AI-Based Virtual Sensing of Gaseous Pollutant Emissions at the Tailpipe of a High-Performance Vehicle

AI-based Virtual Sensing of Gaseous Pollutant Emissions at the Tailpipe of a High-performance Vehicle

Authors: E. Giovannardi, A. Brusa, B. Petrone, N. Cavina, R. Tonelli, I. Kitsopanidis

Abstract

This scientific publication presents the application of artificial intelligence (AI) techniques as a virtual sensor for tailpipe emissions of CO, NOx, and HC in a high-performance vehicle. The study aims to address critical challenges faced in real industrial applications, including signal alignment and signal dynamics management. A comprehensive pre-processing pipeline is proposed to tackle these issues, and a Light Gradient Boosting Machine (LightGBM) model is employed to estimate emissions during Real Driving Emissions (RDE)-like cycles. The research compares two modeling approaches: a direct model and a two-stage model with separate models for the engine and the aftertreatment. The findings suggest that the direct model strikes the best balance between simplicity and accuracy. Furthermore, the study investigates two sensor setups: a standard configuration and an optimized one which incorporates an additional lambda probe in the exhaust line after the main catalyst. The results indicate a significant enhancement in performance for NOx and CO estimations with the introduction of the third lambda probe, while HC results remain relatively unchanged. Additionally, the AI model is tested on two different Electronic Control Unit (ECU) software calibrations, yielding excellent results in both cases. This suggests that machine learning models are robust to control software variation and can be used to optimize software calibrations in a virtual environment, reducing the reliance on extensive experimental testing. Moreover, the AI model's performance demonstrates compatibility with real-time implementation. In conclusion, this work establishes the viability and efficiency of artificial intelligence techniques in accurately estimating tailpipe emissions from an engine in an industrial context. The study showcases the potential for AI to contribute to emission estimation and optimization processes, offering a promising pathway for an innovative industrial practice.

Keywords: virtual sensors, machine learning, artificial intelligence, emission prediction, industrial application, tailpipe emissions

1) Introduction

Engine manufacturers are being faced with a challenging task under the EU7 legislation, which mandates the respect of emissions limits under unusual and critical driving conditions, including instances of aggressive driving styles. Moreover, it introduces requirements for OEMs (Original Equipment Manufacturer) to achieve real-time monitoring of pollutant emissions [1]. However, a major obstacle is the unavailability of direct sensors that can measure emissions while the vehicle is in operation, as they are usually too heavy and voluminous for use in uncontrolled environments. Additionally, the homologation process requires extensive testing that covers nearly every possible condition that may be encountered during the engine's future employment. These emissions tests are costly, time-consuming, and challenging to reproduce accurately [2], [3].

Engine manufacturers must therefore seek innovative solutions to optimize powertrain development and comply with the EU7 regulations. One promising approach is the use of virtual sensors [4], which can indirectly estimate pollutant emissions leveraging signals already available to the engine control unit (ECU). To this end, data-driven techniques, such as machine learning and deep learning, are used to simulate the complex thermal-fluid dynamics and chemical reactions within an engine. Although an analytical model could theoretically represent such a system, the complexity of thermochemical reactions makes this approach infeasible within a reasonable timeframe [5].

Similarly, after-treatment systems can be modeled using CFD and chemical kinetics to simulate the chemical reactions that take place within the system [6]. However, these models require detailed information about system components and are computationally expensive [7]–[9].

Considering this, the use of data-driven techniques aimed at virtual sensing is a viable approach to minimize the need for physical tests. This, in turn, curtails both the duration and expenses involved in the developmental phase [10]. Additionally, such techniques make possible to measure emissions in real-time or offline in virtual environments across various driving conditions that may not be feasible to replicate using conventional measurement systems, such as PEMS, or in standard laboratory facilities [11].

Previous research has explored solutions for virtual sensing of pollutant emissions [12], [13]. The primary focus of those works has been on predicting emissions at the exit of the engine and usually considering compression ignited engines [14]–[16]. This has proven useful for identifying critical maneuvers that result in increased emissions and improving engine calibration in these specific cases. The authors of this study have also investigated various data-driven regressors for this purpose on NOx [17], CO and HC [18] at the engine-out section. Also, particulate matter and particulate number is a concern for the legislation and some studies have already proved the advantages of adopting virtual sensing for PN prediction at the engine-out section [19].

However, the virtual sensing of gaseous emissions such as NOx, HC and CO at the tailpipe section is of particular interest for homologation purposes [20]. As such, this study aims to address this challenge by presenting a methodology for virtually sensing emissions at the tailpipe exit using information from the ECU's sensors and actuations, exploiting a Light Gradient Boosting Machine-based regressor [21].

2) Research Structure

The authors introduce a data-driven modeling workflow for the real-time prediction of tailpipe-out gaseous emissions. The methodology was applied and validated for predicting CO, NOx, and HC on experimental driving cycles.

Sections 3-5 provide a detailed description of the experimental setup, the sensors installed in the exhaust line for the tests as well as the experimental campaign performed to generate the dataset for the study.

Sections 6 and 7 are focused on the main criticalities of the experimental data, such as the problem of signals synchronization, and propose a pre-processing pipeline to address those issues, including a synchronization procedure based on local lagged cross-correlation and a sliding window for features engineering.

Section 8 presents a discussion on the results about: the comparison between direct and two-stage modeling, the importance of the third lambda probe as additional features, and the model's robustness to changes in ECU software calibration.

The concluding Section 9 outlines the observations regarding applicability and results, with particular emphasis on the main advantages highlighted in the study.

3) Experimental Setup

The study presented is based on data coming from experimental tests on a vehicle equipped with a high-performance naturally aspirated V12 engine, whose main specifications are reported in Table I.

Table I. Engine main specifications

Engine specifications	
Engine Type	V12
Displacement [cc]	6495.6 cc
Aspiration	Naturally Aspirated
Combustion System	DI Spark-ignition
Number of cylinders	12 (6 per bank)
Valves per cyl [#]	4 (2 int + 2 exh)
Bore x Stroke [mm]	94.0 x 78.0

A simplified scheme of the system under study is reported in Figure 1.

