emissions but this is unlikely to be the only factor and further work is necessary to obtain a better understanding of

the underlying phenomena.

3.2. Emission maps for the LT, the NLR and the MLP models

Figure 4 displays a comparison of the emission maps as a function of vehicle speed and acceleration in urban driving conditions for the three studied models. The emission maps for the three models are an average over 400 bootstraps. It should be noticed that the range of speed and acceleration displayed in Figure 4 is significantly larger than typical driving conditions found in the dataset. The white dashed lines represent the fifth and ninety-fifth percentiles of the speed and acceleration distribution in the available dataset (this range is hereafter referred to as 90% dataset range). Predictions outside this range are depicted for comparison purposes but should not be considered as robust or accurate.

It can be observed from Figure 4 that the emission map of the LT_1 model, which is a simple aggregation of the individual datapoints, presents a relatively rough profile characterised by sharp transitions from high to low emissions at relatively similar speed and acceleration. These sharp transitions are due to a lack of data in some bins and although some variability inherent to the vehicle or driving style might occasionally lead to such discrepancies, it is not representative of real world driving emissions (e.g. emission maps from (O'Driscoll et al., 2016)). All the three models highlight that high emissions occur mainly during acceleration events, which is consistent with the literature (Gallus et al., 2017). The emission maps of the NLR_1 model and the MLP_1 model present extremely similar patterns for all the six main clusters, especially in the 90% dataset range. Once trained, the MLP_1 model tends to converge to a formulation that is almost identical to the NLR_1 model. The general accuracy of the 3 models was computed as the weighted average of the joint probability distribution of speed and acceleration:

$$\overline{e} = \int_{a_{\min}}^{a_{\max}} \int_{v_{\min}}^{v_{\max}} e f_{\mathrm{v,a}} \mathrm{d}v\mathrm{d}a \tag{4}$$

where $f_{\mathrm{v,a}}$ is the joint probability distribution of speed and acceleration and e for each cluster is defined as:

$$e = \frac{\sigma_{\mathrm{E}}}{\overline{E}} \tag{5}$$

where σ_{E} is the standard deviation for each bin over the 400 bootstraps and $\overline{E}$ is the weighted average emission for each bin over the 400 bootstraps, over all clusters, based on trips in each cluster.

For the LT_1 model, $\overline{e}$ is equal to 0.80, while for the MLP_1 and NLR_1 model, it is equal to 0.45 and 0.39 respectively. Not surprisingly, $\overline{e}$ is the highest for the LT model that shows the poorest performances and the highest standard deviation per bin, which can be explained by the lack of data in some bins. $\overline{e}$ is slightly higher for the MLP model

compared to the NLR model, which can be explained by the nature of both models: the NLR model is deterministic and therefore converges almost systematically towards the same coefficients, whereas, to prevent the over-fitting of the neural network, a dropout rate of 0.1 was used, introducing some stochasticity to each MLP training run, which increases slightly the final standard deviation and therefore $\overline{e}$.

The near equivalence between the NLR_1 and MLP_1 models suggests that with the limited set of inputs available (instantaneous vehicle speed and acceleration) a better representation cannot be achieved. If only instantaneous speed and acceleration are available and if it is possible to assign the vehicle to one cluster, we recommend therefore opting for simplicity and using the NLR_1 model rather than a neural network. Although further work is necessary to confirm this, neural networks are likely to be an appealing tool if a more detailed set of inputs is available (e.g. engine load, engine speed).

For an analysis of the convergence of the MLP_1 model, see Section S6.5 in SI. It is shown that for all the pixels within the 90% dataset range, the MLP_1 model converges to a unique solution (standard deviation almost equal to zero) within the 90% dataset range. Outside this range, the lack of data makes the model less stable, due to difficulties to converge to a single result.

3.3. Comparison of models performance against the dataset

The general predictive skills of each of the three instantaneous emissions models are assessed in Figure 5 through the comparison of the measured and predicted emission factors for the previously defined 3 x 20 unseen tests. Although it can be argued that the emission factor does not fully represent the emitting behaviour since it does not represent the variability in instantaneous emissions, it is still a good indicator of the average performance of a model. Results for the generalised models (LT_2, NLR_2, MLP_2) are depicted in green, while results for the calibrated models (LT_1, NLR_1, MLP_1) are depicted in magenta.

It can be observed in Figure 5 that the LT_1 model (Figure 5.a) systematically underestimates the emission factors by more than 20%, which is significant. In particular, it does not appear capable of predicting emissions factors over 0.2 g/km, thereby severely underpredicting the emissions of the worst polluters. On the other hand, the NLR_1 model (Figure 5.b) and the MLP_1 model (Figure 5.c) show similar performance with relative errors below 20% (with one exception). Considering that only the vehicle speed and acceleration measured by an embedded GPS were used as inputs to the models, this performance can be considered as satisfactory. However, it is extremely likely that improved performance would be achieved with a wider range of inputs (e.g. engine speed). One interesting feature to notice is the ability of the NLR_1 model to capture the emission factor, implying that meaningfully grouping vehicles together allows surprisingly simple emission models to be used. For accuracy reasons, we recommend to always prefer the NLR_1 and the MLP_1 models to the LT_1 model.

