POLITECNICO DI TORINO
Repository ISTITUZIONALE

Data-Driven Model for Real-Time Estimation of NOx in a Heavy-Duty Diesel Engine

(Article begins on next page)

26 June 2024

Data-Driven Model for Real-Time Estimation of NOx in a Heavy-Duty Diesel Engine

Alessandro Falai [1,2,*,†] and Daniela Anna Misul [1,2,*,†]

1   Department of Energy, Politecnico di Torino, Corso Duca Degli Abruzzi 24, 10129 Turin, Italy
2   Interdepartmental Center for Automotive Research and Sustainable Mobility (CARS@PoliTO),
    Politecnico di Torino, Corso Duca Degli Abruzzi 24, 10129 Turin, Italy
*   Correspondence: alessandro.falai@polito.it (A.F.); daniela.misul@polito.it (D.A.M.)
†   These authors contributed equally to this work.

Abstract: The automotive sector is greatly contributing to pollutant emissions and recent regulations introduced the need for a major control of, and reduction of, internal combustion engine emissions. Artificial intelligence (AI) algorithms have proven to hold the potential to be the thrust in the state-of-the-art for engine-out emission prediction, thus enabling tailored calibration modes and control solutions. More specifically, the scientific literature has recently witnessed strong efforts in AI applications for the development of nitrogen oxides (NOx) virtual sensors. These latter replace physical sensors and exploit AI algorithms to estimate NOx concentrations in real-time. Still, the calibration of the algorithms, together with the appropriate choice of the specific metric, strongly affects the prediction capability. In the present paper, a machine learning-based virtual sensor for NOx monitoring in diesel engines was developed, based on the Extreme Gradient Boosting (XGBoost) machine learning algorithm. The latter is commonly used in the literature to deploy virtual sensors due to its high performance, flexibility and robustness. An experimental campaign was carried out to collect data from the engine test bench, as well as from the engine electronic control unit (ECU), for the development and calibration of the virtual sensor at steady-state conditions. The virtual sensor has, since then, been tested throughout on an on-road driving mission to assess its prediction performance in dynamic conditions. In stationary conditions, its prediction accuracy was around 98%, whereas it was 85% in transient conditions. The present study shows that AI-based virtual sensors have the potential to significantly improve the accuracy and reliability of NOx monitoring in diesel engines, and can, therefore, play a key role in reducing NOx emissions and improving air quality.

Keywords: compression ignition engine; experimental data processing and analysis; model validation; machine learning; combustion modeling

1. Introduction

Air pollution caused by vehicle emissions, including nitrogen oxide (NOx), is a major environmental and public health concern. Strict regulations have been implemented globally to reduce NOx emissions from vehicles, and car manufacturers are actively seeking ways to optimize the performance of their internal combustion engines (ICEs) to meet these regulations [1].

The most well-known techniques for reducing and controlling engine-out NOx basically involve the following: the use of Exhaust Gas Recirculation (EGR) [2–6]; the deployment of advanced combustion technology, such as homogeneous charge compression ignition (HCCI) [7,8] and premixed charge compression ignition (PCCI) [9,10], which can achieve low NOx emissions by carefully controlling the air–fuel mixture and the combustion process; different injection strategies targeting pressure and timing [11,12]. As far as the vehicle tailpipe emissions are concerned, the main NOx reduction techniques involve the following: filters and catalysts, such as Selective Catalytic Reduction (SCR)

[13], which uses a catalytic converter to reduce NOx emissions by injecting a reducing agent, such as urea, into the exhaust stream; Lean NOx Trap (LNT) [14], which exploits a catalytic converter to temporarily store NOx emissions for reduction at specific engine operating conditions; Diesel Particulate Filter (DPF) [15], which uses a filter to capture particulate matter from the exhaust stream and can also help in reducing NOx emissions. Still, the target NOx reduction performance can only be accomplished by means of reliable NOx sensing.

Several models can be used for NOx prediction in engines, and engine maps exploitation is a well-established approach to calibrate the model, which can be tested on-road during transient conditions [16–18]. Engine maps are the graphical representations of the relationship between various engine operating parameters (e.g., load, speed, fuel flow rate) and engine performance characteristics (e.g., power output, emissions). These maps can be used to predict the NOx emissions of an engine under a given set of operating conditions [19,20]. A method exploiting the engine maps for NOx prediction was developed in [21]. An engine map was implemented in the electronic control unit (ECU) and represented by mathematical models to compute NOx emissions as a function of the observed operating conditions of the engine. The model was calibrated using engine test bench data, and tested on-road to determine its prediction capability under real-world conditions. Applications related to the virtual NOx sensing, exploiting the engine map operating points at steady-state conditions [22], can be found in the literature, together with applications derived from experimental test bench acquisition [23]. However, little research has been conducted on the virtual sensing and prediction of NOx during on-road real-world driving conditions, accounting for the effects of transient phenomena.

The optimization of NOx emissions remains a challenge in the present and virtual sensing techniques are widely used [24,25] to avoid the disadvantages of physical solid-state sensors to retrieve engine-out NOx information in diesel engine applications. Recently, several works have aimed at designing virtual sensors exploiting machine learning algorithms [26,27] in order to predict the engine-out NOx levels, based on the engine operating conditions, as from the ECU. Machine learning algorithms appear to be very promising as virtual measurement tools due to their ability to detect strongly non-linear behavior in the analyzed physical system [28,29]. Several machine learning algorithms have been applied to the prediction of engine NOx for virtual sensor applications, including linear regression [30], support vector machines (SVM) [31], and artificial neural networks (ANNs) [32]. Among these machine learning models, the XGBoost has proved to be particularly effective, outperforming other machine learning models, such as gradient boosting (GBT) and random forest (RF) [33]. This is due to the XGBoost's ability to handle large datasets and to cope with 'missing values', thus making it well-suited for virtual sensor applications where data may be incomplete or noisy. Moreover, the XGBoost was found to be a powerful and efficient machine learning tool for a variety of applications, including the prediction of engine-out NOx levels [34]. Recent studies also explored the use of ensemble models, such as random forest and adaptive boosting, in addition to the XGBoost [35,36]. These models combine the predictions of multiple individual models to achieve improved accuracy, whilst attaining a robustness equivalent to the that of the single models. Moreover, the ensemble methods mainly rely on randomization techniques, thus creating many different solutions to the considered problem. In this framework, the XGBoost algorithm has exhibited great prediction performance when compared to other ensemble models, such as GBT and RF, [37] causing to be selected for real-time applications. Moreover, some Extreme Gradient Boosting Regression Tree (XGBoost) models have been created to investigate the accuracy in estimating physical parameters, such as tailpipe NOx emissions [38,39]. However, the literature has limited research on the control and monitoring of engine-out NOx emissions under transient conditions.

The present paper, hence, focused on the use of the gradient boosting algorithm known as extreme gradient boosting (XGBoost) for real-time engine-out NOx sensing in real-world driving conditions. The XGBoost-based virtual sensor was calibrated in steady-

state conditions and then validated in transient on-road traces following an experimental campaign conducted on the test bench and on the road.

2. Materials and Methods

Virtual sensing techniques were exploited as valid alternatives to physical measurement instruments, due to economic or application limits of the latter. The present study focused on the machine learning-based virtual sensor development for nitrogen oxides prediction in a diesel engine application. At the very beginning, a robustness analysis was conducted on the acquired steady-state data for the performance evaluation of the machine learning algorithms. Then, XGBoost prediction capabilities were verified and validated over a real-world driving mission through an experimental campaign conducted on-road. The case study and the involved datasets, as well as the proposed method for model training and validation, were investigated.