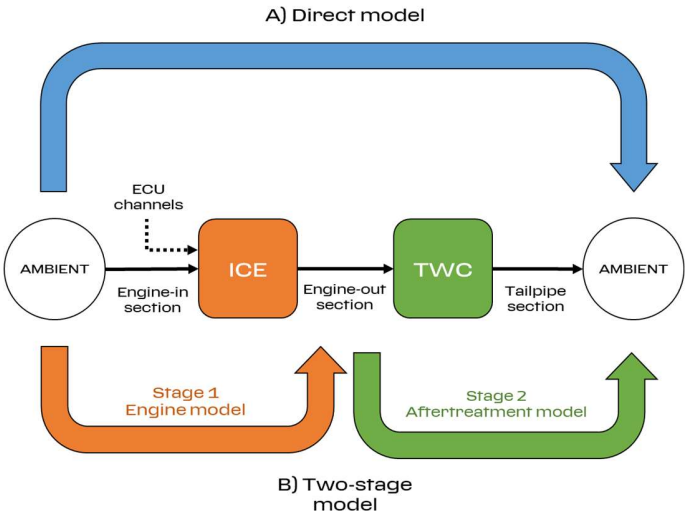

Figure 1. Scheme of the system under study, with direct and two-stage modeling approaches

The engine-in section represents the section upstream of the engine, here intended in a more general way as the ensemble of operating conditions, actuations and sensors information that are collected by the ECU.
The engine-out section represents the connection between the engine and the after-treatment system where emissions concentration achieves its maximum, before being partially eliminated by the catalyst.
The tailpipe section represents the interface between the exhaust pipes and the external environment where the Portable Emissions Measurement Systems (PEMS), as well as any other measurement system, are installed to measure the emissions for homologation purposes.
The emissions at the tailpipe section can be modeled with two different approaches: direct model and two-stage model. Both approaches are evaluated to find the best one, and further details about the comparison and the type of models are provided in Section 8.

4) Exhaust line setup

Exhaust pipeline

The exhaust system configuration is shown in the scheme of Figure 2 featuring two catalytic converters: the pre-cat and the main cat. These converters are responsible for reducing gaseous emissions in the exhaust gases.

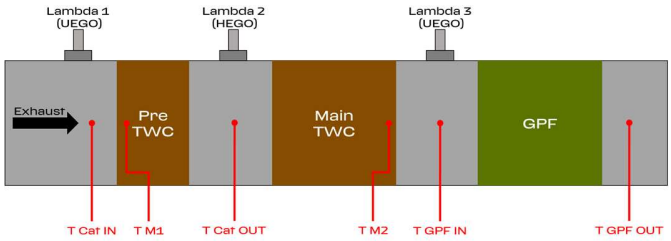

Figure 2. Functional layout and sensors of the exhaust line

Six thermocouples have been installed at various sections of the line. They are positioned upstream and downstream of the catalytic converters and GPF (Gasoline Particulate Filter), as well as at the exhaust sampling locations to measure emissions accurately.
Regarding sensor installation, the standard setup includes two lambda probes. The first probe, a UEGO (Universal Exhaust Gas Oxygen) type, is placed after the engine to monitor the air-fuel ratio. The second probe, an HEGO (Heated Exhaust Gas Oxygen) type, is located downstream of the pre-cat to diagnose the Oxygen Storage Capacity (OSC) of the catalyst. The optimized setup includes an additional third lambda probe, which is a UEGO type. It is situated downstream of the main catalyst to measure its oxygen content, allowing for monitoring of its saturation level and thus its conversion efficiency.
This study aims to compare the accuracy of the emission models with and without the third lambda probe as an input. Furthermore, using the Feature Importance Permutation (FIP) algorithm [22], the significance of this additional probe in relation to the model's performance is quantified.

Catalytic converter

The catalytic converter is a crucial component of automobile exhaust systems due to its ability to convert toxic exhaust gases into less harmful substances. To optimize catalyst conversion efficiency, the air fuel ratio must be maintained near the stoichiometric point [23]. While an excess of oxygen can enhance CO and HC oxidation, it can also impair the catalytic converter's ability to reduce NOx emissions [24] and vice-versa.
In this regard, lambda sensors (oxygen sensors) play a crucial role in monitoring critical thresholds for optimal catalyst working conditions. However, the empirical behavior of the catalytic converter indicates that its efficiency at preventing emission peaks is related to the exhaust level in previous time steps. This historical information can be leveraged to estimate the magnitude and duration of emission peaks. The catalyst's temperature significantly affects its chemical efficiency, with low temperatures during the cold phase resulting in reduced efficiency. The gas flow speed also impacts catalyst efficiency, as the catalyst requires a specific amount of time for each reaction. When the exhaust flow approaches the maximum capacity of the chamber, part of the exhausts may bypass the catalyst, reducing its effectiveness.

5) Experimental Campaign

The activity was conducted using a set of driving cycles with an RDE-like speed profile reproduced on a test bench. The speed profile covered the typical operating ranges of an RDE cycle (urban, extra-urban, and motorway) and was carried out in the vicinity of the Maranello plant, covering a route lasting approximately 30 minutes. The recorded speed track was subsequently reproduced in the Ferrari emission laboratories on a roller bench using a prototype vehicle. Ten driving cycles were conducted using two different ECU software: a first-attempt base calibration, which will be referred to as software A, and an optimized calibration for emissions reduction, which will be referred to as software B. In order to produce a fair final evaluation, two cycles (one software A and one software B) were withheld from the training set, resulting in the final dataset reported in Table II.

Table II. Experimental dataset

	Training cycles	Test cycles
Software A	3 (90 min)	1 (30 min)
Software B	5 (150 min)	1 (30 min)

6a) System Insight and Issues

The first premise to highlight is that emissions (target output of the model) are measured with a dedicated system, that can be a PEMS or a laboratory analyzer. Some ECU channels (inputs of the model) come from different kinds of sensors which retrieve data about intermediate stages of the system, such as the temperature of the gases or the oxygen concentration in the exhausts before and after the catalyst. These pieces of information, merged with values derived from direct actuation, compose a set of many time-series.