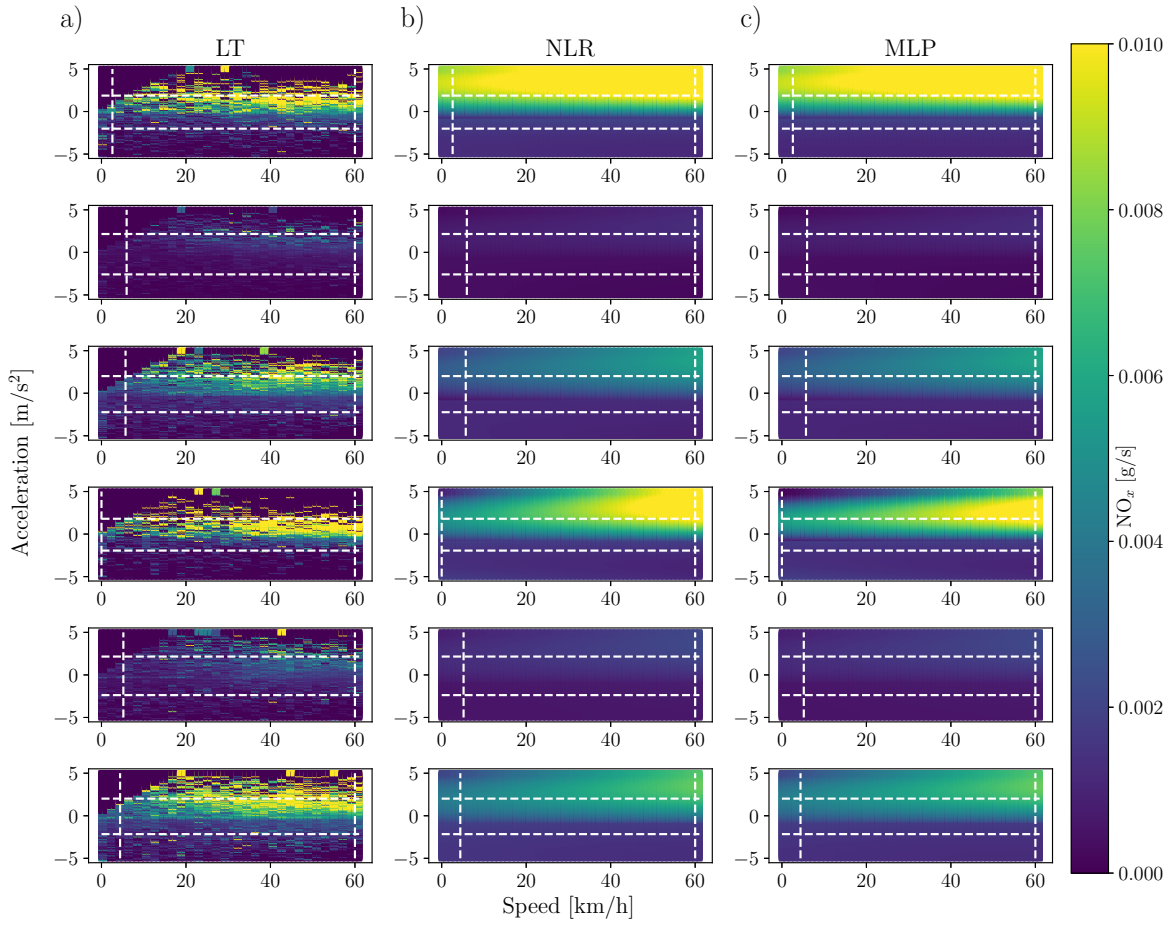


Figure 4: Comparison of the emission maps for the main clusters for the three models; (a) LT_1, (b) NLR_1 and (c) MLP_1 models. White dash lines represent the 5^{th} and 95^{th} percentiles. Clusters are presented in numerical order (1, 4, 5, 8, 10, 11) from the top to the bottom of the graph.