In this section, the proposed method, composed of sequential steps, is discussed and shown in Figure 1.
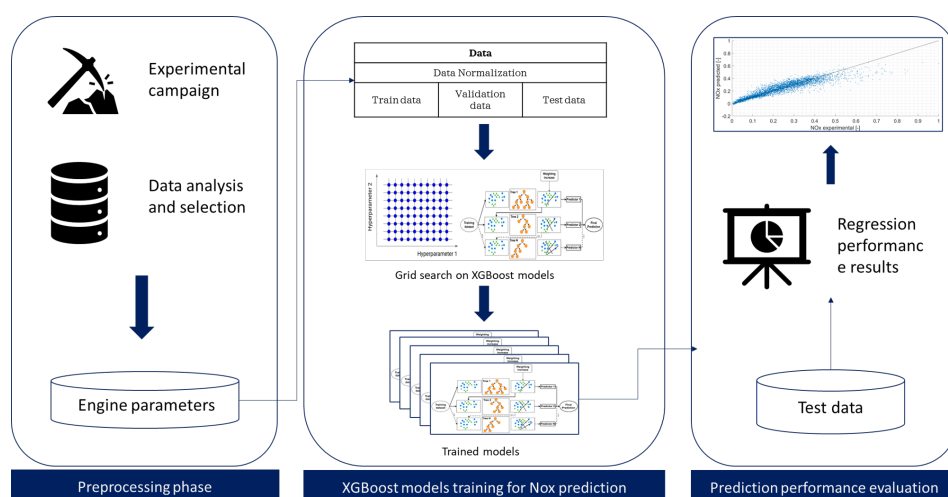

Figure 1. The proposed method for the model's performance evaluation.

In the preprocessing phase, the engine data values acquired from the experimental campaigns were analyzed, handled and 'cleaned'. In the following steps, related to the XGBoost models training phase, the data were normalized and split into training, validation and test sets when necessary, and then employed to perform the learning process of several XGBoost architectures. The grid search algorithm was used as a powerful hyperparameters tuning technique to find the most accurate architecture for each specific case study. Given that the performance of an AI model is heavily dependent on the hyperparameter values, a combination of grid search and cross validation approaches [40] was used to identify the optimal values for a specific model. As with any machine learning model, the XGBoost was defined by a set of hyperparameters that had to be specified in order to tailor the model for the specific application. In the final step, the performance of the best XGBoost architecture was evaluated by considering a test dataset according to different metrics, i.e., *RMSE* and coefficient of determination R^2.

2.1. Preprocessing Phase

The case study for the deployed AI-based virtual sensor was an 11 L diesel engine for heavy-duty application, tested both at steady state and under transient conditions. The details of the experimental campaign, together with a deep insight into the experimental set-up, as well as the engine behavior, can be found in [41]. The main engine parameters at steady-state conditions were acquired by means of an engine test bench equipped with physical sensors mounted on the acquisition system and through the variable acquired by the electronic control unit (ECU) sensors and maps, whereas in the transient case a

prototype heavy-duty (HD) vehicle was tested on-road and the main engine parameters were acquired by the on-board system. The main characteristics of the case study are shown in Table 1.

Table 1. Main engine parameters and test characteristics.

Segment of Application	Displacement	Turbocharger
Heavy-duty vehicles & trucks	11 L	VGT type
Fuel Injection System	**Engine Params**	**n° Engine Points**
High pressure common rail	14	4711 [1]

[1] The operating engine points are related to the steady state tests over the engine map.

The main engine parameters acquired for the different operating conditions follow:

- ω_{eng}: engine speed in rpm;
- Q_{tot}: total amount of injected fuel inside the cylinders in $mm^3/(cycle \times cylinder)$;
- P_{rail}: pressure in the rail system in bar;
- Q_{main}: amount of injected fuel during the main injection in $mm^3/(cycle \times cylinder)$;
- Q_{pil}: amount of injected fuel during the pilot injection in $mm^3/(cycle \times cylinder)$;
- SOI_{main}: start of the injection of the main injection in degree;
- SOI_{pil}: start of the injection of the pilot injection in degree;
- $IMAP$: pressure value in the intake manifold in bar;
- $IMAT$: temperature value in the intake manifold in K;
- O_2: oxygen concentration in the chamber in %;
- Q_{air}: amount of air introduced in the chamber in $Kg/(cycle \times cylinder)$;
- λ: ratio between air and fuel quantities.
- EGR: amount of exhaust gas recirculated in $Kg/(cycle \times cylinder)$;
- NOx: amount of nitrogen oxides emitted in ppm

The experimental campaign was carried out at Politecnico di Torino on an engine test bench equipped with an ELIN APA 100 AC dynamometer. The raw engine-out gaseous emissions were measured by means of an AVL AMAi60- endowed with two complete trains for the simultaneous measurement of the gaseous concentrations of the main species, both at the intake and exhaust manifolds. The test engine was equipped with thermocouples and piezoresistive pressure transducers to gather temperature and pressure at many points, including upstream and downstream from the turbine, compressor, and intercooler, as well as in the intake manifold and EGR circuit. [17] The experimental acquisitions were conducted according to different engine strategies that are not presented, due to confidentiality issues. The data are, hence, consistently reported in a normalized form. The data from the engine experiment were exploited for the machine learning model and the real-time virtual NOx sensor development. The engine steady state operating points are presented in terms of total injected fuel quantity versus engine speed in Figure 2.
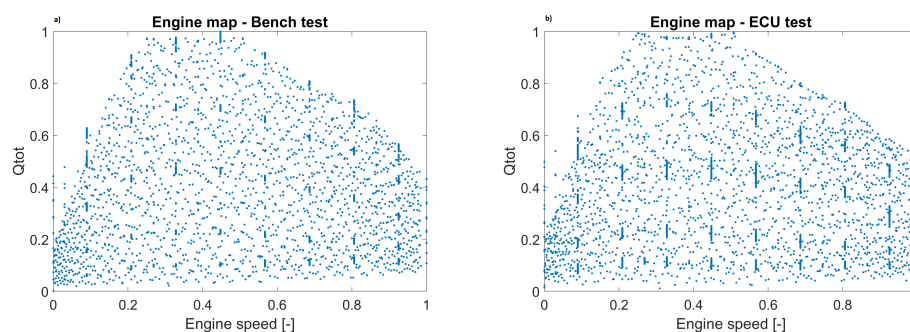

Figure 2. Engine operating points in stationary conditions through (**a**) test bench and (**b**) ECU acquisition systems.

The common engine map shape and the full load (FL) curve can be recognized in the graphs, as well as the presence of different strategies adopted during the experimental campaign bringing an unusual grouping phenomena. Given that the experimental acquisition occurred at steady-state testing conditions, the characteristic time associated with transient events was basically extinguished for each individual engine point depicted on the map. The road tests, on the other hand, were dominated by transient events with a large variation rate. Therefore, the sampling frequency which the engine variables were acquired at was set to 100 Hz.

