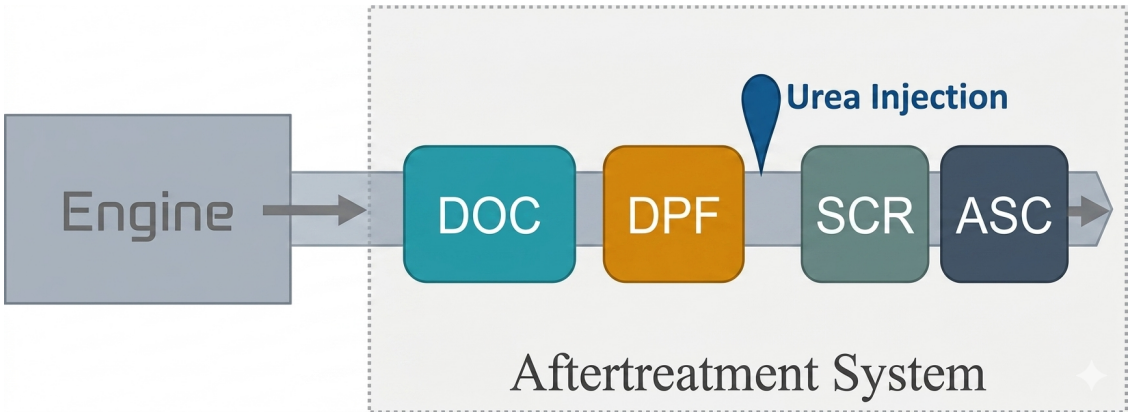
CHALMERS
UNIVERSITY OF TECHNOLOGY



Physics-Informed Machine Learning model for Emission Prediction in Exhaust Aftertreatment Systems

Prediction of NO_x, CO, HC and NH_3 Emissions for Volvo Penta Off-Road Diesel Engines

Master's thesis in Systems, Control and Mechatronics

LINNEA PERSSON

DEPARTMENT OF MECHANICS AND MARITIME SCIENCES

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Master's thesis 2026

Physics-Informed Machine Learning model for Emission Prediction in Exhaust Aftertreatment Systems

Prediction of NO_x, CO, HC and NH_3 Emissions for Volvo Penta Off-Road Diesel Engines

LINNEA PERSSON

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Mechanics and Maritime Sciences
Division of Mechanical Engineering
Chalmers University of Technology
Gothenburg, Sweden 2026

Physics-Informed Machine Learning model for Emission Prediction in Exhaust Aftertreatment Systems
Prediction of NO_x, CO, HC and NH_3 Emissions for Volvo Penta Off-Road Diesel Engines
LINNEA PERSSON

Supervisor: Emma Westerborn & Axel Hansson, Volvo Penta
Examiner: Jonas Sjöblom, Chalmers University of Technology

Master's Thesis 2026
Department of Mechanics and Maritime Sciences
Division of Mechanical Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Simplified illustration of a diesel exhaust aftertreatment system (EATS) consisting of DOC, DPF, SCR, and ASC used for emission reduction in modern industrial diesel engine.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2026

Physics-Informed Machine Learning model for Emission Prediction in Exhaust Aftertreatment Systems
Prediction of NO_x, CO, HC and NH_3 Emissions for Volvo Penta Off-Road Diesel Engines
LINNEA PERSSON
Department of Mechanics and Maritime Sciences
Chalmers University of Technology

Abstract

Modern industrial diesel engines rely on advanced exhaust aftertreatment systems (EATS) to comply with increasingly stringent emission regulations such as EU Stage V, reducing emissions of nitrogen oxides (NO_x), hydrocarbons (HC), carbon monoxide (CO), and ammonia slip (NH_3). Among these systems, selective catalytic reduction (SCR) plays a central role in NO_x reduction through urea-based ammonia dosing. However, SCR development, calibration, and catalyst sizing still depend heavily on expensive and time-consuming physical engine and rig testing. This creates a strong industrial need for reliable predictive models that can complement physical testing through early-stage virtual evaluation.

This thesis investigates the use of physics-informed feature engineering combined with machine learning for predicting system-out emissions in Volvo Penta off-road diesel engine platforms. The objective is to develop a supervised regression framework capable of predicting system-out emissions based on engine-out conditions and exhaust aftertreatment parameters, while improving robustness and physical interpretability compared to purely data-driven approaches. Particular focus is placed on NO_x and NH_3 prediction, since these emissions are most strongly linked to SCR behaviour and catalyst dynamics.

The study uses existing experimental test data from Volvo Penta engines in the D5–D13 platform range, representing industrial and off-road applications. Separate XGBoost regression models were developed for each target variable using both directly measured signals and derived physics-informed features related to SCR behaviour, catalyst thermal conditions, flow dynamics, and ammonia availability. Model performance was evaluated using ShuffleSplit cross-validation together with validation on unseen datasets, primarily using the coefficient of determination (R^2) and Mean Absolute Error (MAE).

The results show that physics-informed feature engineering improved both model accuracy and robustness compared to baseline models using only directly measured inputs. The strongest improvements were observed for NO_x prediction, where physics-informed features improved model accuracy and robustness across different operating conditions and engine platforms. The final models also achieved strong predictive performance for HC and CO, while NH_3 prediction remained more challenging due to the complexity of ammonia storage and slip behavior. Feature importance analysis further confirmed that the learned relationships were consistent with known SCR

physics, improving confidence in model interpretability and engineering relevance.

The developed framework demonstrates that physics-informed feature engineering combined with machine learning can provide a practical and computationally efficient approach for EATS emission prediction. The results also show clear potential for supporting early-stage virtual evaluation, SCR catalyst sizing, and aftertreatment system optimisation within industrial engine development.