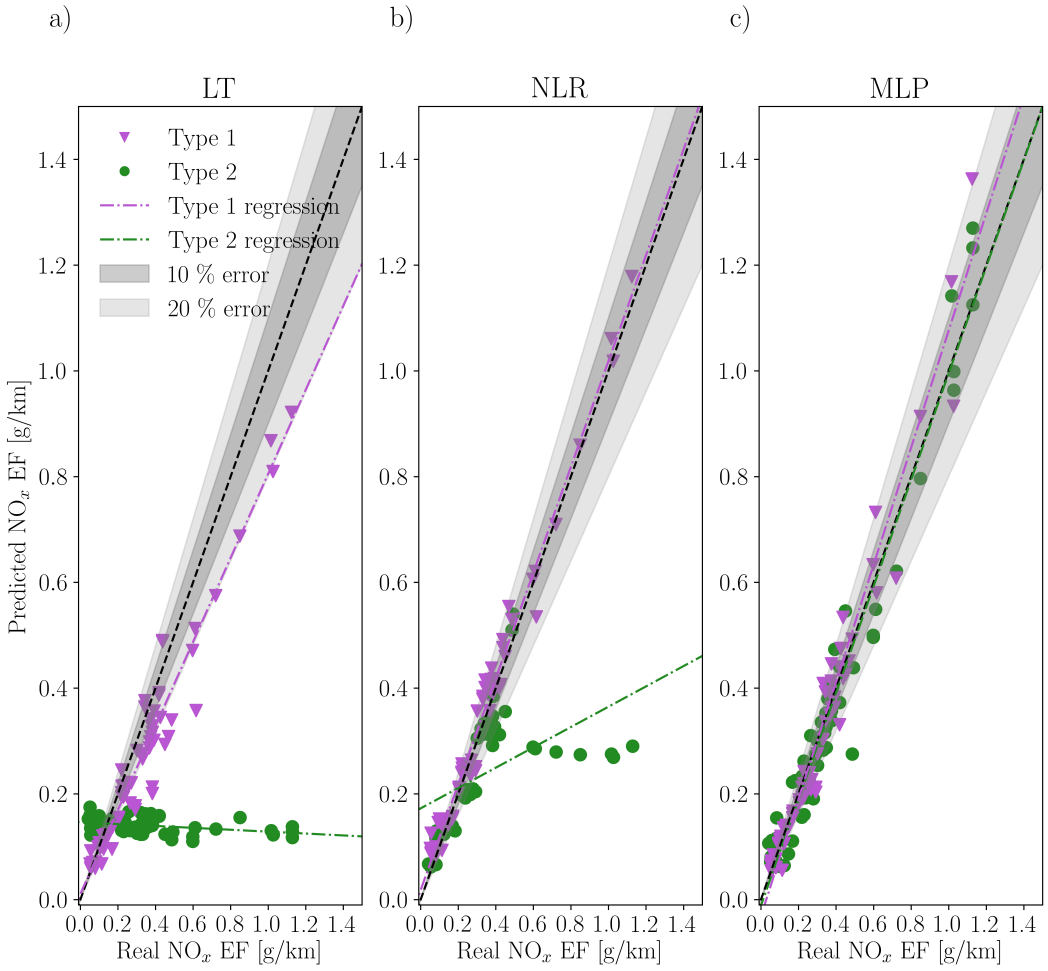


Figure 5: Comparison of the measured and predicted emission factors for three instantaneous emission models, (a) a look-up table, (b) the non-linear regression model and (c) the MLP developed in this study. The results for the LT_1, NLR_1 and MLP_1 models (or clustered models) are depicted with magenta dots while the results for the LT_2, NLR_2 and MLP_2 models (or general models) are depicted with green triangles. The black dashed line represents the identity line.

It is also clearly evident in Figure 5.a that the LT_2 and the NLR_2 models are unable to correctly predict the emission factors for vehicles not in one of the main clusters. In fact, all the tests presenting a relative error significantly larger than 20% in Figure 5.b do not belong to any of the main clusters and the emission factors were extrapolated as presented in Equation 3. This reflects the fact that a simple inverse distance weighting function does not allow the generalisation of the LT_1 and the NLR_1 models. It can therefore be concluded that the relationship between the models of the different clusters is non-linear and cannot be easily inferred. If a test does not fall in any of the main clusters, the development of an individual model should be preferred. In contrast to the LT_2 and the NLR_2 models, the MLP_2 model, depicted in green in Figure 5.c performs well in predicting the emission factor, within the range of a 20%

relative error. This demonstrates that an MLP model, given the distances to the centroids of the main clusters in addition to the instantaneous speed and acceleration, is able to build a complex non-linear generalised model, at the cost of a minimal decrease in accuracy compared to a model directly calibrated for each cluster.

The good performance of the NLR_1, MLP_1 and MLP_2 as well as the poor performance of the LT_1, LT_2 and NLR_2 models described in Figure 5 are also reflected in Figure 6, which shows boxplots of the measured and predicted emissions factors, fractional bias (FB) and Normalised Mean Squared Error (NMSE). It can be observed in Figure 6.a that the LT_1 and LT_2 models present relatively high fractional bias and normalised mean squared errors, reflecting not only the systematic underestimation of the average emissions but also the inability to capture peaks in emission. The NLR_1 and NLR_2 models achieve a higher level of accuracy compared to the LT models. Figure 6.b shows that the fractional bias is close to zero for both cases, demonstrating that the NLR_1 and NLR_2 models are not systematically biased. The higher fractional bias of the NLR_2 model reflects its inability to predict emissions for tests that do not belong to any of the main clusters. Although lower than for the LT_1 and LT_2 models, the normalised mean squared error is around 0.3, which shows the difficulty of capturing the peaks in emission. The MLP_1 and MLP_2 models perform more accurately compared to the LT_1 and LT_2 models. Both the MLP_1 and the MLP_2 models perform well; they do not show any systematic error (as reflected by the fractional bias) but are not able to capture the peaks in emission (as reflected by the normalised mean squared error). None of the models are able to accurately capture all the peaks in emissions. However, it should be noted that those peaks are extremely challenging to capture even when variables like engine speed, gear ratio or EGR rate are known. Vehicle speed and acceleration do not characterise those peaks sufficiently to allow them to be predicted accurately. Nevertheless, these peaks generally account for only a small fraction of the total emissions and although they can be important locally, we consider that not capturing these features represents only a minor drawback. The robustness of the results through the runs is assessed in Table S3 in SI. Results are almost identical through the runs and can therefore be considered as robust.