2.2. XGBoost Models' Training

The development and assessment of the NOx predictor were based on the XGBoost algorithm and moved the steady state analysis, where the ML models were defined and trained on the engine operating points detected with the two different acquisition systems, i.e., test bench and ECU. A sensitivity analysis was carried out over the size of datasets exploited for the models' learning processes. Specifically, among the whole available data, the size of training and test datasets were changed in order to assess for the robustness of the selected models through a physical data fitting procedure. Considering that machine learning algorithms are data-driven models, the more data supplied to the training phase, the better the learning performance. Still, increasing the number of points during the training might lead to overfitting. As a result, reducing the training dataset dimension in favor of an increased number of test datasets allows for assessing the models' robustness, despite a loss in accuracy. Furthermore, in the creation of the predictors, a feature extraction approach was employed to identify the most influential and weighted variables. This approach, also known as feature importance, is regarded as a critical preprocessing step in the creation of the ML models [42].

In the processing phase, all datasets employed over the different analysis were split into train, validation and test sets in order to ensure a reliable learning process for the models and to evaluate the prediction performance. This particular mathematical problem belongs to the class of supervised learning, due to the presence of desired output exploited in the learning process [43]. The logic for the developed ML model, together with the XGBoost architectures and the hyperparameters involved, are shown in Figure 3.
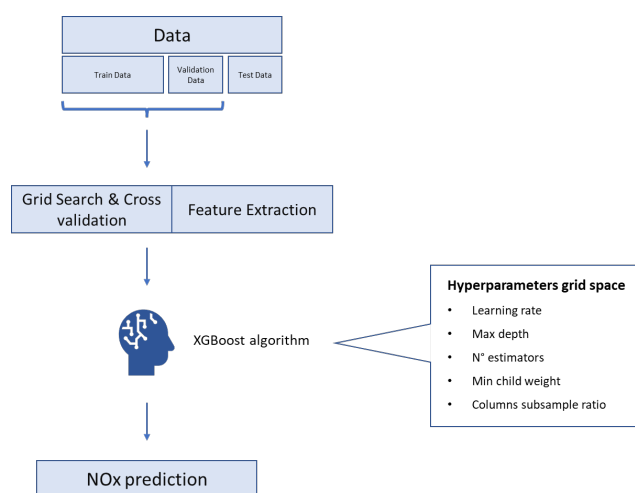
Figure 3. The workflow of XGBoost learning process. The involved Hperparameters in the XGBoost model are the following: the *n_estimators*, which specifies the number of decision trees to be boosted; the *max_depth*, which limits the depth in each tree's growth; the *Learning_rate*, which is a regularization parameter that shrinks feature weights in each boosting step; the *min_child_weight*, which corresponds to the minimum number of instances needed to be in each node of the tree; the *columns_subsample_ratio*, which is the subsample ratio of columns when constructing each tree [44,45].

Grid search and k-fold cross-validation techniques were combined to perform hyperparameter tuning in order to determine optimal values for the given model [40]. All different parameters were fed into a parameter grid, and based on a scoring metric (accuracy), the best combination was identified. The k-fold cross-validation process was allowed to perform the models' learning, evaluating each parameter combination over different datasets, i.e., validation set. The procedure involves partitioning the original training data into k subsets, where k is a positive integer, typically 5 or 10. The model is then trained on k-1 folds before being evaluated on the remaining fold. This procedure is repeated k times, with each test set fold being utilized once. The model's performance is measured by its average performance throughout all k iterations. In this investigation, the k parameter was assigned a value of 10. This strategy prevents overfitting, a typical issue that arises in machine learning when a model performs well on training data but fails to generalize to new, unknown data. Using k-fold cross-validation, the model is trained and assessed on multiple different subsets of data, thus yielding a more accurate performance estimation [46].

The XGBoost (eXtreme Gradient Boosting) is a popular and powerful machine learning algorithm that has been widely used for classification and regression tasks. The XGBoost is an optimized gradient-boosting decision tree algorithm, specifically designed to handle large datasets and to perform well with a high number of features. The algorithm is an ensemble method that combines decision trees to build a more powerful model. At its core, the XGBoost predicts the target variable by generating decision trees and optimizing the predictions by modifying the weights applied to each decision tree. Gradient descent is used in the optimization process to minimize the loss function, which assesses the difference between the anticipated and actual target values. During each iteration of the gradient descent optimization procedure, the XGBoost updates the weights of each decision tree, based on the gradient of the loss function. The technique employs a special algorithm for creating trees that divides the data into subsets, and the tree is built in a recursive fashion until a stopping criterion is reached [47]. There are a variety of loss functions that can be used, depending on the mathematical problem task. In the present study, the XGBoost-based predictor was laid on a regression task. Therefore, common loss functions included the mean squared error (MSE) and mean absolute error (MAE). These loss functions measure the difference between the predicted and the actual values and penalize large errors more heavily than small ones. In the present study, the MSE loss function was employed for the optimization process of the XGBoost parameters.

Feature selection was exploited as a machine learning approach that helps to determine which features in the experimental data are most significant in delivering the predictions. This is particularly useful if the number of variables is large and finding the most relevant ones is an objective of the analysis, or if is important to know which features are driving the model predictions. In the case of the XGBoost algorithm, feature importance is relevant for several reasons:

- Model interpretability: by identifying the most important features, a better understanding of how the model is producing the predictions, and how different features are interacting with each other, can be gained. This can be helpful to interpret the results of the models and to make more informed decisions based on its predictions.
- Feature selection: identifying the most important features can also be useful for feature selection, which is the process of selecting a subset of features to use in the models. By selecting the most important features, the performance of the models can be potentially improved by eliminating less important features that may be adding noise or reducing the model's ability to generalize.
- Model debugging: if poor model performance is experienced, identifying the most important features can be helpful to debug the model and identify potential issues. For example, if a particular feature is not statistically very important, it can be removed from the model.

Basically, the feature importance algorithm identifies, for each feature, a corresponding weight that represents the total gain of splitting data along the feature. It is calculated as the sum of the improvement in the loss function for all splits that use the selected feature. Features with higher weights are generally considered to be more important. The feature extraction technique was applied to the engine parameters listed in Section 2.1. The output of the algorithm was the computation of the relative importance of each parameter for NOx formation, and the results are presented in the Result section.

2.3. Prediction Performance Evaluation

Each selected model was exploited as a virtual predictor of NOx pollutant and the prediction performance investigated through different metrics. Together with the test data, the performance of all trained XGBoost architectures for each case study was analyzed, based on the following:

- the coefficent of determination R^2;

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x}_i)^2} \tag{1}$$

- the normalized RMSE considering the test dataset;

$$RMSE_{norm} = \sqrt{\frac{\sum_{i=1}^{n}(x_{n,i} - \hat{x_{n,i}})^2}{n}} \tag{2}$$

- the real RMSE in *ppm* considering the test dataset;

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{n}} \tag{3}$$

where x_i is the experimental data, $\hat{x}_i$ is the model estimated value, $\bar{x}_i$ is the mean of experimental measured data, $x_{n,i}$ are the normalized experimental data, $\hat{x_{n,i}}$ are the normalized estimated value and n is dataset sample size. The performance evaluation of the machine learning models and the selection of the best hyperparamenter values are widely discussed and analysed in the Results and Discussion section.

Using the high-level Python programming language and a PC with an Intel(R) Core(TM) i7-8700 processor at 3.20 GHz and 64GB RAM architecture, the virtual sensor was designed and the complete data processing phase was executed.