Several issues are related to this system:

- Lack of synchronization between emission measurement systems and ECU actuations, introducing a temporal offset in the whole experiment.
- Mass transport that affects the time delay between the measured emissions and the operating conditions controlled by the ECU. This effect is much more evident when the target is to measure emissions at the exit of the tailpipe, because the air first and the exhausts then have to pass through the engine, the exhaust manifold, the pipes until the catalyst. The whole process duration is estimated to be between 2 and 5 seconds.
- The effect of mass transport is not constant, because when engine speed and load change, the air speed changes as well and, consequently, the shift due to the transport of mass varies.
- The effect of mass transport affects especially the output emissions, but also all the inputs that are measured by sensors installed in the path from the engine to the tailpipe (as lambda probes).

For the aforementioned reasons, the alignment between tailpipe emissions and ECU channels is not a trivial task and a part of the pre-processing must be dedicated to it [25], [26].

6b) Alignment technique

It can be deduced that achieving a faithful representation of the examined system, with a temporal correspondence between input signals (ECU channels) and the resulting output (measured emissions), necessitates a two-step realignment process. First, a realignment is required among the different ECU channels, followed by a subsequent realignment between these channels and the measured emissions. The initial realignment pertains to various sensors (e.g., lambda probes and thermocouples) positioned at different locations in the exhaust line, thus capturing system information under distinct spatiotemporal conditions. As an example, Figure 3 shows the misalignment of the three lambda probes on the lean transition that occurs during a cut-off maneuver. The lean transition is read by the probes with a delay that is related to their position along the exhaust line, because of the mass transport.

The second realignment addresses the temporal delay that invariably influences the emission measurements in relation to the corresponding input operating conditions. Since such a delay is much more significant than the previous one, for the purpose of this study the misalignment among the ECU channels is assumed to be negligible compared to the misalignment between the measured emissions and the ECU channels.

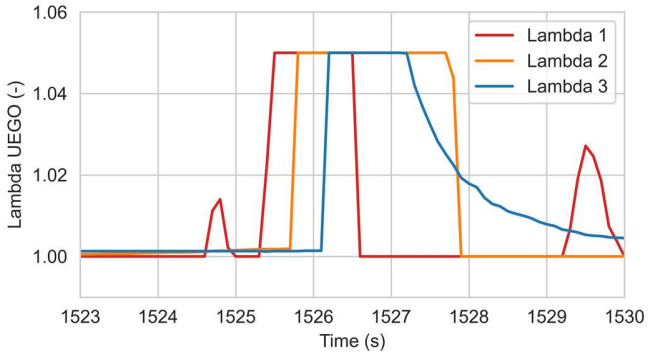

Figure 3. Misalignment on the three lambda probes in the lean transition during a cut-off

To address this issue, an alignment technique is developed that correlates the emissions with the third lambda probe.

For exhaust gas analysis, a shared analyzer system is employed across multiple test cells within the laboratory. The extraction of exhaust gases for emission measurement occurs downstream of the third lambda probe. These gases are conveyed to the analyzers through a duct whose length depends on the distance between the test cell and the analyzers themselves. Accounting for these considerations, the proposed alignment technique utilizes the third lambda probe channels as time reference for defining the time shift to apply to the emissions channels.

Among the ECU channels, the third lambda probe is chosen as the reference due to its proximity to the emission extraction point. Additionally, the third lambda probe, positioned immediately downstream of the primary catalytic converter, provides an indirect measure of the converter's filling state and, consequently, its emission reduction efficiency. Notably, the signal captured by the third lambda probe exhibits a strong correlation with emissions, particularly regarding CO and NOx. Under standard operating conditions, the aftertreatment system, specifically the TWC, has the capacity to absorb or release oxygen to facilitate emission reduction. Consequently, the third lambda probe detects a signal indicative of the stoichiometric ratio. However, if the catalytic system becomes fully saturated with O2, such as during a series of cut-offs, it loses the ability to absorb oxygen, resulting in a decrease in NOx reduction efficiency. In this scenario, unabsorbed excess air (and oxygen) from combustion passes through the catalyst and is detected by the third lambda probe, leading

to readings indicating a lean mixture. Conversely, when the catalyst is completely empty of O2, the third lambda probe detects a rich mixture, and due to the absence of oxidation reactions, an increase in CO emissions is observed in the exhaust. Consequently, the third lambda probe is particularly correlated to CO and NOx peaks (Figure 4).
To achieve alignment, a technique based on local lagged cross-correlation between the target channel and the third lambda probe has been developed. As an example, the methodology applied to NOx is presented.

The method adopted aims to leverage the information contained in these individual critical events to achieve the most robust alignment possible. As previously emphasized, a peak in emissions is typically determined by a deviation (either increasing or decreasing) from the unitary value read by the third lambda probe.

The alignment algorithm consists of the following steps:

1. Peaks detection: peaks in emissions and their corresponding peaks on the third lambda probe, meeting specific prominence and maximum amplitude criteria, are identified.
2. Window application: a 10-second window centered around the maximum value of the lambda peak is constructed.
3. Optimal local time-shift calculation: the lagged cross-correlation algorithm is applied with variable shift applied to the emissions within the window. The shift that maximizes the Pearson coefficient [27] between the lambda signal and the target signal is selected as optimal. The process is repeated, and the optimal local shifts are found for each peak.
4. Optimal global time-shift calculation: each peak exhibits a distinct local optimal shift depending on the instantaneous boundary conditions, such as mass transport, exhaust gas flow rate, and flow velocity at the exhaust. A weighted average of the local shifts relative to each peak is calculated, with weights proportional to the amplitude of the peaks.

This approach assigns greater importance to higher peaks and yields a unique shift value to align the target signal along the entire driving cycle. A scheme of the alignment procedure applied to NOx emissions is reported in Figure 4.

It is important to note that the local lagged cross-correlation method does not aim to accurately represent the physical phenomenon of misalignment between the ECU channels and the emission trace. Instead, it proposes a mathematical criterion that allows to align the emission target signal on various test scenarios.