3.4. Case study and comparison with COPERT5

Finally, the performance of the NLR_1 and MLP_1 models were assessed on a case study, and compared to the COPERT5 model. The evaluation of the three models was conducted on the previously defined 3 x 20 unseen tests to eliminate any potential bias in favour of the NLR_1 and MLP_1 models that would result from using the same dataset to develop and test the model. The NLR_1 and MLP_1 models were applied as described in sections 2.3 and 2.4 to compute the instantaneous NO_x emissions. COPERT5, which is widely used in European countries, is an emission model estimating emissions based on the average-speed along a link. COPERT5 performance increases as road link increases above 400 m (Samaras et al., 2014). In this analysis, road links were taken as uniform 1 km lengths, similarly to O'Driscoll et al. (2016). The average speed at each point was computed as the rolling average speed of

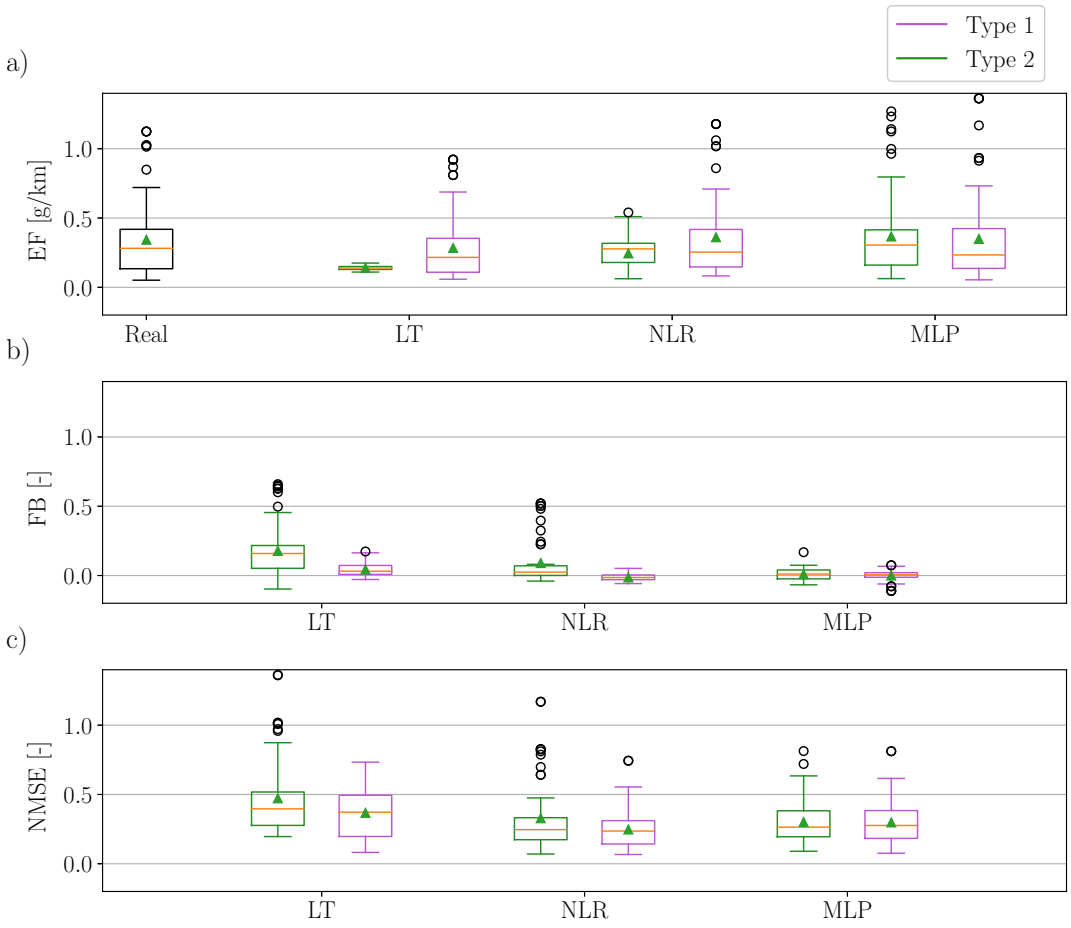


Figure 6: Boxplots of the measured and predicted emission factors (a), boxplots of the Fractional Bias (b) and boxplots of the Normalised Mean Squared Error (c) for the three studied instantaneous models. The results for the LT_1, NLR_1 and MLP_1 models (or clustered models) are depicted in magenta while the results for the LT_2, NLR_2 and MLP_2 models (or general models) are depicted in green. The median of the data is represented by an orange line while the mean is depicted with a green triangle.