2.4. Virtual NOx Sensing in Steady-State Conditions

Virtual sensing is a frequently adopted approach to ensure sensorless solutions in the automotive field capable of properly calibrating the main engine parameter, thus allowing for consumption and emissions reduction. Still, considering the transient dynamic conditions that an engine encounters during regular on-road driving missions, the virtual NOx prediction sensor should be calibrated in stationary conditions throughout the entire engine operating range. Such operations require a considerable amount of experimental readings but would, in turn, enable the electronic control unit (ECU) to properly control the engine during real driving missions. The present paper, hence, moved from the development and calibration of an AI-based virtual sensor through the steady-state data set. It is worth remembering that two different sets of acquisition were available from the experimental campaign, namely, a set of data collected at the test bench (Bench) and a set of data directly acquired from the engine control unit (ECU). The ML virtual sensor was, hence, developed considering different combinations of the available data sets for the training and testing phase, to better investigate the algorithm response to the use of specific inputs. Moreover, a sensitivity analysis was carried out on the train and test sub-set sizes (test dataset size) in order to prove the robustness of the methodology. The different

cases analyzed are reported in Table 2, whereas the XGBoost model performance results are widely discussed in the Results and Discussion section.

Table 2. Steady-state case analysis based on datasets and train–test split sensitivity.

Case Study	Train–Test	Test Dataset Size [%]
#1	Bench–Bench	$[10; 50; 95]$
#2	ECU–ECU	$[10; 50; 95]$
#3	Bench–ECU	NO [1]

[1] The datasets involved in the training and test phases were derived from two different acquisition systems. Hence, a train–test split technique was not adopted.

In the Bench–Bench and ECU–ECU examples, the datasets used for the model learning were distinct and derived from the same experiments, i.e., the Bench–Bench case was based on the engine bench test data, whereas the ECU–ECU case was based on engine data collected from the ECU. In the last scenario examined, Bench–ECU, the training dataset and the test dataset were derived from the two applications. Therefore, there was no train–test split and all dataset samples were utilized for the different learning phases. Given that the bench acquisition system detects real events with very accurate physical sensors, whereas the engine map implements data tables, the model performance evaluation in Case Study 3 was dependent on the extent to which the physical phenomena were detected by the virtual sensor, as a potential real-time estimator to be implemented on ECU.

2.5. On-Road NOx Prediction

After calibrating and testing the virtual sensor in steady-state conditions to assess for the model robustness and NOx prediction performance, validation under dynamic conditions was carried out. In order for the virtual sensor to accurately predict NOx emissions under a range of operating conditions, it was important to train the model using engine map data and to test it on real-driving missions. Engine map data, which comprises engine speed and load information, replicates a variety of steady-state scenarios the engine may experience. However, these conditions may not precisely represent the dynamic, real-world operating conditions that the engine would face when driving on a road. By testing the sensor during an on-road driving mission, it is, in turn, possible to identify how well the virtual sensor performs under actual driving scenarios and to assess its ability to properly predict real-time emissions. As for integrating the virtual sensor in a control system, any potential issues or limitations that need to be addressed can, hence, be identified.

Therefore, for this investigation, stationary experimental data on the ECU and on the bench were utilized to train the algorithm, whilst on-road-based measurements were used to validate the algorithm and examine the accuracy of NOx prediction during real-world driving missions. During experimental road tests, the vehicle powered by a compression ignition engine was equipped with a physical measuring system for engine-out management and monitoring. The development and assessment of a virtual NOx sensor for real-time applications would enable the replacement of real sensors, possibly resulting in cost and reliability improvements. The real driving mission performed by the vehicle during the experimental campaign occurred in Piedmont and accounted for a variety of operating conditions and maneuvers related to the constraints and requirements of realistic urban and extra-urban scenarios. The cases under investigation are summarized in Table 3.

Table 3. Dynamic on-road case analysis based on training phase over different acquisition systems.

Case Study	Train–Test
#1	Bench–On-road mission
#2	ECU–On-road mission

The *Bench–Onroad* case is related to the training conducted on the engine test bench at steady state conditions and *ECU–Onraod* studies the learning process on stationary engine ECU data and validation on transient conditions. Since the development of the virtual sensor was oriented toward a future real-time implementation and could, thus, be implemented in the ECU, it was crucial to examine the two data collection systems independently, given the heterogeneous nature of the variables. Thus, the engine map-based model's prediction of performance may be compared to the real collection of physical sensors.

3. Results and Discussion

In this work, an XGBoost-based NOx virtual sensor was developed considering an engine steady-state calibration for real-time NOx monitoring during on-road driving missions. A preliminary study was conducted under stationary conditions to assess the model's capability to capture NOx phenomena under various operating conditions. In addition, a robustness study was conducted by progressively lowering the training data size and reversely increasing the test data one. In this way, the magnitude of noise could be varied and the quality of the model performance inferred. Consequently, the XGBoost model for the real-time estimation of NOx in on-road applications was built by training the model under steady-state conditions, from which the entire engine map domain might be completely covered during experiments, and testing it on real-world driving data.

3.1. NOx Prediction in Steady-State Conditions

The case study results concerning the training and test phases on bench test conditions (*Bench–Bench*), and taking into account the sensitivity analysis conducted on the train–test split, are summarized in Table 4 and the regression results are shown in Figure 4.

Table 4. Summary of performance prediction results for test case *Bench–Bench*. The test size is here reported for the sensitivity analysis, considering the different percentage values 10%, 50% and 95%.

Test Size [%]	10	50	95
R^2	0.98	0.97	0.85
RMSEnorm	0.017	0.022	0.052
RMSE[*ppm*]	60.0	77.4	186.5
q	0.0034	0.0062	0.0224
m	0.985	0.976	0.912
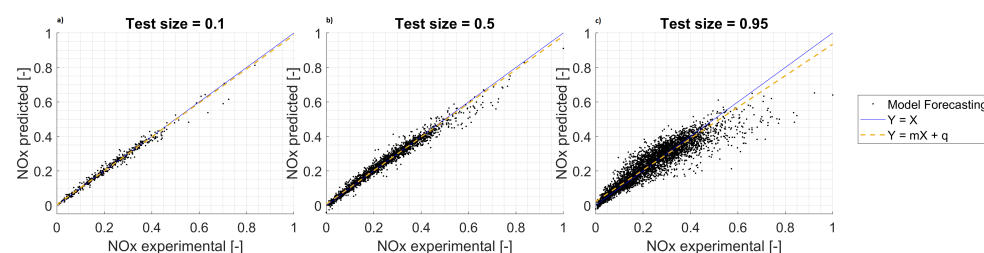
Figure 4. Regression prediction results on test dataset for bench test conditions. Sensitivity analysis on the three different sizes of test dataset: (**a**) 10%, (**b**) 50% and (**c**) 95%.

The considerable reduction in the engine points ascribed to the model learning phase, as the test dataset increased from 0.1 to 0.95, caused the prediction performance to decrease.

When 95% of the data was used for the testing, significant errors were attained, despite the moderate increase in rmse as the split switched from 10% of test size (60.0 ppm) to 50% of test size (77.4 ppm). It is, in fact, worth underlining that the model prediction performance remained high, despite the halving of the training data set, thus, indicating a remarkable capacity to properly capture the physical phenomena. Additionally, the regression line in Figure 4 demonstrates the rise in noise brought on by the test size-appropriate sensitivity analysis.

The case study results concerning the training and test phases on ECU test conditions (*ECU–ECU*) are summarized in Table 5 and the regression results are shown in Figure 5.

Table 5. Summary of performance prediction results for test case *ECU–ECU*. The test size is here reported for the sensitivity analysis, considering the different percentage values 10%, 50% and 95%.