Among the simplifications introduced by the proposed methodology, the following are noteworthy:

- A constant delay in emissions is assumed throughout the entire test duration.
- The effect of operating conditions on mass transport and temporal misalignment of the measured channels is considered negligible.
- The measurement delay between lambda probes in different sections of the exhaust line and ECU actions is considered negligible.
- The proposed method requires that at least one peak of emissions is detected during the driving cycle.

However, the obtained results validate the robustness of this alignment procedure, and the simplifications introduced can be deemed acceptable.

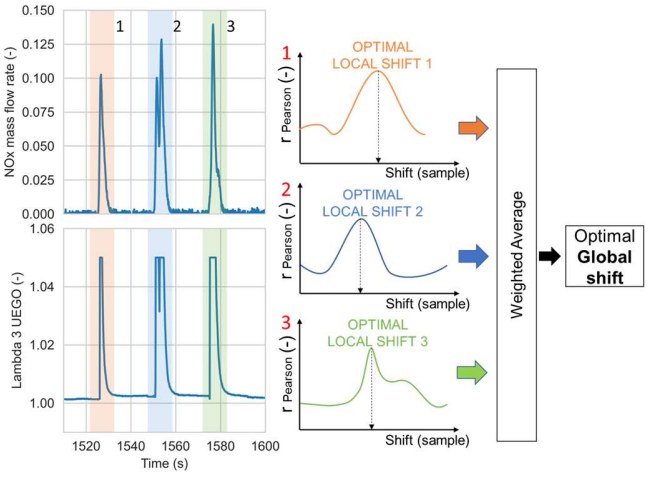

Figure 4. Synchronization procedure with local lagged cross-correlation between Lambda 3 and NOx

7) Features Engineering

Since the acquisition frequency of the ECU channels can vary, all the channels have been resampled to 10 Hz, chosen as a uniform sampling frequency. The complete set of features used as input for the model is reported in Table III.

Table III. List of input features available

Features	Description
Engine speed	-
Engine load	-
P inj	Injection pressure
T inj	Injection duration
SA	Spark Advance
IVO	Intake Valve Opening
EVC	Exhaust Valve Closing
Lambda 1 (UEGO)	1st lambda probe, downstream of the engine
Lambda 2 (HEGO)	2nd lambda probe, downstream of the pre-cat
Lambda 2 (UEGO)	2nd lambda probe, linearized mathematically
Lambda 3 (UEGO)	3rd lambda probe, downstream of the main cat
T Water	Engine coolant temperature
T Cat IN	Exhausts temperature before the pre-cat
T Cat OUT	Exhausts temperature after the pre-cat
T GPF OUT	Exhausts temperature after the GPF
T M1	Exhausts temperature inside the pre-cat
T M2	Exhausts temperature inside the main cat

The LightGBM model utilizes those signals as its input features. They comprise engine speed, and load and engine actuations, including spark advance, injection timing, injection pressure, and valves opening and closing angles. Additionally, the model considers measurements of the gas temperature in different sections of the exhaust line, the coolant temperature, and the three lambda sensors.
The second lambda sensor in the system is a HEGO sensor. However, by leveraging the knowledge of its characteristic curve across various temperatures, it is possible to compute the linear value of the lambda ratio. It is worth noting that this calculation is more accurate the more the lambda value is close to one, given that HEGO sensors characteristic tends to saturate for lambda values significantly different from the unity.

An analysis based on the Feature Importance Permutation (FIP) algorithm has been made to highlight the most relevant features for the prediction of each pollutant species, with a particular focus on the importance of the third lambda probe.
To account for the temporal dependence of the output values, a backward sliding window of fixed width has been applied to the input features. The width of the sliding window specifies the number of contiguous samples from each input channel that are used to predict emissions in a specific time sample *n*. The function in (1) defines the calculated emissions at sample n and relies on the inputs from sample $n-w$ up to sample n, where w denotes the window width as measured in number of samples.

$$y(n) = f(x(n), x(n-1), \dots, x(n-w)) \quad (1)$$

Previous research [28] reported a sensitivity analysis that demonstrated that increasing the window width led to improved accuracy, but also resulted in a concomitant increase in the computational costs. Based on the findings of this analysis, the optimal balance was achieved with a window width of 50 samples, which corresponds to a time frame of 5 seconds.

8) Results

Direct Model Vs Two-stage Model
The system being examined consists of two distinct components: the engine, which generates emissions during the combustion process, and the after-treatment system, which reduces or eliminates those emissions before they are released into the environment, as shown in Figure 1.
This study compares two different approaches for developing a virtual sensor of the tailpipe emissions: the first, known as the "direct model", treats the engine and after-treatment system as a unique black box, receiving input solely from the ECU and aiming at estimating the emissions at the tailpipe. In contrast, the second method, known as the "two-stage model", considers the engine and the after-treatment system as two separate systems. First, the instantaneous engine-out emissions are estimated using the first-stage model that replicates the engine system. These results are then used as an additional input for the second-stage model, which emulates the behavior of the after-treatment system. The ultimate objective of the second-stage model is to estimate the emissions at the tailpipe using input from the ECU and the engine-out emissions predicted by the first-stage model.

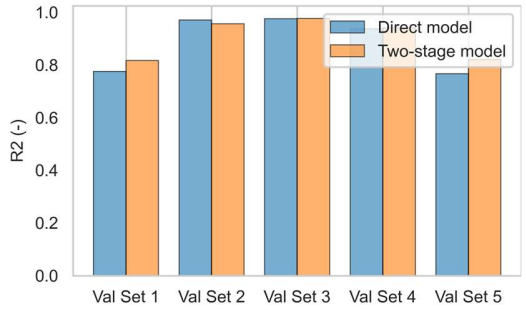

Figure 5. Comparison between direct and two-stage model on a 5-fold cross validation

The results obtained from the 5-fold cross-validation analysis (Figure 5) indicate that the direct model exhibits slightly inferior performance compared to the two-stage model across all validation subsets. This outcome was expected, as the two-stage model incorporates also the engine-out emissions within the input. However, given the marginal improvement in performance offered by the two-stage model, the decision has been made to proceed with the direct model. This choice primarily arises from practical considerations during implementation. The two-stage model necessitates the experimental measurement of both tailpipe emissions (target of the second model) and engine-out emissions (target of the first model) in the training phase. This requirement entails a more extensive effort in the experimental campaign, which may not always align with the company's timelines. Furthermore, the two-stage model involves the concatenation of two models, thereby introducing two distinct sources of error. In contrast, the direct model has demonstrated nearly identical performance while maintaining a lower degree of complexity by virtue of the use of a single model.