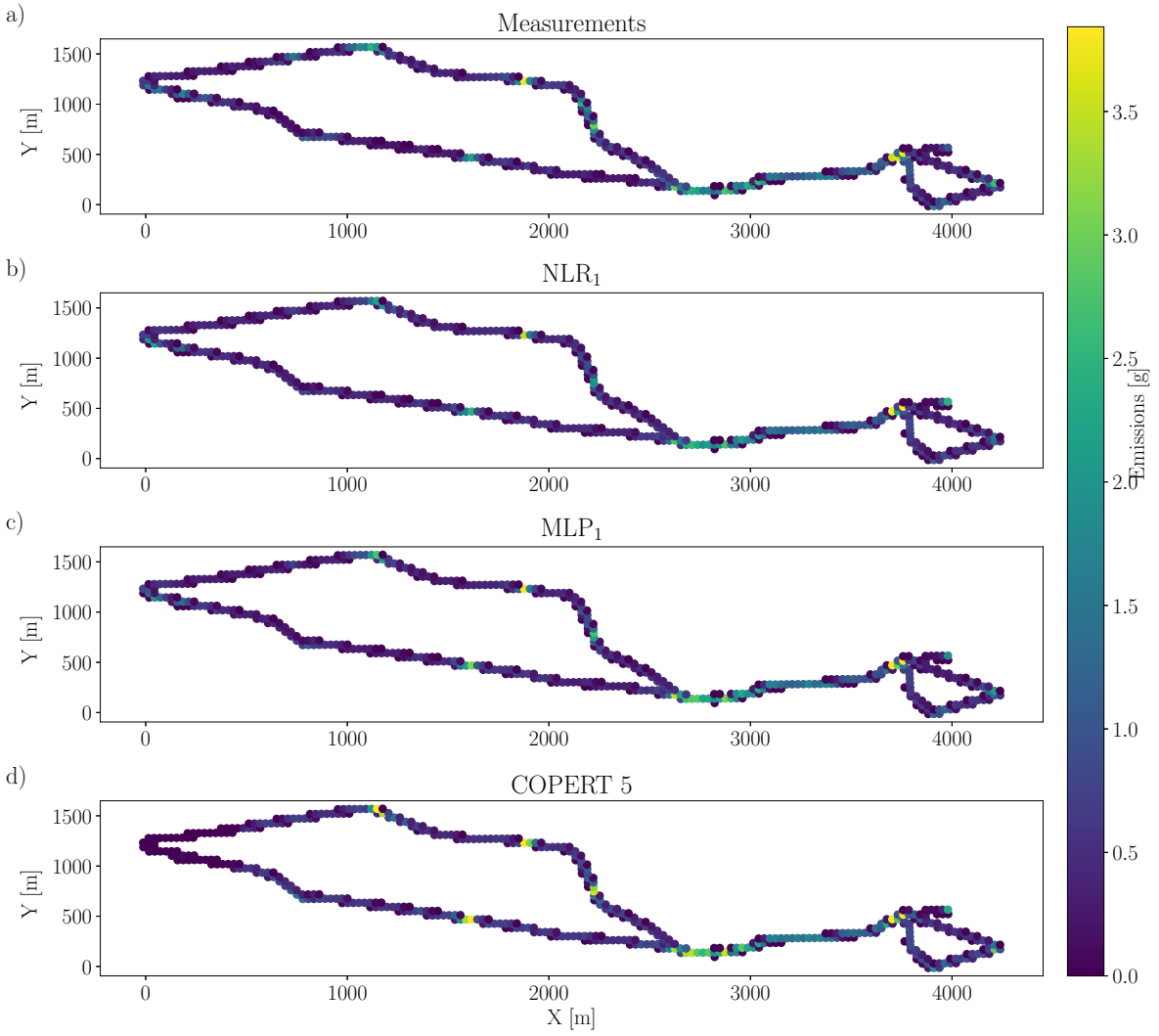


Figure 7: Aggregated emission maps of the unseen tests: (a) are the measurements, (b) the NLR_1 model, (c) the MLP_1 model and (d) the COPERT 5 model.

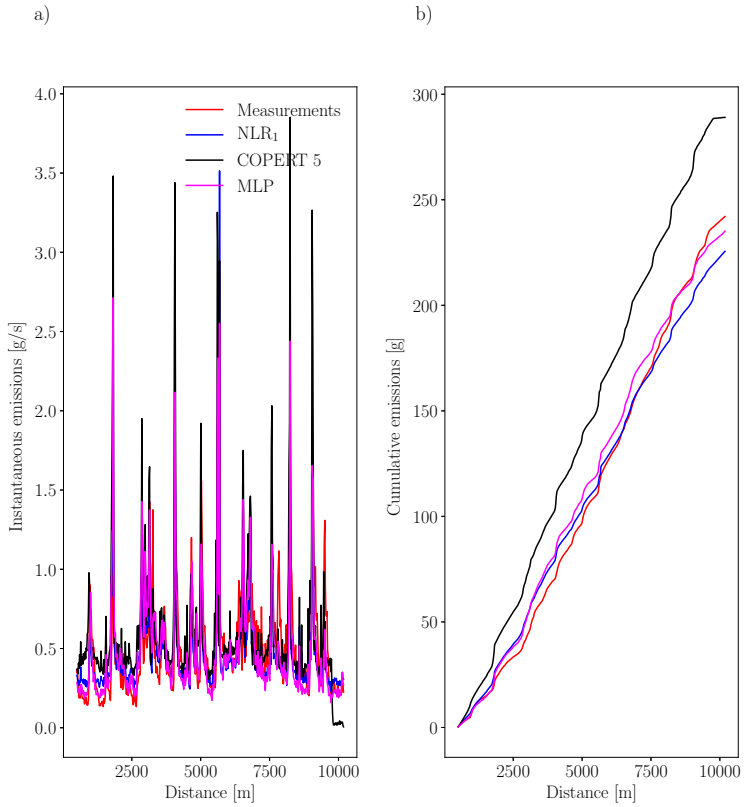


Figure 8: Aggregated measured and modelled emissions as the function of the distance travelled: (a) are the instantaneous emissions and (b) are the cumulative emissions.

the 1 km link centred on that specific data point. We note that this is not the usual way to use COPERT5. However, this is motivated by the wish to compare fairly this average-speed model with the two instantaneous emission models adapted or developed for this study. The first and last 500 meters of each test route were therefore discarded. The specific NO_x emission factor for each data point was computed as described in the (National Atmospheric Emissions Inventory, 2019). The instantaneous emissions were finally computed by multiplying the obtained emission factor by the travelled distance. The measured and estimated instantaneous emissions for the 3 x 20 tests were aggregated at a 10 meter level, based on the computed travelled distance. In other words, all the measured or predicted points between 0 and 10, 10 and 20, ..., meters in the direction of travel were aggregated together.