Keywords: Physics-informed machine learning, Exhaust aftertreatment systems, Selective Catalytic Reduction, XGBoost, NO_x prediction, SCR sizing, Virtual testing, Volvo Penta

Acknowledgements

I would like to express my sincere gratitude to Volvo Penta for giving me the opportunity to carry out this master's thesis within such an interesting and technically challenging field.

I would especially like to thank the departments of Exhaust After Treatment Systems and CAE Simulations at Volvo Penta, as well as everyone involved within both departments, for their support, valuable discussions, technical insights, and practical help throughout the project.

A special thank you goes to my supervisors, Emma Westerborn and Axel Hansson, for their invaluable support, guidance, and encouragement throughout the entire project. Their knowledge, feedback, and continuous engagement have been essential for this thesis and highly appreciated.

Finally, I would like to thank my examiner, Jonas Sjöblom, for his guidance, valuable feedback, and academic support throughout the thesis process.

Thank you all for your contributions to this master's thesis.

Linnea Persson, Gothenburg, June 2026

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

ANR	Ammonia-to-NO_x Ratio
ASC	Ammonia Slip Catalyst
CO	Carbon monoxide
DOC	Diesel Oxidation Catalyst
DPF	Diesel Particulate Filter
EATS	Exhaust Aftertreatment System
EU	European Union
GHSV	Gas Hourly Space Velocity
HC	Hydrocarbons
H_2O	Water
MAE	Mean Absolute Error
NH_3	Ammonia
NO	Nitric oxide
NO_2	Nitrogen dioxide
NO_x	Nitrogen oxides
NRMM	Non-road Mobile Machinery
N_2O	nitrous oxide
PIML	Physics-Informed Machine Learning
PINN	Physics-Informed Neural Network
PM	Particulate Matter
PN	Particle Number
R^2	Coefficient of Determination
RMSE	Root Mean Square Error
SCR	Selective Catalytic Reduction

Contents

List of Figures

List of Tables

1
Introduction

This chapter introduces the background and industrial motivation of the thesis, focusing on emission prediction and exhaust aftertreatment system development for Volvo Penta industrial diesel engines under increasingly stringent emission regulations. Furthermore, the problem formulation, research aim, research questions, and scope of the study are presented.

1.1 Background

Increasingly stringent global emission regulations have made exhaust aftertreatment systems (EATS) a critical part of modern industrial and off-road diesel engine development [1, 2]. For manufacturers such as Volvo Penta, compliance with regulations such as EU Stage V requires effective control of emissions including nitrogen oxides (NO_x), hydrocarbons (HC), carbon monoxide (CO), particulate matter (PM), and particle number (PN), while maintaining both environmental performance and product competitiveness [3, 4]. In addition to regulated emissions, ammonia slip (NH_3) must also be carefully controlled, since excessive NH_3 release reduces overall environmental performance and indicates suboptimal SCR catalyst operation and dosing control [5].

A modern EATS typically consists of several integrated components, including a diesel oxidation catalyst (DOC), diesel particulate filter (DPF), selective catalytic reduction (SCR) system, and ammonia slip catalyst (ASC), which together reduce harmful emissions before they are released into the atmosphere [6]. Among these, the selective catalytic reduction (SCR) system is particularly important, as it is primarily responsible for NO_x reduction and strongly influences both catalyst sizing and urea dosing strategy [5]. Since compliance with NO_x emission limits remains one of the most critical regulatory requirements for modern diesel engines, SCR system performance becomes central to both engine development and final product validation [7].

To reduce development time, cost, and dependency on extensive physical testing, virtual testing has become increasingly important in industrial engine development. Existing virtual testing frameworks are widely used for combustion analysis, calibration, and performance evaluation, while EATS are often represented using simplified assumptions or incomplete sub-models, making accurate prediction of catalyst efficiency, tailpipe emissions, and overall aftertreatment performance more challenging

[7]. This limits the ability to fully evaluate catalyst behavior and emission compliance under realistic operating conditions.

At the same time, large amounts of experimental data from combined engine and aftertreatment test campaigns are available within industrial development environments. These datasets contain valuable information on engine-out emissions, exhaust temperatures, catalyst loading, urea dosing, and system-out emissions across a wide range of operating conditions [7]. This creates an opportunity to improve predictive modelling of exhaust aftertreatment behaviour through data-driven and physics-informed modelling approaches [7, 8].

Reliable emission prediction models could support several important engineering tasks, including virtual testing, calibration transfer between engine platforms, and early-stage catalyst sizing studies before costly physical prototype testing is required [7]. In particular, predicting how changes in catalyst volume affect NO_x conversion and NH_3 slip could provide valuable guidance for determining the minimum catalyst size required to meet emission legislation without unnecessary material cost or packaging constraints [5].

1.2 Problem Formulation

As described in the background, accurate prediction of exhaust aftertreatment performance is essential for both emission compliance and catalyst dimensioning. However, current modelling approaches still present major limitations when applied to industrial development and virtual testing.

Traditional first-principles models can describe catalyst behaviour with high physical accuracy since they are based on reaction kinetics, thermodynamics, and transport phenomena. However, they are often computationally expensive, highly complex, and difficult to calibrate for industrial applications involving multiple engine platforms and large operating ranges [7]. In addition, they typically require catalyst-specific parameters that are difficult to obtain during early design phases, making them less suitable for rapid engineering decisions and large-scale virtual testing.