Features correlation analysis
To explore the relationships among these features, a correlation analysis is conducted using hierarchical clustering. This analytical approach effectively emphasizes the interdependence among the different input signals. The clusters are formed utilizing the Ward's linkage method [29]. By employing this technique, valuable insights into the data structure are obtained, facilitating a better understanding of how these signals relate to each other.

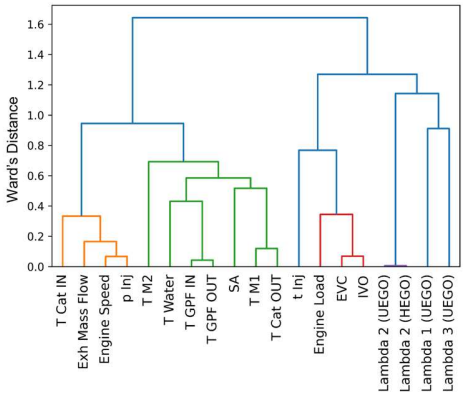

Figure 6. Dendrogram of hierarchical clustering on the input features

The dendrogram in Figure 6 shows how clusters are composed using U-shaped links to connect parent and child clusters. The U-link's curve indicates a cluster merge, while its arms show the specific merged clusters and their distance apart. This distance is also the cophenetic distance between the original observations in the child clusters.
The temperatures are grouped together within the same cluster due to their high degree of correlation, which imparts redundant information. Likewise, the second lambda sensor (referred to as HEGO) and its associated linearized value (referred to as UEGO) are clustered together. This clustering is done to address their strong interdependency. Another cluster merges engine load and valve opening/closing angles. This is driven by the fact that the ECU controls valves actuations at varying angles in response to the engine's load conditions. Finally, the temperature at the catalytic converter inlet, precisely measured downstream from the exhaust manifold, exhibits significant correlation with both the exhaust gas flow rate and the engine speed. These variables serve as indicators of the thermal power generated by the engine itself, thereby warranting their inclusion in the same cluster.no conversion

Importance of features and of the third lambda sensor
The FIP algorithm results reported in Figure 7, show the importance of the selected input feature for the prediction of each pollutant species. The third lambda probe has been found to have a significant impact especially on the estimation of CO and NOx emissions. This implies that the installation of this sensor in experimental setups is essential to

obtain a correct estimate of these species. The other two lambda probes are not given much weight, except for the second probe, which has a significant weight for CO only. This result may seem counterintuitive, because in extreme conditions of catalyst saturation or depletion, the aftertreatment system should have almost zero conversion efficiency, and the emissions should almost entirely bypass it. In this condition, theoretically, the air-fuel ratio measured at the engine outlet (at the first lambda probe) should be a good indicator of the emissions generated. However, this does not emerge from the FIP analysis, probably because the after-treatment system under study is homologated according to the last emissions legislation and is equipped with a very high catalytic volume and a control system able to avoid low conversion efficiency conditions. Therefore, the experimental tests remain in low catalytic efficiency for a negligible amount of time compared to the duration of the entire test, making this phenomenon not influent on the overall test.

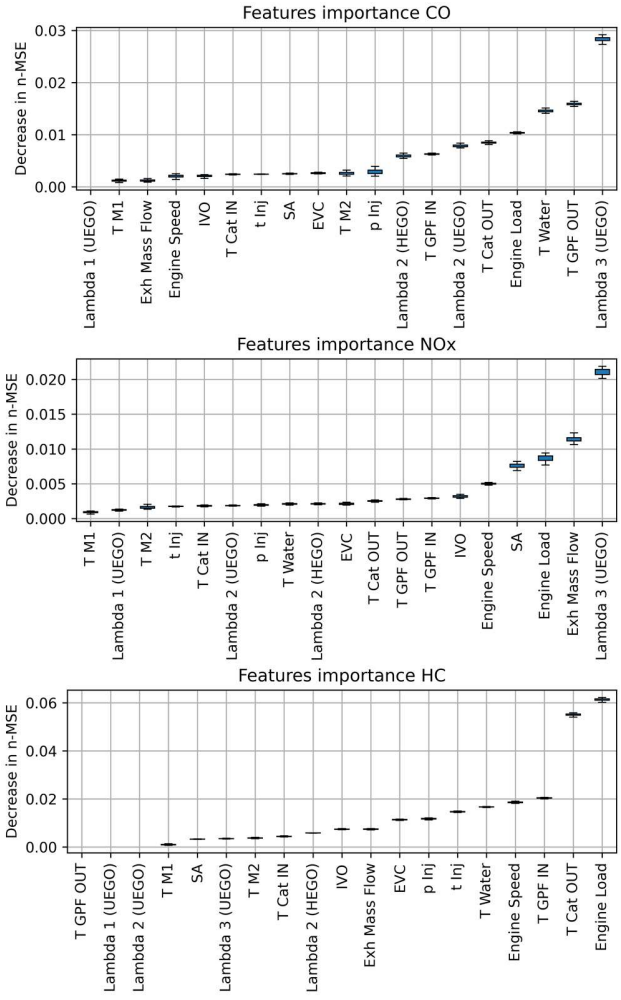

Figure 7. Features importance for CO, NOx and HC prediction

On the other hand, for HC emissions, the algorithm considers the lambda probes to be less important. This is because, during the driving cycle, there are no phases, outside of the cold start, where significant events occur in terms of HC production at the tailpipe. For this reason, the physical phenomenon that primarily influences HC production is strongly linked to the operating conditions, and in particular to the engine load, and to achievement of the light-off temperature of the catalytic converter, rather than the lambda value measured by the downstream probe (Figure 7).