Figures 7 and 8 present a comparison of the NLR_1, the MLP_1 and COPERT5 models predicted instantaneous emissions aggregated at a 10 meters level for all the 3 x 20 unseen tests, as described earlier. It can be observed in Figures 7 and 8 that the agreement between the measurements and the predictions from both the NLR_1 model and the MLP_1 model is good overall. Both models give very similar predictions, which confirms results presented in section

Table 3: Mean and standard deviation of the FB and NMSE statistical indices for the 60 unseen tests used for the case study

	NLR_1	MLP_1	COPERT5
EF error [%]	-1.90 (s.d. 19.13)	-2.49 (s.d. 15.13)	27.26 (s.d.) 91.59
FB [-]	0.01 (s.d. 0.03)	0.007 (s.d. 0.03)	0.05 (s.d. 0.11)
NMSE [-]	0.27 (s.d. 0.13)	0.31 (s.d. 0.14)	0.38 (s.d. 0.14)

3.2. Although the NLR_1 model misses some of the peaks in emission, the overall trends are well modelled and the relative error in total emission prediction is only of -6.8%. The MLP_1 model also modelled very well the overall trends and underestimates only by 2.9% the total NO_x emissions. The COPERT5 model predictions also show a satisfactory agreement with measurements, despite presenting larger residuals than the NLR_1 and the MLP_1 models (see Section S6.6 in SI). A closer inspection at Figures 7 and 8 also shows the lower accuracy presented by the COPERT5 model compared to the MLP_1 and the NLR_1 models. Overall, the COPERT5 model overestimates the total emissions by 19.4%.

It is worth emphasising that the use of a rolling average to compute the average speed used to evaluate the emissions is not performed during a typical application of COPERT5 but might represent an improvement of the model at an aggregate level. However, as reflected in Table 3, which presents the mean and standard deviation of the relative error in emission factor prediction, the FB and NMSE over the 3 x 20 unseen tests, the individual instantaneous NO_x emissions predicted by the COPERT5 model clearly presents unsatisfactory results compared to the NLR_1 and the MLP_1 models. COPERT5 should therefore not be used to evaluate the performance of a specific vehicle but can provide a reasonable estimate of the total emissions of a fleet over a portion of the network, which corresponds to its normal usage.

4. Conclusion

In this study, machine learning techniques were applied to the large dataset of PEMS measurements in urban driving conditions described in (O'Driscoll et al., 2016). A clustering analysis yielded (i) an update of existing emission models (a look-up table model and a non-linear regression) and (ii) the development of a new emission model (a multi-layer perceptron). A comparison between the updated emission models and the new emission model is provided.

The clustering analysis enabled the grouping of vehicles presenting similar emitting behaviours and allowed us to develop accurate instantaneous NO_x emissions models that sit in-between vehicle-specific and average-fleet models in terms of complexity. Analysis of the clusters reveals that: (i) no obvious internal variable (such as engine size, vehicle weight, vehicle manufacturer or NO_x aftertreatment technology) or external variable (external temperature or driving dynamic) is able to explain the clustering and could be used to predict, *a priori*, the cluster in which a specific

non tested vehicle would fall; (ii) different tests of the same vehicle along the same test route by a trained driver can be assigned to different clusters. When the discrepancies in emission factor between the repeated tests classified in different clusters (7 out of 17 vehicles) was smaller than 20-30%, no fundamental difference between the time-series was noticed and real-world variability constitutes the main explanation for these differences. Higher discrepancies in emission factors were correlated with visible differences, localised on specific parts of the tests, between the time-series of the repeated tests.

Analysis and comparison of the updated emission models with the new MLP model shows that: (i) the clustering works well in producing instantaneous NO_x emissions models with a reduced complexity compared to vehicle-specific models; (ii) a relatively limited set of inputs (vehicle instantaneous speed and acceleration) can accurately predict NO_x emissions, even though the peaks in emissions can be missed; (iii) the NLR_1 and the MLP_1 models present similar performance, and given that the NLR_1 model is less computationally expensive, we recommend it as a preferred choice (coefficients are given in the SI); (iv) for non-clustered vehicles, it is either needed to calibrate a specific model as a linear interpolation from the other clusters was proven to not work or to use directly the MLP_2 model. If a more detailed set of inputs is available (i.e. engine speed, EGR rate), neural networks are likely to present better performances than a simple non-linear regression and are worth exploring. We emphasise that a model is only valid under the conditions for which it was designed and tested, which means that extrapolating outside this range will likely provide wrong or biased estimates. The main challenge in operational use of the presented models is their reliance on a prior clustering analysis. Future work should focus on resolving this issue and is likely to require evidence and knowledge of emissions characteristics of engine families shared across manufacturers. In the future, the outputs of RDE testing, which aims to better represent real-world driving conditions, could be used for to adapt the models.