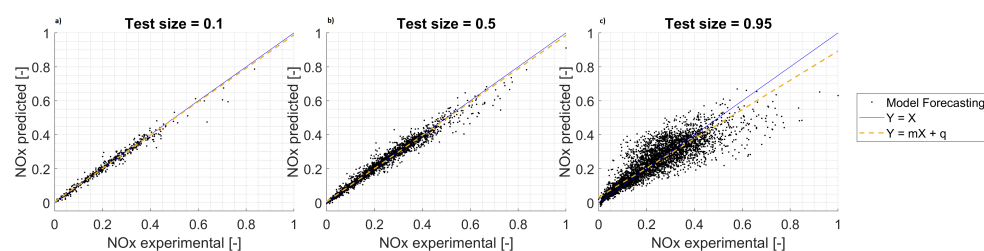| Test Size [%]	10	50	95
R^2	0.98	0.97	0.76
RMSEnorm	0.021	0.024	0.067
RMSE[*ppm*]	73.3	84.9	237.7
q	0.0038	0.0047	0.0309
m	0.985	0.9802	0.862

Figure 5. Regression prediction results test dataset for ECU test conditions. Sensitivity analysis on the three different size of test dataset: (**a**) 10%, (**b**) 50% and (**c**) 95%.

As for the previous case, the accuracy of NOx prediction tended to decline (and the error to grow) as the test size increased and the training size decreased. The surprisingly slight decrease in R2, 0.98 for 90% and 0.97 for 50% of the full dataset, as a result of nearly half the amount of data available for model training, is noteworthy. This was due to the high predictive power of these machine learning methods, especially considering that the data in this particular application were collected under stationary conditions. Therefore, it was more efficient to capture NOx generation phenomena beginning with steady-state engine operating conditions.

Finally, the case study results concerning the training conducted on Bench test conditions and test conducted on ECU test conditions (*Bench–ECU*) are summarized in Table 6 and the regression results are shown in Figure 6.

Table 6. Summary of performance prediction results for test case *Bench–ECU*.

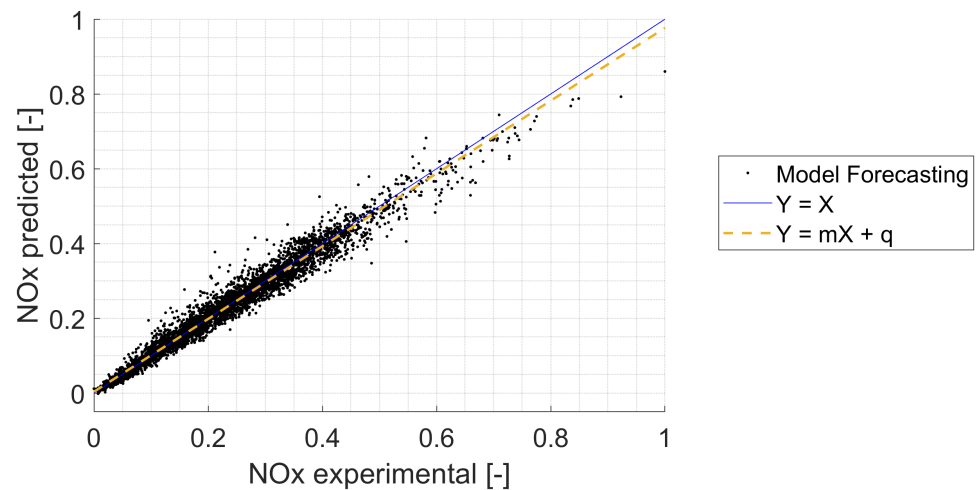|	
R^2	0.97
RMSEnorm	0.024
RMSE[*ppm*]	86.7
q	0.0037
m	0.972

Figure 6. Regression prediction results of training conducted on bench test conditions and test carried out under ECU test conditions.

In terms of the robustness of the virtual sensor, the findings obtained in this example were consistent with those obtained in the previous cases, as can be observed. In addition, the Bench–ECU example demonstrates how the formation phenomenology of nitrogen oxides was accurately estimated (R^2 of about 98%) by the developed virtual sensor which was tested on ECU engine data.

The more features a model contains, the more complex it is (and the sparser the data), and, therefore, the more susceptible it is to variance-induced errors. For this reason, the feature importance approaches belong to the feature engineering process which entails selecting the minimum required features to generate a feasible model [48]. Given that the formation of nitrogen oxides in diesel engines is strongly influenced by the specific engine operating conditions, the feature extraction approach simplifies the models and can provide an assessment of the model performance. Indeed, the robustness of the predictions can also be verified by comparing the statistically most relevant engine variables for the models to the empirical knowledge of the physical phenomenon. Considering the full availability of data during the training phase of the models, i.e., a test size equal to 10% for both *Bench–Bench* and *ECU–ECU* case studies, the feature extraction approach outputs are shown in Table 7. For the same train–test split value, the hyperparameters of the best XGBoost architecture for each considered analysis are reported in Table 8.

Table 7. The feature importance outputs for the three steady-state cases study which highlight the relative importance of each variable. The green labeling identifies the most influencial engine variables up to reaching 90% of the threshold in the feature importance algorithm for the XGBoost application.

Variable	Bench–Bench	ECU–ECU	Bench–ECU
1	$SOI_{main}(13.9\%)$	$SOI_{main}(13.5\%)$	$SOI_{main}(14.2\%)$
2	$P_{rail}(13.4\%)$	$P_{rail}(13.4\%)$	$\omega_{eng}(13.3\%)$
3	$\omega_{eng}(13.3\%)$	$\omega_{eng}(13.0\%)$	$P_{rail}(13.3\%)$
4	$\lambda(9.5\%)$	$\lambda(10.7\%)$	$\lambda(9.7\%)$
5	$O_2(9.3\%)$	$IMAT(7.9\%)$	$O_2(9.6\%)$
6	$IMAT(7.5\%)$	$Q_{tot}(7.9\%)$	$IMAT(7.7\%)$
7	$Q_{tot}(7.4\%)$	$O_2(6.3\%)$	$EGR(7.4\%)$
8	$EGR(6.8\%)$	$Q_{main}(6.3\%)$	$Q_{tot}(6.5\%)$
9	$IMAP(5.6\%)$	$IMAP(6.1\%)$	$Q_{air}(5.9\%)$
10	$Q_{air}(5.5\%)$	$SOI_{pil}(5.4\%)$	$IMAP(5.1\%)$
11	$SOI_{pil}(4.5\%)$	$Q_{air}(4.6\%)$	$SOI_{pil}(3.8\%)$
12	$Q_{main}(2.6\%)$	$EGR(4.1\%)$	$Q_{main}(2.7\%)$
13	$Q_{pil}(0.7\%)$	$Q_{pil}(0.8\%)$	$Q_{pil}(0.8\%)$

According to the feature importance definition, the variables SOI_{pil}, Q_{main} and Q_{pil} were discarded as they proved not to be significantly affected by large variations. The algorithm, therefore, did not consider SOI_{pil} and Q_{pil} to be particularly significant in the evolution of NOx emissions. Furthermore, since the Q_{main} was strictly proportional to Q_{tot}, only one of them was taken into account.

Table 8. Best XGBoost architectures and grid values setup for the three case studies under investigation. The mean squared error (MSE) evaluation metric was used to obtain the optimal value by the GridsearchCV algorithm.