Purely data-driven machine learning models offer an attractive alternative because they can learn directly from available test data and provide fast prediction performance. However, such models often behave as black-box systems with limited interpretability and reduced reliability when operating outside the conditions represented in the training data [7, 8]. This limits their robustness and extrapolation capability in industrial applications where physical consistency is essential.

Physics-informed machine learning (PIML) offers a potential solution by combining data-driven modelling with known physical and chemical principles, such as reaction kinetics, thermal dynamics, and flow behaviour [9, 8]. By embedding physical knowledge into machine learning models, it may be possible to achieve accurate and generalizable representations of exhaust aftertreatment systems while reducing

calibration effort and computational cost. This improves prediction accuracy, robustness, and interpretability while maintaining computational efficiency suitable for industrial implementation.

For Volvo Penta, this challenge is directly connected to practical engineering decisions such as catalyst sizing, urea dosing calibration, and validation of emission compliance across multiple engine platforms. A more reliable modelling framework could reduce dependency on expensive rig testing, improve virtual testing capability, and provide stronger support for future SCR upscaling or downsizing studies.

The central challenge addressed in this thesis is therefore to develop a machine learning model for system-out emission prediction that remains both accurate and physically meaningful while being robust enough for industrial use. In particular, the work investigates whether physics-informed feature engineering can improve prediction performance for NO_x, CO, HC, and NH_3 emissions compared to purely data-driven approaches, and whether such a model can provide a reliable foundation for future virtual testing and SCR sizing applications.

1.3 Aim

The aim of this thesis is to develop and evaluate a physics-informed machine learning model for predicting system-out emissions in EATS for Volvo Penta industrial off-road diesel engines. The work focuses on predicting NO_x, NH_3, HC, and CO emissions downstream of the aftertreatment system using existing experimental engine and EATS test data from multiple engine platforms within the D5–D13 range.

A central objective is to investigate whether the incorporation of physical and chemical domain knowledge can improve prediction accuracy, robustness, and generalisation compared to purely data-driven machine learning approaches. This is achieved by comparing a baseline model using directly measured input variables with a physics-informed model including engineered features derived from SCR-related physical relationships such as catalyst residence time, ammonia availability, thermal behaviour, and temperature-dependent reaction kinetics.

In the longer term, the developed modelling framework is intended to support future virtual testing and SCR sizing studies, where accurate prediction of NO_x conversion and NH_3 slip is essential for balancing emission compliance, catalyst volume, packaging constraints, and system cost.

1.4 Research Questions

Based on the aim of this thesis, the following research questions are addressed:

- How can machine learning models be developed to accurately predict system-out NO_x, NH_3, HC, and CO emissions in exhaust aftertreatment systems using existing experimental data from Volvo Penta industrial diesel engines?

- Which engine and exhaust aftertreatment parameters are most relevant for reliable emission prediction, and how should the experimental data be preprocessed and structured for machine learning applications?

- How does physics-informed feature engineering influence prediction accuracy, robustness, and generalisation compared to purely data-driven machine learning models, particularly for SCR-related emissions such as NO_x and NH_3?

- Can the developed modelling framework provide a reliable foundation for future virtual testing and preliminary SCR sizing studies, including catalyst upscaling and downsizing based on emission compliance requirements?

1.5 Scope and Limitations

This thesis is based entirely on existing experimental engine and exhaust aftertreatment system data provided by Volvo Penta. No new engine testing, laboratory measurements, or physical validation of resized catalyst systems were conducted. All model development, training, and evaluation therefore depend on the quality, availability, and operating range of the existing datasets.

The study focuses on selected Volvo Penta industrial diesel engine platforms within the D5–D13 range and includes only diesel engines with diesel-based exhaust aftertreatment systems. Alternative fuels, hybrid powertrains, gasoline engines, and other propulsion concepts are outside the scope of this work.

The primary focus is placed on predicting system-out NO_x, NH_3, HC, and CO emissions, with particular emphasis on SCR-related emissions such as NO_x conversion and ammonia slip. The study included three representative EATS size configurations, selected based on data availability and industrial relevance.

The work focuses on machine learning-based prediction models and physics-informed feature engineering rather than detailed first-principles reactor models or high-fidelity simulations such as CFD. Only one machine learning model structure (XGBoost) was evaluated, using two model configurations: a purely data-driven baseline model and a physics-informed model including additional engineered features derived from SCR-related physical and chemical relationships. Advanced deep learning approaches such as Physics-Informed Neural Networks (PINNs), as well as full

optimisation of model architecture, were not included within the scope of this thesis.

The validation strategy was based on predefined training and validation datasets across the different engine platforms together with ShuffleSplit cross-validation during model development. More advanced validation strategies such as leave-one-engine-out validation or GroupKFold were not included within the scope of this work.

The developed models were evaluated in an offline simulation environment. Real-time implementation, embedded deployment, controller integration, certification-related validation, and full catalyst redesign or validated SCR dimensioning were outside the scope of the study, although the developed framework was evaluated as a foundation for future SCR sizing studies through model-based catalyst volume sensitivity analysis.

Finally, industrial confidentiality limits the level of technical detail, catalyst specifications, and data transparency that can be presented in the final report.

2
Theory

This chapter presents the theoretical background required to understand the methodology and results of this thesis. The focus is placed on diesel engine emissions, exhaust aftertreatment systems (EATS), and modeling approaches for SCR-related emission prediction.

The chapter begins with an overview of emission legislation for non-road mobile machinery (NRMM), with particular focus on the EU Stage V regulation relevant for the Volvo Penta engine platforms considered in this work. This is followed by diesel engine emissions and the main EATS components, particularly the SCR system.