Model prediction with different ECU software

This section presents the performance of three direct models for estimating NOx, CO, and HC mass flow rates and cumulative masses at the tailpipe.

Figure 8, Figure 9 and Figure 10 display the results obtained by the model, which are compared to the experimental measurements from two driving cycles, referred to as A and B respectively for the test with the base ECU software calibration and for the test with the optimized calibration. Notably, it is crucial to emphasize that HC emissions (Figure 10) are primarily formed during the initial phase of the driving cycles, before the catalyst reaches its light-off temperature, while are negligible in the hot phase. This can be partially due to the higher conversion efficiency of the HC in slightly rich conditions [30] and to side reactions occurring in the TWC in presence of water [31], that lead to abatement of HC and formation of CO (2).

$$C_3H_6 + 3H_2O \rightarrow 3CO + 6H_2 \quad (2)$$

This observation underscores the importance of the catalyst's state and temperature for a correct estimation of HC emissions. This finding is supported by the results obtained from the FIP algorithm, which indicates the exhausts temperature after the catalyst as the second most important feature. It is important to note that even if this measurement does not directly indicate the catalyst's temperature, it provides a qualitative indication of the exothermal phenomena occurring inside it. The comparison makes a distinction between scenarios with and without the inclusion of the third lambda probe as an input, represented respectively by red dashed lines and blue dashed lines.

The mass flow rate, estimated using virtual sensor models, exhibits a strong correlation with the corresponding experimental data for all pollutant species when the third lambda probe is included within the inputs. Furthermore, the virtual sensor models excel in estimating the most critical maneuvers where peaks of emissions are produced, with remarkable precision in terms of timing and, despite some minor errors, the models accurately predict the magnitude of the peaks as well.

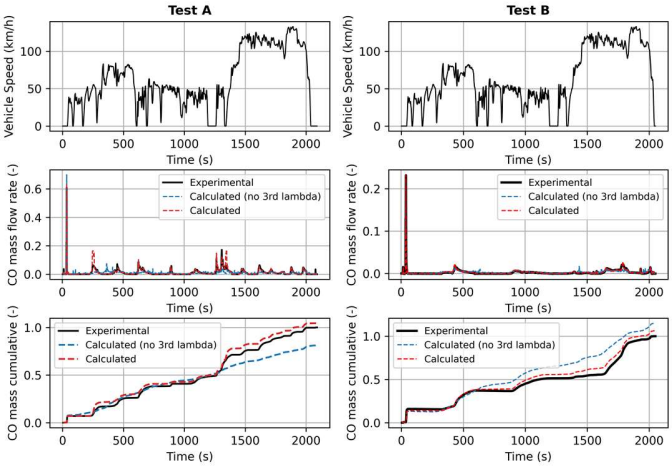

Figure 8. Experimental and calculated CO mass flow rate and cumulative mass on test cycles

In terms of cumulative emissions, the virtual sensor models successfully capture the overall trend. However, in certain cases, there is a slight deviation in the total cumulative value due to the inaccurate prediction of isolated peaks, which affects the cumulative calculation. Nonetheless, the overall trend of cumulative emissions is well-modeled, indicating the capability of virtual sensors to capture the long-term emission patterns and highlighting the robustness and

reliability of the virtual sensing technique in capturing the dynamic nature of pollutant emissions during various driving scenarios.
Removing the third lambda probe leads to a significant drop of accuracy for CO and NOx prediction. Indeed, the prediction without the additional lambda sensor (blue dashed line) diverges significantly more from the experimental signal (black solid line) than the prediction with the third lambda probe (red dashed line), which is almost superimposed on the experimental signal in terms of both mass flow rate and cumulative mass (Figure 8 and Figure 9).
On the contrary, HC prediction is unaffected by the presence of the third lambda probe, because, in this case the main phenomenon causing HC formation is the cold start phase.

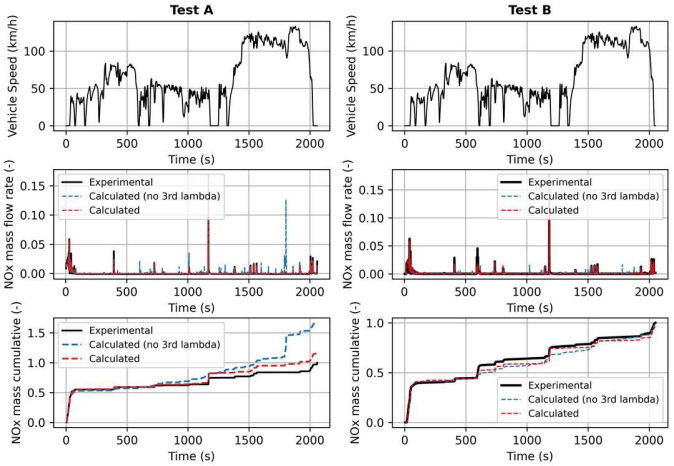

Figure 9. Experimental and calculated NOx mass flow rate and cumulative mass on test cycles

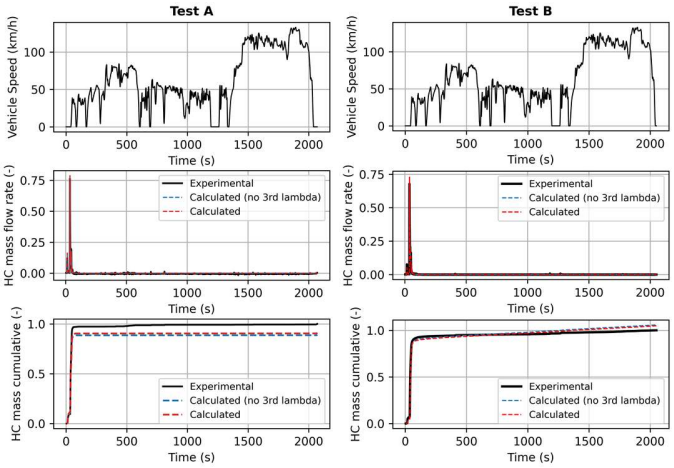

Figure 10. Experimental and calculated HC mass flow rate and cumulative mass on test cycles

Table IV reports the percentage errors on total masses made by the virtual sensors for all the species considering both test cycles and both model configurations, with and without the additional lambda probe. In the case of CO, the additional lambda sensor leads to an error reduction of approximately 14% for test A and 9% for test B. For NOx, the error reduction is even 40% for test A, while it is negligible for test B. However, this is only a consideration made on the total amount of mass, which is affected by error compensation. This means that in the total mass calculation, overestimations are compensated by underestimations. However, looking at the general trend of the two predictions in test B, the one with the third lambda sensor is much more coherent with the experimental signal.