Hyperparameters	Grid Values	Bench–Bench	ECU–ECU	Bench–ECU
learning rate	[0.1, 0.05, 0.08, 0.1, 0.15]	0.08	0.08	0.08
max_depth	[3, 4, 5]	4	4	4
n° estimators	[500, 1000, 1200, 1500]	1500	1500	1500
min_child_weight	[3, 5, 7]	3	7	7
colsample_bytree	[0.5, 0.7, 1.0]	1	1	1

As can be seen, the most relevant engine variables identified by the feature extraction process were approximately the same in the three cases analyzed. First, the variables most relevant for predicting NOx emissions are likely to be consistent across different types of data sources and testing environments. For instance, variables such as engine load, fuel flow, and exhaust gas temperature are likely to be important for predicting NOx emissions, regardless of whether the data are collected from a test bench or from the vehicle's engine control unit. Moreover, the most relevant engine variables selected by XGBoost were similar across the three different cases because the data used to train and test the model were similar. For instance, the dataset used to train the model on the test bench was similar to the dataset used to test the model on the vehicle ECU. Therefore, the most relevant variables were similar in both cases. For the same reasons, the descriptive patterns of the data detected by the algorithm would be very similar. Therefore, the architectures of the XGBoost in the three cases were very similar to one another.

3.2. On-Road NOx Prediction

Once the model was assessed in terms of its high predictive performance and its robustness in capturing phenomenological events under steady-state conditions, a virtual NOx sensor was built for real-time applications taking advantage of the experimental data collected on-road. As opposed to the bench tests, where data wre collected while the engine was running under controlled conditions to allow for consistent and repeatable testing conditions, on-road data were acquired while the vehicle was being driven on actual roads under realistic scenarios. This data reflected the real working conditions of the engine, including environmental parameters, such as temperature, humidity, and altitude, as well as the dynamic load variations that occurred during the driving mission. The on-road measurements also included the vehicle's speed and acceleration, which might impact engine performance. Consequently, the design of the virtual NOx sensor was based on its calibration in stationary conditions (either bench or ECU) and validated by means of real-world experimental tests. The considered engine parameters were identified by the feature importance process during the previous steps of the analysis. Therefore, the learning phase was covered by the feature importance output from the steady-state conditions. The predictive performance results for the two case studies are listed in Table 9 outlining the same statistical metrics.

Table 9. Performance prediction results over on-road test driving mission.

Case Study	#1 Bench—On-Road Mission	#2 ECU—On-Road Mission
R^2	0.76	0.75
RMSEnorm	0.054	0.055
RMSE[*ppm*]	192.0	194.9
m	0.92	0.97
q	-0.007	-0.017

From the values of the metrics, a noticeable reduction in predictive performance, of up to 75% accuracy, and a major shift in the regression line could be observed (q was significantly higher than the stationary cases), as well as a large increment in the error. This was mostly due to the fact that the operating conditions encountered by the engine during a road driving mission are significantly different from those encountered during the model learning phase described by stationary conditions. The predicted NOx in both case studies are shown alongside the experimental traces in Figure 7, where three different time windows [50 s 250 s], [500 s 700 s], [850 s 1050 s] of the missions, with a length equal to 200 s, are highlighted. Three fairly large time windows that corresponded to different operating conditions are shown so that the forecasting ability of the model could be tested in almost all real-world conditions the vehicle would face.

Observably, the anticipated NOx signals accurately represented the qualitative and quantitative trend of the experimental unit, as well as its transient trend during the driving mission. In the [50,250] time window, some mission points that overestimated or underestimated the experimental results led to negative NOx values. This was likely to be ascribed to the great occurrence of operating points where the lambda values reached extremely high values beyond the definition domain of the learning phase, as well as the rapid spikes in the fuel injected quantity values. Moreover, the signal was extremely spiky due to the fact that the model was an instantaneous measurement sensor and each point of the mission had operating conditions in terms of engine variables that were very different from those of the subsequent instants. The latter, together with the fact that the NOx formation dynamics with their characteristic times were not captured in the learning phase, connected to the differences in the experimental and predicted signals.
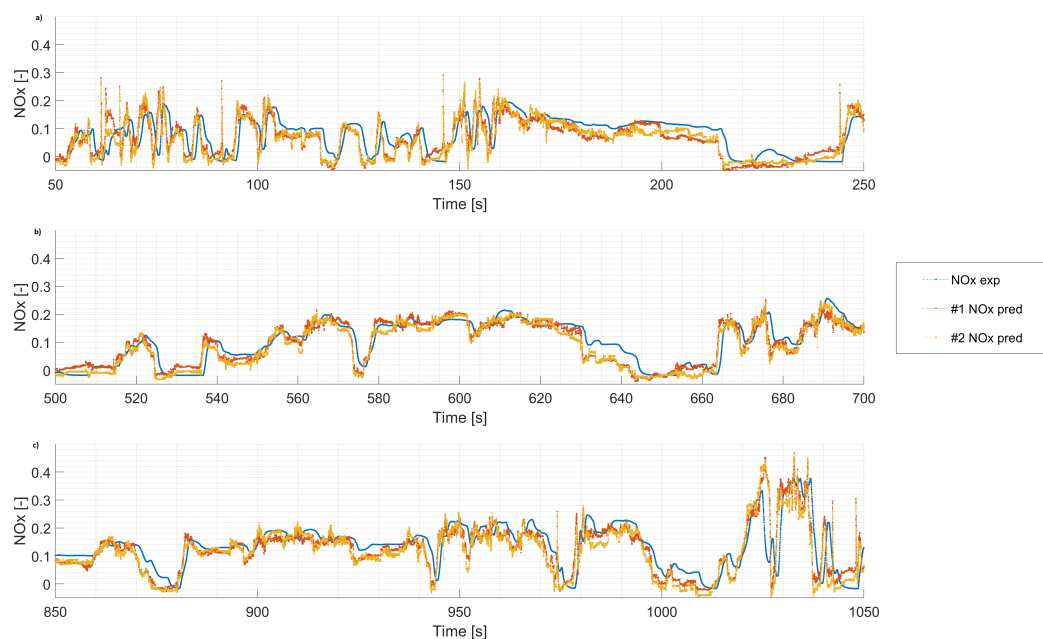
Figure 7. The NOx predicted signals for cases study #1 and #2. (**a**–**c**) depict three distinct 200 s-length windows of driving mission.

As seen in Table 9 and Figure 7, the lower R2 values could be attributable to the anticipation of the predicted signals relative to the observed signal. In real-time virtual sensing applications, the anticipation or delay in time of a predicted signal causes a drastic reduction in performance, leading, in this case, to incorrect emission control. Such behaviour is connected to the line and the measurement system's characteristic time delay. The latter is produced by the time lag between the observed phenomenon, i.e., NOx measured level, and the time frame such a level would refer to. As a matter of fact, the predicted NOx were referred to the implemented engine parameters at the given time-step, whereas the measured NOx would have been produced by the engine setting at a previous time frame. The latter is connected to the time needed for the exhaust gases to travel from the engine discharge to the acquisition probes, as well as to the characteristic sensor time lag. Under steady-state conditions, such delay is basically erased by the stationary characteristic of the test, whereas it stems from transient conditions. By delaying the NOx signal output by the model by 1 s [49], the anticipated signal almost perfectly matched the experimental signal (see Figure 8 referring to one time window).
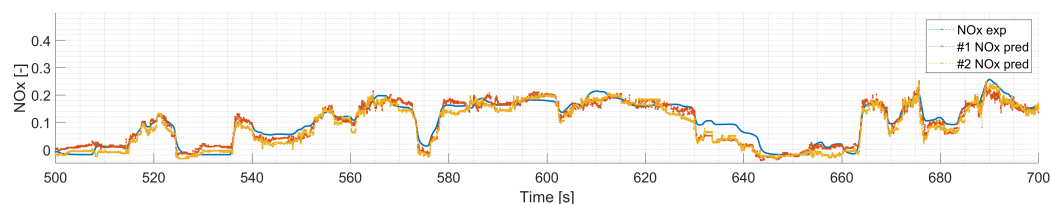

Figure 8. The NOx predicted signals for cases studies #1 and #2, taking into account the line delay of measurement system.