The chapter also introduces the Volvo Penta engine platforms included in the study together with the modeling approaches used for exhaust aftertreatment prediction. Special attention is given to physics-informed feature engineering and XGBoost-based modeling, since these form the methodological foundation of the developed model.

2.1 Emission Legislation

Emission legislation sets legal limits for pollutants released from internal combustion engines in order to protect human health, improve air quality, and reduce environmental impact [2, 10]. Increasingly strict emission limits have driven the development of modern diesel engines and advanced EATS capable of reducing harmful emissions [1, 6].

This thesis focuses on industrial and off-road diesel engines operating under the European Union regulations for NRMM. Emission limits vary depending on engine application and regulation category, such as on-road vehicles, marine engines, stationary power generation, and non-road machinery [10, 1].

EU Stage V is the current emission standard for NRMM engines and the most relevant regulation for the Volvo Penta engine platforms considered in this thesis. It sets strict limits for pollutants such as NO_x, hydrocarbons (HC), carbon monoxide (CO), particulate matter (PM), and particle number (PN), making advanced aftertreatment systems necessary for compliance [10, 1]. A more detailed description of the EU Stage V regulation is presented in Section 2.1.3.

As emission limits become stricter, accurate aftertreatment modeling becomes increasingly important for predicting NO_x conversion, NH_3 slip, and catalyst behaviour under realistic operating conditions [5, 7].

2.1.1 Purpose of Emission Legislation

For diesel engine manufacturers, emission legislation creates a continuous need to improve both combustion performance and exhaust aftertreatment efficiency. One of the main engineering challenges is the trade-off between reducing NO_x emissions and limiting particulate emissions. Lower combustion temperatures can reduce NO_x formation, but often increase soot, HC, and CO emissions [3]. Emission compliance therefore depends on both engine calibration and aftertreatment performance, rather than combustion improvements alone [1, 6].

For modern NRMM engines under EU Stage V, emission compliance directly affects catalyst sizing and urea dosing strategies. The aftertreatment system must achieve sufficient NO_x reduction without causing excessive NH_3 emissions or unnecessary catalyst oversizing [5]. Since larger catalyst volumes increase both cost and packaging complexity, accurate prediction of catalyst performance becomes important for efficient system design [7].

As emission limits become stricter, manufacturers must ensure compliance under increasingly realistic operating conditions rather than only during certification testing. This increases the importance of predictive modeling and virtual testing, which can support earlier optimization of aftertreatment systems before expensive physical testing is performed [7].

2.1.2 Regulated Emissions

The main regulated emissions relevant for diesel engines are summarized below:

- **Nitrogen oxides (NO_x):** Mainly nitric oxide (NO) and nitrogen dioxide (NO_2), formed at high combustion temperatures. NO_x is one of the most important regulated emissions in diesel engines due to its environmental and health impacts. Since lean diesel combustion naturally promotes NO_x formation, efficient aftertreatment is required for emission compliance [3].

- **Hydrocarbons (HC):** Unburned fuel components caused by incomplete combustion, especially at low combustion temperatures or unstable operating conditions [3].

- **Carbon monoxide (CO):** Formed during incomplete combustion when there is not enough oxygen for full oxidation to CO_2.[11, 3].

- **Ammonia (NH_3):** An unwanted emission that may occur during NO_x reduction. Excess NH_3 emissions are commonly referred to as ammonia slip [5].

- **Particulate matter (PM):** Mainly soot particles formed during incomplete combustion in regions with limited oxygen. PM emissions are strictly regulated due to their environmental and health impacts [1].

2.1.3 EU Stage V Emission Standard

EU Stage V was introduced through Regulation (EU) 2016/1628 and is the current emission standard for non-road diesel engines in the European market [10, 1].

The regulation replaced the previous Stage IV standard and was introduced progressively from 2019 depending on engine category and power class [10, 1].

One of the major changes in Stage V compared to Stage IV is the introduction of stricter particulate matter (PM) and particle number (PN) limits, together with continued strict NO_x requirements [1, 13].

Table 2.1 summarizes the Stage V emission limits for the compression ignition engine categories most relevant to the Volvo Penta engine platforms considered in this thesis, particularly the 56–130 kW and 130–560 kW power classes corresponding to the D5, D8, D11, and D13 engines.

Table 2.1: EU Stage V emission limits for compression ignition NRMM engines relevant to the engine platforms considered in this thesis. Adapted from Regulation (EU) 2016/1628 and DieselNet [10, 1].

Net Power (kW)	**CO (g/kWh)**	**HC (g/kWh)**	**NO_x (g/kWh)**	**PM (g/kWh)**	**PN (#/kWh)**
$19 \leq P < 37$	5.0	4.7 (HC + NO_x)		0.015	1×10^{12}
$37 \leq P < 56$	5.0	4.7 (HC + NO_x)		0.015	1×10^{12}
$56 \leq P < 130$	5.0	0.19	0.40	0.015	1×10^{12}
$130 \leq P \leq 560$	3.5	0.19	0.40	0.015	1×10^{12}
$P > 560$	3.5	0.19	3.5	0.045	–

2.1.4 Future Emission Requirements

Emission legislation for non-road diesel engines is expected to become stricter beyond the current EU Stage V regulation [12, 1]. Future requirements are expected to place greater focus on transient operation, real-world emissions, cold-start performance, and long-term compliance under realistic operating conditions rather than only certification testing [12].

In addition to conventional pollutants such as NO_x, PM, HC, and CO, future regulations are also expected to increase attention on secondary emissions and overall

system efficiency. This increases the importance of predictive modeling and virtual testing for efficient development and optimization of modern aftertreatment systems.