As anticipated, the relative error on HC does not depend on the introduction of the third lambda sensor, as already stated in the previous paragraph.
Overall, the performances with ECU software A and B are comparable, proving the robustness of the model to different calibrations.

One concluding aspect worth contemplating pertains to the computational time, which is remarkably reduced, requiring a mere approximate of 1.5 seconds for estimating emissions throughout an entire 30-minute driving cycle. The employed models were executed on a computing system equipped with an Intel(R) Xeon(R) CPU @ 3.20GHz and 64GB of RAM.

Table IV. Summary of relative errors on cumulative emissions

	Relative error on cumulative CO		**Relative error on cumulative NOx**		**Relative error on cumulative HC**	
	No 3rd lambda	*Default*	*No 3rd lambda*	*Default*	*No 3rd lambda*	*Default*
Test A	-18.8%	4.5%	66.8%	16.8%	-9.8%	-8.8%
Test B	15.1%	6.4%	-10.1%	-11.9%	4.7%	4.9%

9) Conclusions

The capability to predict tailpipe emissions is a significant advantage for OEMs in terms of development time and costs. The main advantage of machine learning models is their enormous computational speed, which allows driving cycles to be simulated in a short period of time, reducing the number of required experimental tests and enabling real-time implementation.
This study has highlighted some critical issues associated with the industrial implementation of machine learning techniques for emission prediction, including: sampling of signals acquired at different frequencies, lack of synchronization between emissions (measured with a dedicated system) and ECU signals, or the problem of managing the dynamics of the signals.
A methodology has been presented that includes a pre-processing pipeline aimed at overcoming these critical issues, proposing an alignment based on local lagged cross-correlation and the introduction of a sliding window to take into account the dynamics of the inputs.
The goodness of this pipeline, applied on top of a LightGBM-based regressor, was evaluated on a real industrial application, where several considerations were made on two types of modeling (direct and two-stage), highlighting minimal differences between the two approaches and preferring the direct methodology for its greater simplicity.
In addition, the importance of input features has been investigated, highlighting that, especially for CO and NOx, it is essential to install a UEGO lambda probe downstream of the main catalyst to obtain a reliable prediction.
Finally, the comparison between two different ECU software calibrations has proved that this modeling workflow is robust to software changes and can therefore be used, in the development phase, to quantify in a virtual environment the impact that a particular calibration has on emissions.

With a view to a future ECU implementation, this study has also shown that such an application is compatible with real-time implementation thanks to the very fast computational time of the LightGBM. However, the challenge of applying data-driven algorithms to virtual sensor technology for gaseous emissions estimation remains a complex task.

Some steps need to be implemented in real-time: data acquisition from ECU, data pre-processing (resampling and alignment), estimation of CO, NOx and HC through LightGBM models, identification of possible anomalies or critical conditions and communication of the results to the final device in real-time (OBD or ECU).
Despite these aspects have not been directly examined in this work, the possibility of integrating the pipeline within the ECU for OBM (on-board monitoring) applications remains open.

References

[1] Z. C. Samaras *et al.*, "A European Regulatory Perspective towards a Euro 7 Proposal," in *SAE Technical Papers*, 2022. doi: 10.4271/2022-37-0032.

[2] V. Valverde Morales, M. Clairotte, J. Pavlovic, B. Giechaskiel, and P. Bonnel, "On-Road Emissions of Euro 6d-TEMP Vehicles: Consequences of the Entry into Force of the RDE Regulation in Europe," in *SAE Technical Papers*, 2020. doi: 10.4271/2020-01-2219.

[3] Z. Toumasatos, A. Raptopoulos-Chatzistefanou, D. Kolokotronis, P. Pistikopoulos, Z. Samaras, and L. Ntziachristos, "The role of the driving dynamics beyond RDE limits and DPF regeneration events on pollutant emissions of a Euro 6d-temp passenger vehicle," *J Aerosol Sci*, vol. 161, 2022, doi: 10.1016/j.jaerosci.2021.105947.

[4] D. Martin, N. Kühl, and G. Satzger, "Virtual Sensors," *Business and Information Systems Engineering*, vol. 63, no. 3, 2021, doi: 10.1007/s12599-021-00689-w.

[5] G. F. Scocozza *et al.*, "Development and Validation of a Virtual Sensor for Estimating the Maximum in-Cylinder Pressure of SI and GCI Engines," in *SAE Technical Papers*, 2021. doi: 10.4271/2021-24-0026.

[6] A. Della Torre, G. Montenegro, and A. Onorati, "CFD framework for the modeling of aftertreatment systems: Application to the study of an electrically heated DOC for Diesel e," 2018. doi: 10.1007/978-3-658-21194-3_25.

[7] F. Ranuzzi, N. Cavina, A. Brusa, M. De Cesare, and M. Panciroli, "Development and software in the loop validation of a Model-based water injection combustion controller for a GDI TC engine," in *SAE Technical Papers*, 2019. doi: 10.4271/2019-01-1174.

[8] A. Brusa *et al.*, "Development and experimental validation of an adaptive, piston-damage-based combustion control system for si engines: Part 1—evaluating open-loop chain performance," *Energies (Basel)*, vol. 14, no. 17, 2021, doi: 10.3390/en14175367.

[9] A. Brusa *et al.*, "Development and experimental validation of an adaptive, piston-damage-based combustion control system for SI engines: Part 2-implementation of adaptive strategies," *Energies (Basel)*, vol. 14, no. 17, 2021, doi: 10.3390/en14175342.