Taking this fact into account, the value of R2 changed to 85% for both case studies, thus emphasizing the real-time capabilities of the virtual sensor at predicting NOx emissions with a high level of accuracy [50].

4. Conclusions

The virtual sensor developed in the present study is based on the XGBoost algorithm and proved to be highly accurate in predicting NOx emissions for real-time diesel engine applications. The calibration of the virtual sensor was performed under steady-state conditions and then tested on an on-road driving mission, resulting in an accuracy of 85%. The virtual sensor was also found to be highly accurate in stationary conditions, with prediction results up to 98% in terms of accuracy when tested on the experimental data acquired at the engine test bench and through the ECU maps. The results indicate that the virtual sensor holds great potential to be a valuable tool for real-time emission monitoring and control in diesel engine applications. The virtual sensor can, in fact, be integrated into the Engine Control Unit (ECU), where it could be used, in conjunction with the ECU control system, to adjust the engine parameters in real-time to minimize NOx emissions. Future developments of the present paper may address different driving conditions (e.g., high speed, high load, and low speed) to further assess the capabilities of the virtual sensor in accurately predicting NOx emissions. The logic could also be transferred to AI deep networks, which could further improve the prediction accuracy and broaden the virtual sensor capabilities. For example, deep networks could be used to make the virtual sensor more capable of predicting NOx emissions in real-time under a broader range of conditions and to determine the characteristic time of NOx emissions phenomena by analyzing time-dependent events. In conclusion, the virtual sensor developed in the present study shows great promise as a tool for real-time emission monitoring and control in diesel engine applications.

Nomenclature

The following abbreviations were used in this manuscript:

AI	Artificial Intelligence
NOx	Nitrogen Oxides
XGBoost	Extreme Gradient Boosting
ECU	Electronic Control Unit
ICE	Internal Combustion Engine
EGR	Exhaust Gas Ricirculation
HCCI	Homogeneous Charge Compression Ignition
PCCI	Premixed Charge Compression Ignition
SCR	Selective Catalytic Reduction
LNT	Lean NOx Trap
ANN	Artificial Neural Network
SVM	Support Vector Machine
RMSE	Root Mean Squared Error
R^2	Coefficient of determination
HD	Heavy Duty
ω_{eng}	Engine speed
Q_{tot}	Total injected fuel quantity
P_{rail}	Rail pressure
Q_{main}	Fuel quantity of main injection
Q_{pil}	Fuel quantity of pilot injection
SOI_{main}	Start of injection of main injection
SOI_{pil}	Start of injection of pilot injection
$IMAP$	Intake Manifold Air Pressure
$IMAT$	Intake Manifold Air Temperature
O_2	Oxygen concentration
Q_{air}	Air quantity
λ	Air and fuel ratio
FL	Full load
ML	Machine Learning
MSE	Mean Squared Error
MAE	Mean Absolut Error

References

1. Regulation (EC) No 715/2007 of the European Parliament and of the Council of 20 June 2007 on Type Approval of Motor Vehicles with Respect to Emissions from Light Passenger and Commercial Vehicles (Euro 5 and Euro 6) and on Access to Vehicle Repair and Maintenance Information (Text with EEA Relevance). Available online: https://eur-lex.europa.eu/eli/reg/2007/715/oj (accessed on 22 December 2022)

2. Kannan, C.; Vijayakumar, T. Application of exhaust gas recirculation for NOx reduction in CI engines. In *NOx Emission Control Technologies in Stationary and Automotive Internal Combustion Engines*; Ashok, B., Ed.; Elsevier: Amsterdam, The Netherlands, 2022; pp. 189–222.

3. Baratta, M.; Finesso, R.; Misul, D.; Spessa, E. Comparison between Internal and External EGR Performance on a Heavy Duty Diesel Engine by Means of a Refined 1D Fluid-Dynamic Engine Model. *SAE Int. J. Engines* **2015**, *8*, 1977–1992. [CrossRef]

4. Jain, A.; Singh, A.P.; Agarwal, A.K. Effect of split fuel injection and EGR on NOx and PM emission reduction in a low temperature combustion (LTC) mode diesel engine. *Energy* **2017** *122*, 249–264. [CrossRef]

5. Maiboom, A.; Tauzia, X.; Hètet, J.-F. Influence of high rates of supplemental cooled EGR on NOx and PM emissions of an automotive HSDI diesel engine using an LP EGR loop. *Int. J. Energy Res.* **2008**, *32*, 1383–1398. [CrossRef]

6. Maiboom, A.; Tauzia, X. NOx and PM emissions reduction on an automotive HSDI Diesel engine with water-in-diesel emulsion and EGR: An experimental study. *Fuel* **2011**, *90*, 3179–3192. [CrossRef]

7. Stanglmaier, H.R.; Roberts, C.E. Homogeneous Charge Compression Ignition (HCCI): Benefits, Compromises, and Future Engine Applications. *SAE Int.* **1999**, *108*, 2138–2145.

8. Puškár, M.; Kopas, M. System based on thermal control of the HCCI technology developed for reduction of the vehicle NOX emissions in order to fulfil the future standard Euro 7. *Sci. Total Environ.* **2018**, *643*, 674–680. [CrossRef]

9. Torregrosa, A.J.; Broatch, A.; Garcìa, A.; Mònico, L.F. Sensitivity of combustion noise and NOx and soot emissions to pilot injection in PCCI Diesel engines. *Appl. Energy* **2013** *104*, 149–157. [CrossRef]

10. Aoyama, T.; Hattori, Y.; Mizuta, J.; Sato, Y. *An Experimental Study on Premixed-Charge Compression Ignition Gasoline Engine*; SAE Technical Paper; SAE: Warrendale, PA, USA, 1996. [CrossRef]

11. Cao, D.N.; Hoangm A.T.; Luu, H.Q.; Bui, V.G.; Tran, T.T.H. Effects of injection pressure on the NOx and PM emission control of diesel engine: A review under the aspect of PCCI combustion condition. *Energy Sources* **2020**, 1–18. [CrossRef]

12. Buyukkaya, E.; Muhammet, C. Experimental study of NOx emissions and injection timing of a low heat rejection diesel engine. *Int. J. Therm. Sci.* **2008**, *47*, 1096–1106. [CrossRef]

13. Fang, H.L.; DaCosta, H.F.M. Urea thermolysis and NOx reduction with and without SCR catalysts. *Appl. Catal. B Environ.* **2003** *46*, 17–34. [CrossRef]

14. Nova, I.; Lietti, L.; Forzatti, P. Mechanistic aspects of the reduction of stored NOx over Pt–Ba/Al$_2$O$_3$ lean NOx trap systems. *Catal. Today* **2008** *136*, 128–135. [CrossRef]

15. Jiao, P.; Li, Z.; Shen, B.; Zhang, W.; Kong, X.; Jiang, R. Research of DPF regeneration with NOx-PM coupled chemical reaction. *Appl. Therm. Eng.* **2017**, 110, 737–745. [CrossRef]

16. Hofmann, L.; Rusch, K.; Fischer, S.; Lemire, B. Onboard Emissions Monitoring on a HD Truck with an SCR System Using Nox Sensors. *J. Fuels Lubr.* **2004** *113*, 559–572.