2.2 Diesel Engine and Engine-Out Emissions

Diesel engine operating conditions strongly influence exhaust aftertreatment performance. Parameters such as engine-out NO_x, exhaust gas temperature, exhaust mass flow, and the NO/NO_2 ratio directly affect the efficiency of downstream aftertreatment systems [3, 5].

Diesel engines use compression ignition and typically operate under lean combustion conditions with high combustion temperatures and excess oxygen, which promote NO_x formation, mainly NO and NO_2 [16, 3].

Strategies used to reduce NO_x emissions often increase PM emissions, creating the well-known NO_x–PM trade-off in diesel engines [14]. Aftertreatment performance depends strongly on engine-out emissions, exhaust temperature, and exhaust flow, since these define the operating conditions entering the aftertreatment system [5].

2.3 Exhaust Aftertreatment System (EATS)

A typical diesel EATS consists of several components arranged in series, where each component targets different pollutants using catalytic or physical processes. The main components include Diesel Oxidation Catalysts (DOC), Diesel Particulate Filters (DPF), Selective Catalytic Reduction (SCR) systems, and Ammonia Slip Catalysts (ASC) [17].

Figure 2.1 illustrates a typical EATS layout used in modern industrial diesel engines.

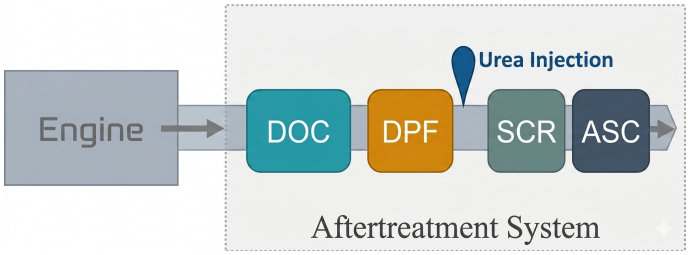


Figure 2.1: Typical EATS layout consisting of DOC, DPF, SCR, and ASC.

2.3.1 Diesel Oxidation Catalyst (DOC)

The DOC is typically the first component in the EATS, placed directly downstream of the engine, and is primarily used to oxidize CO and unburned HC emissions. It also converts part of the NO in the exhaust into NO_2 [11, 6].

The main oxidation reactions in the DOC are [11, 6]:

Carbon monoxide oxidation

$$\mathrm{CO} + \tfrac{1}{2}\mathrm{O_2} \rightarrow \mathrm{CO_2} \tag{2.1}$$

Hydrocarbon oxidation

$$\mathrm{C_xH_y} + \left(\mathrm{x} + \tfrac{\mathrm{y}}{4}\right)\mathrm{O_2} \rightarrow \mathrm{xCO_2} + \tfrac{\mathrm{y}}{2}\mathrm{H_2O} \tag{2.2}$$

Nitric oxide oxidation

$$\mathrm{NO} + \tfrac{1}{2}\mathrm{O_2} \rightarrow \mathrm{NO_2} \tag{2.3}$$

Although the DOC does not reduce total NO_x, the produced NO_2 affects downstream aftertreatment performance and contributes to improved NO_x conversion at lower temperatures [5, 18].

DOC performance is strongly temperature dependent. Efficient oxidation of CO and HC typically requires exhaust temperatures above approximately 200 °C, while lower temperatures significantly reduce conversion efficiency [11].

2.3.2 Diesel Particulate Filter (DPF)

The DPF is used to reduce PM emissions by capturing soot particles from the exhaust gas before they are released to the atmosphere [18, 1]. DPF is typically placed downstream of the DOC, where the NO_2 produced in the DOC can support passive soot oxidation at lower temperatures [11, 18].

As soot accumulates inside the filter, regeneration is required to remove the stored soot and maintain acceptable exhaust backpressure [18]. The main soot oxidation reactions are [18]:

Oxidation by oxygen

$$\mathrm{C_{(s)}} + \mathrm{O_2} \rightarrow \mathrm{CO_2} \tag{2.4}$$

NO_2-assisted oxidation

$$\mathrm{C_{(s)}} + \mathrm{NO_2} \rightarrow \mathrm{CO} + \mathrm{NO} \tag{2.5}$$

$$\mathrm{CO} + \tfrac{1}{2}\mathrm{O_2} \rightarrow \mathrm{CO_2} \tag{2.6}$$

The DPF also influences downstream aftertreatment conditions through its effects on exhaust temperature, pressure drop, and exhaust composition [5, 18].

2.3.3 Selective Catalytic Reduction (SCR)

SCR is the main aftertreatment technology used to reduce NO_x emissions in modern diesel engines [5, 2]. The SCR catalyst is typically placed downstream of the DOC and DPF, where exhaust temperature and gas composition are suitable for efficient

NO_x conversion. SCR performance depends strongly on engine-out NO_x concentration, exhaust temperature, exhaust mass flow, catalyst volume, residence time, and the NO/NO_2 ratio entering the catalyst [5, 19].

The SCR system works by injecting an aqueous urea solution upstream of the catalyst. The urea decomposes and forms NH_3, which acts as the reducing agent. Over the catalyst surface, NO_x is selectively converted into nitrogen (N_2) and water (H_2O) in the presence of excess oxygen [5, 20].