Advanced emissions models that are accurate and fast enough to be used in real-time will enable a range of intelligent transport systems services including real-time traffic management and signal control optimisation to minimise emissions. This study is the first to demonstrate feasibility and apply machine learning techniques to a large dataset of real-world vehicle NO_x emissions. The new models that we presented rely on simple inputs of vehicle speed and acceleration, which could be extracted from existing sources including traffic cameras and vehicle tracking devices and therefore deployed immediately.

Acknowledgements

This work was supported by Innovate UK (project ref: 103304). Data was provided by Emissions Analytics. Clémence M. A. Le Cornec is supported by a Skempton scholarship at the Department of Civil & Environment Engineering at Imperial College London.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2015). TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Software available from tensorflow.org.

Alonso, J. M., Alvarruiz, F., Desantes, J. M., Hernandez, L., Hernandez, V., and Molto, G. (2006). Combining Neural Networks and Genetic Algorithms to Predict and Reduce Diesel Engine Emissions. *IEEE Transactions on evolutionary computation*.

Andersen, Z. J., Kristiansen, L. C., Andersen, K. K., Olsen, T. S., Hvidberg, M., Jensen, S. S., Ketzel, M., Loft, S., Sorensen, M., Tjonneland, A., Overvad, K., and Raaschou-Nielsen, O. (2012). Stoke and Long-Term Exposure to Outdoor Air Pollution From Nitrogen Dioxide - A cohort study. *Stroke*, 43:320–325.

Atkinson, C. M., Long, T. W., and Hanzevack, E. L. (1998). Virtual Sensing: A Neural Network-based Intelligent Performance and Emissions Prediction System for On-Board Diagnostics and Engine Control. *Electronic Engine Controls 1998: Diagnostics and Controls*.

Chang, J. C. and Hanna, S. R. (2004). Air quality model performance evaluation. *Meteorology and Atmospheric Physics*, 87(1):167–196.

de Lucas, A., Duran, A., Carmona, M., and Lapuerta, M. (2001). Modeling diesel particulate emissions with neural networks. *Fuel*, 80(4):539–548.

Department for Transport (2016). Vehicle emissions testing programme: data and conclusions. https://www.gov.uk/government/publications/vehicle-emissions-testing-programme-conclusions", [Online; accessed 13-January-2020].

EEA (2017). Air quality in Europe — 2017 report. Technical report, European Environment Agency.

European Commission (2018). Testing of emissions from cars. https://europa.eu/rapid/press-release_MEMO-18-3646_en.htm.

Franco, V., Sánchez, F. P., German, J., and Mock, P. (2014). Real-world exhaust emissions from modern diesel cars. *communications*, 49(30):847129–102.

Gallus, J., Kirchner, U., Vogt, R., and Benter, T. (2017). Impact of driving style and road grade on gaseous exhaust emissions of passenger vehicles measured by a Portable Emission Measurement System (PEMS). *Transportation Research Part D: Transport and Environment*, 52:215 – 226.

Grinham, B. and Parry, T. (2018). Vehicle Licensing Statistics: Annual 2017 (Revised). Technical report, Department for Transport - UK government.

Hagena, J. R., Filipi, Z., and Assanis, D. N. (2006). Transient Diesel Emissions: Analysis of Engine Operation During a Tip-In. In *SAE 2006 World Congress and Exhibition*. SAE International.

Hatzopoulou, M. and Miller, E. (2010). Linking an activity-based travel demand model with traffic emission and dispersion models: Transport's contribution to air pollution in Toronto. *Transportation Research Part D: Transport and Environment*, 15(6):315 – 325.

Johri, R. and Filipi, Z. (2014). Neuro-fuzzy model tree approach to virtual sensing of transient diesel soot and NOx emissions. *International Journal of Engine Research*, 15:918–927.

Kesgin, U. (2004). Genetic algorithm and artificial neural network for engine optimisation of efficiency and NOx emission. *Fuel*, 83:885–895.

Kingma, D. P. and Ba, J. (2014). Adam: A Method for Stochastic Optimization. *CoRR*, abs/1412.6980.

Le Cornec, Clemence M. A (2019). Ann.emissions. https://github.com/clemlecornec/ANN.Emissions, [Online; accessed 13-January-2020].

Ligterink, N., Kadijk, G., van Mensch, P., Hausberger, S., and Rexeis, M. (2013). *Investigations and real world emission performance of Euro 6 light-duty vehicles*. Delft: TNO.

Manzie, C., Watson, H., and Halgamuge, S. (2007). Fuel economy improvements for urban driving: Hybrid vs. intelligent vehicles. *Transportation Research Part C: Emerging Technologies*, 15(1):1 – 16.