17. Finesso, R.; Hardy, G.; Maino, C.; Marello, O.; Spessa, E. A New Control-Oriented Semi-Empirical Approach to Predict Engine-Out NOx Emissions in a Euro VI 3.0 L Diesel Engine. *Energies* **2017** *10*, 1978. [CrossRef]

18. Baratta, M.; Kheshtinejad, H.; Laurenzano, D.; Misul, D.; Brunetti, S. Modelling aspects of a CNG injection system to predict its behavior under steady state conditions and throughout driving cycle simulations. *J. Nat. Gas Sci. Eng.* **2015**, *24*, 52–63. [CrossRef]

19. Guardiola, C.; Pla, B.; Blanco-Rodriguez, D.; Calendini, P.O. ECU-oriented models for NOx prediction. Part 1: A mean value engine model for NOx prediction. *J. Automob. Eng.* **2015**, *229*, 992–1015. [CrossRef]

20. Finesso, R.; Misul, D.; Spessa, E. Estimation of the Engine-Out NO2/NOx Ratio in a EURO VI Diesel Engine. *SAE Int.* **2013**. [CrossRef]

21. Atkinson, C.; Mott, G. Dynamic model-based calibration optimization: An introduction and application to diesel engines. In Proceedings of the 2005 SAE World Congress, Detroit, MI, USA, 11–14 April 2005.

22. Donmez, N.; Ozener, O. Modeling of NOx emissions in internal combustion engine. *Int. J. Eng. Res. Adv. Technol.* **2019**, *5*, 36–43. [CrossRef]

23. D'Ambrosio, S.; Finesso, R.; Fu, L.; Mittica, A.; Spessa, E. A control-oriented real-time semi-empirical model for the prediction of NOx emissions in diesel engines. *Appl. Energy* **2014**, *130*, 265–279. [CrossRef]

24. Winkler-Ebner, B.; Hirsch, M.; del Re, L.; Klinger, H.; Mistelberger, W. Comparison of Virtual and Physical NOx-Sensors for Heavy Duty Diesel Engine Application. *SAE Int. J. Engines* **2010**, *3*, 1124–1139. [CrossRef]

25. Stadlbauer, S.; Alberer, D.; Hirsch, M.; Formentin, S.; Benatzky, C.; del Re, L. Evaluation of Virtual NOx Sensor Models for Off Road Heavy Duty Diesel Engines. *SAE Int. J. Commer. Veh.* **2012**, *5*, 128–140. [CrossRef]

26. Fechert, R.; Bäker, B.; Gereke, S.; Atzler, F. Using machine learning methods to develop virtual NOx sensors for vehicle applications. In Proceedings of the Internationales Stuttgarter Symposium, Stuttgart, Germany, 18 August 2020.

27. Kempema, N.; Sharpe, C.; Wu, X.; Shahabi, M.; Kubinski, D. Machine-Learning-Based Emission Models in Gasoline Powertrains Part 2: Virtual Carbon Monoxide. *SAE Int. J. Engines* **2022**, *16*, 2023. [CrossRef]

28. Yuan, X.; Li, L.; Wang, Y. Nonlinear Dynamic Soft Sensor Modeling With Supervised Long Short-Term Memory Network. *IEEE Trans. Ind. Inform.* **2019**, 16, 3168–3176. [CrossRef]

29. Shao, W.; Tian, X. Semi-supervised selective ensemble learning based on distance to model for nonlinear soft sensor development. *Neurocomputing* **2017**, *222*, 91–104. [CrossRef]

30. Liu, C.; Guo, M.; Yang, Y.; Zhang, S. NOx prediction for diesel engine using improved GA-SVR. *Measurement* **2020**, *173*, 108187.

31. Zhang, H.; Zhang, Z.; Liu, Y. NOx emission prediction of diesel engines based on support vector machine. *Measurement* **2018**, *125*, 338–346.

32. Li, X.; Liu, Y.; Zhang, H.; Zhang, Z. NOx prediction for diesel engine using multi-input deep learning network based on LSTM. *Measurement* **2019**, *145*, 36–43

33. Wakjira, T.; Rahmzadeh, A.; Alam, M.S.; Tremblay, R. Explainable machine learning based efficient prediction tool for lateral cyclic response of post-tensioned base rocking steel bridge piers. *Structures* **2022**, *44*, 947–964. [CrossRef]

34. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.

35. Friedman, H.J. Greedy function approximation: A Gradient Boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]

36. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

37. Bentejac, C.; Csorgo, A.; Munoz, M.G. A Comparative Analysis of XGBoost. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [CrossRef]

38. Altug, K.B.; Kucuk, S.E. Predicting Tailpipe NOx Emission using Supervised Learning Algorithms. In Proceedings of the 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies, Ankara, Turkey, 11–13 October 2019.

39. Hu, L.; Wang, C.; Ye, Z.; Wang, S. Estimating gaseous pollutants from bus emissions: A hybrid model based on GRU and XGBoost. *Sci. Total Environ.* **2021**, *783*, 146870. [CrossRef] [PubMed]

40. Great Learning. Hyperparameter Tuning with GridSearchCV. 2022. Available online: https://www.mygreatlearning.com/blog/gridsearchcv/ (accessed on 12 December 2022).

41. Finesso, R.; Marello, O. Calculation of Intake Oxygen Concentration through Intake CO_2 Measurement and Evaluation of Its Effect on Nitrogen Oxide Prediction Accuracy in a Heavy-Duty Diesel Engine. *Energies* **2022**, *15*, 342. [CrossRef]

42. Towards Data Science. Feature Selection Techniques in Machine Learning with Python. 2018. Available online: https://towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python-f24e7da3f36e (accessed on 27 December 2022).

43. Towards Data Science. A Brief Introduction to Supervised Learning. 2019. Available online: https://towardsdatascience.com/a-brief-introduction-to-supervised-learning-54a3e3932590 (accessed on 27 December 2022).

44. Towards Data Science. A Guide to XGBoost Hyperparameters. 2021. Available online: https://towardsdatascience.com/a-guide-to-xgboost-hyperparameters-87980c7f44a9 (accessed on 27 December 2022).

45. DMLC XGBoost. XGBoost Documentation. Available online: https://xgboost.readthedocs.io/en/stable/index.html (accessed on 27 December 2022).

46. Machine Learning Mastery. A Gentle Introduction to k-Fold Cross-Validation. Available online: https://machinelearningmastery.com/k-fold-cross-validation/ (accessed on 8 February 2023).

47. Medium. XGBoost: A BOOSTING Ensemble. Available online: https://medium.com/almabetter/xgboost-a-boosting-ensemble-b273a71de7a8 (accessed on 27 December 2022).

48. Yellowbrick. Feature Importances. 2022. Available online: https://www.scikit-yb.org/en/latest/api/model_selection/importances.html (accessed on 27 December 2022).

49. Peckham, M.; Parnell, J.; Hammond, M.; Mason, B. The Measurement of Fast Transient Emissions During Real World Driving. *Frontiers* **2020**, *6*, 19. [CrossRef]

50. Sellerei, T.; Ferrarese, C.; Franzetti, J.; Suarez-Bertoa, R.; Manara, D. Real-Time Measurement of NOx Emissions from Modern Diesel Vehicles Using On-Board Sensors. *Energies* **2022**, *15*, 8766. [CrossRef]