Several reaction pathways contribute to NO_x reduction depending on the relative amounts of NO and NO_2 entering the catalyst [5]:

Standard SCR reaction (NO-dominated exhaust)

$$4NO + 4NH_3 + O_2 \rightarrow 4N_2 + 6H_2O \tag{2.7}$$

Fast SCR reaction (NO/NO_2 mixture)

$$NO + NO_2 + 2NH_3 \rightarrow 2N_2 + 3H_2O \tag{2.8}$$

NO_2 SCR reaction

$$6NO_2 + 8NH_3 \rightarrow 7N_2 + 12H_2O \tag{2.9}$$

Among these reactions, the fast SCR reaction provides the highest NO_x conversion efficiency at lower temperatures. The NO/NO_2 ratio entering the catalyst therefore strongly affects SCR performance [5, 11].

SCR efficiency is strongly temperature dependent. At low temperatures, urea decomposition and NO_x conversion become less effective, while high temperatures may lead to unwanted ammonia oxidation [5, 20]. The catalyst therefore has a limited effective operating window, making exhaust temperature and dosing control important.

Exhaust mass flow and catalyst volume also influence SCR performance by affecting residence time inside the catalyst. Higher exhaust flow increases Gas Hourly Space Velocity (GHSV), which reduces the available reaction time for NO_x reduction. Larger catalyst volumes can improve NO_x conversion efficiency but also increase system cost and packaging requirements [15, 5].

2.3.3.1 Urea Dosing

After injection upstream of the SCR catalyst, the urea solution decomposes and forms NH_3 through temperature-dependent reactions [20].

Thermal decomposition of urea

$$(NH_2)_2CO \rightarrow NH_3 + HNCO \tag{2.10}$$

Hydrolysis of isocyanic acid

$$HNCO + H_2O \rightarrow NH_3 + CO_2 \tag{2.11}$$

These reactions determine the amount of ammonia available for downstream SCR reactions [5, 20]. At low temperatures, incomplete decomposition may reduce NO_x conversion efficiency and increase the risk of deposit formation [20].

The amount of injected urea must be carefully controlled to achieve high NO_x conversion without causing excessive NH_3 slip. Too little dosing results in poor NO_x reduction, while excessive dosing may increase NH_3 emissions [5, 15].

Under transient operating conditions, urea dosing becomes more challenging due to variations in exhaust temperature, exhaust flow, and engine-out NO_x [15, 20].

2.3.3.2 Ammonia Storage and Ammonia-to-NO_x Ratio (ANR)

An important property of the SCR catalyst is its ability to temporarily store ammonia (NH_3) on the catalyst surface. This stored ammonia is used during NO_x reduction and helps the system handle changes in engine-out NO_x and operating conditions [5, 15]. The catalyst therefore acts both as a reaction surface and as a temporary NH_3 buffer.

If too little ammonia is stored, NO_x conversion decreases and tailpipe NO_x emissions increase. If too much ammonia is stored, the risk of NH_3 slip increases [5]. This balance is commonly described using the Ammonia-to-NO_x Ratio (ANR), which represents the relationship between available ammonia and incoming NO_x. An ANR close to 1.0 corresponds approximately to stoichiometric conditions for NO_x reduction, while higher values increase the risk of NH_3 slip [15].

Ammonia storage behaviour depends strongly on operating conditions such as exhaust temperature, exhaust flow, and the NO/NO_2 ratio entering the catalyst [5]. Under transient operation, rapidly changing engine-out conditions make ammonia storage behaviour more difficult to control [15].

2.3.4 Ammonia Slip Catalyst (ASC)

Unreacted NH_3 may pass through the SCR catalyst and be released to the atmosphere, a phenomenon known as ammonia slip [5]. To reduce NH_3 emissions, an ASC is typically placed downstream of the SCR catalyst as the final component in the EATS. Its main function is to oxidise remaining NH_3 into mainly N_2 and H_2O [5].

The main desired reaction in the ASC is [5]:

Selective ammonia oxidation

$$4NH_3 + 3O_2 \rightarrow 2N_2 + 6H_2O \tag{2.12}$$

Under unfavourable operating conditions, unwanted side reactions may also occur, forming secondary pollutants such as N_2O and NO [5].

Formation of nitrous oxide

$$2NH_3 + 2O_2 \rightarrow N_2O + 3H_2O \tag{2.13}$$

Formation of nitric oxide

$$4NH_3 + 5O_2 \rightarrow 4NO + 6H_2O \tag{2.14}$$

ASC performance is closely linked to the upstream SCR system and its dosing strategy. Excessive urea dosing may improve NO_x conversion but also increase NH_3 slip and the risk of unwanted side reactions [5].

2.4 Engine Platforms and SCR Sizing

This work considers the Volvo Penta D5, D8, D11, and D13 engine platforms, representing medium- to heavy-duty industrial applications with different exhaust flow rates, thermal conditions, and aftertreatment system sizes [21]. These differences make the platforms relevant for studying SCR behaviour across varying operating conditions.

The D5 is the smallest platform included in this work, while the D13 is the largest and operates with higher exhaust flow and thermal load. The D8 and D11 represent intermediate platforms [21, 5].

As engine size increases, exhaust mass flow and thermal load also increase, which reduces catalyst residence time and increases the required SCR volume to maintain sufficient NO_x conversion. Larger engines therefore require different catalyst sizing strategies than smaller platforms [5, 15].

Figure 2.2 shows the four Volvo Penta engine platforms considered in this thesis: D5, D8, D11, and D13.

Figure 2.2: Volvo Penta engine platforms included in this thesis: D5, D8, D11, and D13.

Table 2.2 summarizes the main technical characteristics of the engine platforms together with their associated SCR catalyst volumes [21].

Table 2.2: Technical overview of the Volvo Penta engine platforms and associated SCR catalyst volumes considered in this thesis.