503 Matzer, C., Hausberger, S., Lipp, S., and Rexeis, M. (2017). A new approach for systematic use of PEMS data in emission simulation. *Journal of*
504 *Earth Sciences and Geotechnical Engineering*, 7(1):33–50.
505 National Atmospheric Emissions Inventory (2019). Emission factors for transport.
506 Ntziachristos, L., Papadimitriou, G., Ligterink, N., and Hausberger, S. (2016). Implications of diesel emissions control failures to emission factors
507 and road transport NOx evolution. *Atmospheric Environment*, 141:542 – 551.
508 Obodeh, O. and Ajuwa, C. I. (2009). Evaluation of Artificial Neural Network Performance in Predicting Diesel Engine NOx Emissions. *European*
509 *Journal of Scientific Research*, 33:642–653.
510 O'Driscoll, R., ApSimon, H. M., Oxley, T., Molden, N., Stettler, M. E., and Thiyagarajah, A. (2016). A Portable Emissions Measurement
511 System (PEMS) study of NOx and primary NO2 emissions from Euro 6 diesel passenger cars and comparison with COPERT emission factors.
512 *Atmospheric Environment*, 145:81–91.
513 O'Driscoll, R., E J Stettler, M., Molden, N., Oxley, T., and M ApSimon, H. (2017). Real world CO2 and NOx emissions from 149 Euro 5 and 6
514 diesel, gasoline and hybrid passenger cars. 621:282–290.
515 O'Keefe, M. P., Simpson, A., Kelly, K. J., and Pedersen, D. S. (2007). Duty cycle characterization and evaluation towards heavy hybrid vehicle
516 applications. Technical report, SAE Technical Paper.
517 Panis, L. I., Broekx, S., and Liu, R. (2006). Modelling instantaneous traffic emission and the influence of traffic speed limits. *Science of the Total*
518 *Environment*, 371(1-3):270–285.
519 Park, H.-S. and Jun, C.-H. (2009). A simple and fast algorithm for K-medoids clustering. *Expert Systems with Applications*, 36(2, Part 2):3336 –
520 3341.
521 Parlak, A., Islamoglu, Y., Yasar, H., and Egrisogut, A. (2006). Application of artificial neural network to predict specific fuel consumption and
522 exhaust temperature for a diesel engine. *Applied Thermal Engineering*, 26(8-9):824–828.
523 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas,
524 J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of*
525 *Machine Learning Research*, 12:2825–2830.
526 Pelkmans, L. and Debal, P. (2006). Comparison of on-road emissions with emissions measured on chassis dynamometer test cycles. *Transportation*
527 *Research Part D: Transport and Environment*, 11(4):233 – 241.
528 Roads Task Force, Technical Note 21 (2012). What is air quality on the road network and how does this vary by road type, location and time of
529 day? Technical report, Transport for London.
530 Salvador, S. and Chan, P. (2007). FastDTW: Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11.5:561
531 – 580.
532 Salvator, Stan and Chan, Philip (2019). fastdtw 0.3.4. https://pypi.org/project/fastdtw/, [Online; accessed 13-January-2020].
533 Samaras, C., Tsokolis, D., Toffolo, S., Garcia-Castro, A., Vock, C., Ntziachristos, L., and Samaras, Z. (2014). Limits of applicability of COPERT
534 model to short links and congested conditions. In *20th International Transport and Air Pollution Conference*.
535 Smit, R., Ntziachristos, L., and Boulter, P. (2010). Validation of road vehicle and traffic emission models - A review and meta-analysis. *Atmospheric*
536 *Environment*, 44:2943–2953.
537 Thompson, G. J., Atkinson, C. M., Clark, N. N., Long, T. W., and Hanzevack, E. (2000). Neural Network Modelling of the Emissions and Perfor-
538 mance of a Heavy-Duty Diesel Engine. *Proceedings of The Institution of Mechanical Engineers Part D - Journal of Automobile Engineering*,
539 214:111–126.
540 Traver, M. L., Atkinson, R. J., and Atkinson, C. M. (1999). Neural Network-Based Diesel Engine Emissions Prediction Using In-Cylinder
541 Combustion Pressure. *Emission Formation Processes in SI and Diesel Engines*.

Velmurugan, T. (2012). Efficiency of k-means and k-medoids algorithms for clustering arbitrary data points. *Int. J. Computer Technology & Applications*, 3(5):1758–1764.

Vineis, P., Gerard, H., Michal, K., Federica, V.-T., Fabrizio, V., Luisa, A., Herman, A., Alison, D., Seymour, G., Pierre, H., Christian, M., Giuseppe, M., Kim, O., Ole, R.-N., Francoise, C.-C., Jacob, L., Heiner, B., Antonia, T., Domenico, P., Marco, P., Vittorio, K., Rosario, T., Salvatore, P., Bas, B.-D.-M. H., H., P. P., Eylin, L. E., A., G. C., Carmen, M., Miren, D., Aurelio, B., Lluis, C., Ramon, Q. J., Goran, B., Bertil, F., E., D. N., J., K. T., Rodolfo, S., Rudolf, K., and Elio, R. (2006). Air pollution and risk of lung cancer in a prospective study in Europe. *International Journal of Cancer*, 119(1):169–174.

Weiss, M., Bonnel, P., Kühlwein, J., Provenza, A., Lambrecht, U., Alessandrini, S., Carriero, M., Colombo, R., Forni, F., Lanappe, G., Lijour, P. L., Manfredi, U., Montigny, F., and Sculati, M. (2012). Will Euro 6 reduce the NOx emissions of new diesel cars? – Insights from on-road tests with Portable Emissions Measurement Systems (PEMS). *Atmospheric Environment*, 62:657 – 665.

Appendix A. Supplementary Material

Supplementary material can be found attached to this article.