[10] B. Sarkar, S. R. Gundlapally, P. Koutsivitis, and S. Wahiduzzaman, "Performance evaluation of neural networks in modeling exhaust gas aftertreatment reactors," *Chemical Engineering Journal*, vol. 433, 2022, doi: 10.1016/j.cej.2021.134366.

[11] T. Donateo and R. Filomena, "Real time estimation of emissions in a diesel vehicle with neural networks," in *E3S Web of Conferences*, 2020. doi: 10.1051/e3sconf/202019706020.

[12] A. Warey, J. Gao, and R. Grover, "Prediction of Engine-Out Emissions Using Deep Convolutional Neural Networks," in *SAE Technical Papers*, 2021. doi: 10.4271/2021-01-0414.

[13] M. Fischer, "Transient NOx Estimation using Artificial Neural Networks," *IFAC Proceedings Volumes*, vol. 46, no. 21, pp. 101–106, 2013, doi: 10.3182/20130904-4-JP-2042.00006.

[14] C. Yu *et al.*, "Deep kernel learning approach to engine emissions modeling," *Data-Centric Engineering*, vol. 1, no. 3, 2020, doi: 10.1017/dce.2020.4.

[15] Y. Yu, Y. Wang, J. Li, M. Fu, A. N. Shah, and C. He, "A Novel Deep Learning Approach to Predict the Instantaneous NO_x Emissions From Diesel Engine," *IEEE Access*, vol. 9, pp. 11002–11013, 2021, doi: 10.1109/ACCESS.2021.3050165.

[16] C. M. A. Le Cornec, N. Molden, M. van Reeuwijk, and M. E. J. Stettler, "Modelling of instantaneous emissions from diesel vehicles with a special focus on NOx: Insights from machine learning techniques," *Science of the Total Environment*, vol. 737, 2020, doi: 10.1016/j.scitotenv.2020.139625.

[17] A. Brusa, E. Giovannardi, M. Barichello, and N. Cavina, "Comparative Evaluation of Data-Driven Approaches to Develop an Engine Surrogate Model for NOx Engine-Out Emissions under Steady-State and Transient Conditions," *Energies (Basel)*, vol. 15, no. 21, 2022, doi: 10.3390/en15218088.

[18] E. Giovannardi, A. Brusa, B. Petrone, N. Cavina, E. Corti, and M. Barichello, "An Enhanced Light Gradient Boosting Regressor for Virtual Sensing of CO, HC and NOx."

[19] L. Pulga, C. Forte, A. Siliato, E. Giovannardi, R. Tonelli, and I. Kitsopanidis, "AI strategies for the development of robust virtual sensors: an industrial case for transient PN emissions in a high performance engine".

[20] K. B. Altug and S. E. Kucuk, "Predicting Tailpipe NOx Emission using Supervised Learning Algorithms," in *3rd International Symposium on Multidisciplinary Studies and Innovative Technologies, ISMSIT 2019 - Proceedings*, 2019. doi: 10.1109/ISMSIT.2019.8932775.

[21] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, 2017.

[22] A. Altmann, L. Toloşi, O. Sander, and T. Lengauer, "Permutation importance: A corrected feature importance measure," *Bioinformatics*, vol. 26, no. 10, 2010, doi: 10.1093/bioinformatics/btq134.

[23] Q. Li *et al.*, "Recent advances in metal/ceria catalysts for air pollution control: Mechanism insight and application," *Environmental Science: Nano*, vol. 8, no. 10. 2021. doi: 10.1039/d1en00561h.

[24] J. B. Heywood, "Pollutant formation and control in spark-ignition engines," *Prog Energy Combust Sci*, vol. 1, no. 4, pp. 135–164, Jan. 1976, doi: 10.1016/0360-1285(76)90012-5.

[25] S. Liu *et al.*, "Brief industry paper: The matter of time - A general and efficient system for precise sensor synchronization in robotic computing," in *Proceedings of the IEEE Real-Time and Embedded Technology and Applications Symposium, RTAS*, 2021. doi: 10.1109/RTAS52030.2021.00040.

[26] T. Dorst, Y. Robin, S. Eichstädt, A. Schütze, and T. Schneider, "Influence of synchronization within a sensor network on machine learning results," *Journal of Sensors and Sensor Systems*, vol. 10, no. 2, 2021, doi: 10.5194/jsss-10-233-2021.

[27] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Springer Topics in Signal Processing*, 2009. doi: 10.1007/978-3-642-00296-0_5.

[28] A. Brusa, E. Giovannardi, M. Barichello, and N. Cavina, "Comparative Evaluation of Data-Driven Approaches to Develop an Engine Surrogate Model for NOx Engine-Out Emissions under Steady-State and Transient Conditions," *Energies (Basel)*, vol. 15, no. 21, 2022, doi: 10.3390/en15218088.

[29] S. Miyamoto, R. Abe, Y. Endo, and J. Takeshita, "Ward method of hierarchical clustering for non-Euclidean similarity measures," in *2015 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, 2015, pp. 60–63. doi: 10.1109/SOCPAR.2015.7492784.

[30] J. Kašpar, P. Fornasiero, and N. Hickey, "Automotive catalytic converters: Current status and some perspectives,"

in *Catalysis Today*, 2003. doi: 10.1016/S0920-5861(02)00384-X.

[31] K. Ramanathan and C. S. Sharma, "Kinetic Parameters Estimation for Three Way Catalyst Modeling," *Ind Eng Chem Res*, vol. 50, no. 17, pp. 9960–9979, Sep. 2011, doi: 10.1021/ie200726j.

Definitions/Abbreviations

ECU	Engine Control Unit
EVC	Exhaust Valve Closing
FIP	Feature Importance Permutation
GPF	Gasoline Particulate Filter
HEGO	Heated Exhaust Gas Oxygen
IVO	Intake Valve Opening
LightGBM	Light Gradient Boosting Machine
OEM	Original Equipment Manufacturer
RDE	Real Driving Emission
SA	Spark Advance
TWC	Three-Way Catalyst
UEGO	Universal Exhaust Gas Oxygen