Engine	Cylinders	Displacement [L]	Power Range [kW]	SCR Volume [L]
D5	4	5.1	105–175	22.8
D8	6	7.7	160–235	34.2
D11	6	10.8	240–315	34.2
D13	6	12.8	320–405	45.6

2.5 Modeling Approaches

Modeling is an important tool in the development of modern diesel engines and EATS, particularly for predicting emissions and evaluating aftertreatment performance under varying operating conditions [7].

For SCR systems, the relationship between engine-out NO_x, exhaust temperature, exhaust mass flow, urea dosing, and catalyst behaviour is highly nonlinear and strongly coupled. Predictive models are therefore useful for analyzing system behaviour, supporting catalyst sizing, and estimating NO_x conversion and NH_3 slip under different operating conditions [5, 15].

Several modeling approaches are used in engine and aftertreatment development. Physics-based models describe reaction chemistry, heat transfer, mass transport, and catalyst kinetics using first-principles equations. These models provide strong physical interpretability but often require many parameters and high computational effort [7]. Data-driven models instead learn system behaviour directly from measurement data and can provide strong predictive performance for complex nonlinear systems [8, 7].

In recent years, physics-informed approaches have become increasingly common. By incorporating physically meaningful variables and engineering knowledge into data-driven models, it is possible to improve robustness and maintain physically reasonable behaviour while preserving the flexibility of machine learning methods [8, 9]. This is particularly relevant for SCR systems, where both physical understanding and large amounts of experimental data are available.

2.5.1 SCR System Complexity and Modeling Challenges

SCR performance depends on several strongly coupled operating conditions, including engine-out NO_x, exhaust temperature, exhaust mass flow, and ammonia availability. Variations in these conditions strongly affect NO_x conversion efficiency and NH_3 slip behaviour, especially during transient operation and low-temperature conditions [5, 15].

Catalyst sizing is an important engineering trade-off. Larger catalyst volumes can

improve NO_x conversion efficiency, but also increase system cost and packaging requirements. Reliable prediction of SCR performance is therefore important for efficient aftertreatment development and catalyst sizing [5, 7].

Traditionally, SCR calibration and catalyst sizing rely heavily on engine test bench measurements and physical prototype validation. While these methods provide reliable results, they require significant time and engineering resources, particularly when multiple engine platforms and operating conditions must be evaluated [7].

Predictive modeling can support earlier design decisions and reduce the need for extensive physical testing. In particular, physics-informed approaches can combine experimental data with physically meaningful relationships to improve both prediction accuracy and engineering robustness [8, 9].

2.5.2 Steady-State and Transient Modeling

Modeling of diesel engines and EATS can generally be divided into steady-state and transient approaches. The choice affects model complexity, required input data, and computational cost [22].

A steady-state model assumes approximately constant operating conditions for a given operating point, where variables such as exhaust temperature, exhaust flow, and engine-out NO_x remain stable over time. Dynamic effects are simplified or neglected, and the model therefore describes the system after transient behaviour has settled [22].

Transient modeling instead describes time-dependent system behaviour during changing operating conditions. This provides a more complete representation of real engine operation, but also increases model complexity and calibration requirements [23, 19].

Steady-state modeling is particularly useful for catalyst sizing, operating point analysis, and comparison between different engine platforms. Since industrial test data are often collected around stationary operating points, steady-state approaches also fit well with regression-based model development using experimental data [7].

2.5.3 Calibration Data and Data Quality

The quality and coverage of the calibration data strongly affect predictive model performance. For SCR systems, the dataset must represent a wide range of operating conditions in order to capture variations in aftertreatment behaviour across different engine loads and thermal conditions [7].

Engine test data are often concentrated around common operating regions, which can create an unbalanced dataset with limited coverage of less frequent operating conditions. A model trained on such data may perform well near dominant operat-

ing points but poorly under other conditions [8, 7].

A useful calibration dataset should therefore provide broad operating point coverage and avoid overrepresentation of highly similar samples. In many cases, a smaller but well-distributed dataset can provide better model performance than a larger dataset with poor operating range coverage [8, 7].

Before model development, the data must also be validated carefully. Raw engine test data may contain measurement errors, missing values, unstable operating points, sensor drift, and physically unrealistic observations [23, 7]. For steady-state SCR modeling, transient operating periods and unstable exhaust conditions should typically be removed. If unrealistic or inconsistent data are included, the model may learn non-physical relationships instead of true system behaviour [7].

Filtering is therefore an important step in improving model robustness. Threshold-based filtering can remove clearly invalid values, while clustering and outlier detection methods can help improve dataset balance and reduce overrepresentation of highly similar operating points [24, 8, 7].

2.5.4 Data-Driven Modeling and XGBoost

Data-driven modeling describes approaches where the relationship between system inputs and outputs is learned directly from measured data instead of being derived from first-principles equations [8, 26]. In SCR applications, this enables prediction of quantities such as NO_x conversion, tailpipe emissions, and NH_3 slip using experimental operating data.

XGBoost (Extreme Gradient Boosting) is a gradient boosting method that builds the final prediction from many sequentially trained decision trees, where each new tree reduces the remaining prediction error from the previous trees. It is particularly suitable for industrial tabular datasets because it handles nonlinear relationships, feature interactions, and noisy measurement data well while requiring limited preprocessing [26, 27].

XGBoost also provides feature importance analysis, which can help identify the most influential variables for emission prediction and model interpretation [26].

However, purely data-driven models also have limitations. Prediction quality depends strongly on dataset quality, operating point coverage, and measurement reliability. In addition, black-box models may produce physically unrealistic predictions outside the training region or when important physical constraints are not represented [8, 7, 25].

For this reason, data-driven models are often combined with physical understanding through physics-informed approaches. This can improve robustness while maintaining strong predictive performance for industrial emission modeling [8, 9].

2.5.5 Physics-Informed Modeling

Physics-informed modeling combines data-driven methods with physical understanding of the system. Instead of relying only on raw measured signals, physically meaningful variables and engineering relationships can be incorporated to improve model robustness, interpretability, and extrapolation capability [8, 25].

For SCR systems, this includes variables related to catalyst loading, exhaust flow, temperature, and ammonia availability. Physics-informed approaches therefore provide a practical balance between physical understanding and computational efficiency for industrial emission modeling [28, 29].

2.6 Feature Engineering Based on Reaction Chemistry

For SCR modeling, physically meaningful input variables are important for improving prediction accuracy and model robustness. Instead of relying only on raw measured signals, additional variables can be constructed from known relationships in the aftertreatment system. These engineered features provide a better representation of catalyst behaviour and improve model interpretability [28, 8].

One important feature group is the incoming NO_x load. Since the catalyst must reduce the total NO_x entering the system, concentration alone is often not enough. A more useful variable is the NO_x mass flow, which combines exhaust flow and NO_x concentration and better represents the actual reduction demand [5, 30].

Another key feature is the ANR, which describes the relationship between available reducing agent and incoming NO_x load [31]:

$$ANR = \frac{\dot{m}_{NH_3,\text{proxy}}}{\dot{m}_{NO_x}} \tag{2.15}$$

where $\dot{m}_{NH_3,\text{proxy}}$ represents the ammonia or urea-related quantity and $\dot{m}_{NO_x}$ is the incoming NO_x mass flow. ANR close to stoichiometric conditions is required for efficient NO_x reduction, while excessive values increase the risk of NH_3 slip [31, 5].

The NO_2 fraction is important because SCR performance depends strongly on the NO/NO_2 ratio entering the catalyst [5].

Temperature-based features are important since catalyst activity and reaction behaviour strongly depend on temperature [5].

Flow- and catalyst-size-related features are also important. GHSV is commonly used to describe catalyst loading and is defined as [5]:

$$GHSV = \frac{\dot{V}_{exhaust}}{V_{cat}} \tag{2.16}$$

where $\dot{V}_{exhaust}$ is the volumetric exhaust flow and V_{cat} is the catalyst volume. Higher GHSV means shorter residence time inside the catalyst, which may reduce NO_x conversion efficiency [15].

Simplified history-related features such as smoothed catalyst temperature or accumulated dosing can also be useful as proxies for thermal inertia and ammonia storage without requiring a full transient catalyst model [23].

2.7 Performance Evaluation Metrics

To evaluate the predictive performance of machine learning models, suitable error metrics are required. Since this thesis focuses on regression for SCR emission prediction, the most relevant metrics are Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2) [32].

MAE measures the average absolute difference between predicted and measured values and is defined as [32]:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2.17}$$

where y_i is the measured value, $\hat{y}_i$ is the predicted value, and n is the total number of observations.

RMSE is defined as [32]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.18}$$

RMSE gives greater weight to large prediction errors because the residuals are squared before averaging.

The coefficient of determination (R^2) is given by [32]:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2.19}$$

where $\bar{y}$ is the mean of the measured values. An R^2 value closer to 1 indicates better predictive performance.

For SCR modeling, these metrics allow comparison of prediction quality across different engine platforms and operating conditions. MAE gives a measure of average error, RMSE highlights large deviations, and R^2 provides an overall measure of model fit.

3
Methods

This thesis was carried out in close collaboration with Volvo Penta and followed a structured, data-driven modelling approach. The methodology was designed to evaluate the use of physics-informed machine learning for EATS modelling using existing experimental engine and aftertreatment data.

The work focused on selecting relevant test data, preprocessing large measurement datasets, creating physics-informed features, and developing machine learning models for predicting downstream emissions, with primary focus on SCR-related NO_x conversion and ammonia slip. The analysis and model development were performed in Python using common data processing and machine learning libraries together with Volvo Penta's internal data platforms and measurement tools.

The methodology combined engineering knowledge of catalyst behaviour with supervised regression modelling to improve both prediction accuracy and physical understanding. This created the basis for evaluating the potential of machine learning as a support tool for virtual testing, catalyst sizing, and EATS development.

3.1 Literature Review

A literature review was conducted to establish the theoretical and methodological foundation of the thesis. The review focused on diesel engine emissions, EATS, SCR chemistry, and machine learning approaches for emission prediction and catalyst modeling.

The review included peer-reviewed journal articles, conference papers, textbooks, technical reports, emission legislation documents, and industrial technical resources. Scientific publications related to machine learning, SCR performance, hybrid modeling, and emission prediction were primarily collected from databases such as ScienceDirect, IEEE Xplore, SpringerLink, and Google Scholar. The literature was used to support the selection of modeling methods, feature engineering strategies, and evaluation approaches used in this work.

The technical background related to diesel emissions, catalyst chemistry, and exhaust aftertreatment technologies was supplemented using established industrial and technical sources such as DieselNet, EU emission legislation documents, and technical documentation from engine and aftertreatment manufacturers. These sources

were primarily used to describe DOC, DPF, SCR, and ASC systems together with the mechanisms influencing NO_x, CO, HC, NH_3, and particulate emissions.

The literature review mainly focused on:

- diesel emissions and exhaust aftertreatment systems,
- SCR system behaviour and catalyst performance,
- steady-state and transient modeling approaches,
- data-driven and physics-informed machine learning methods for industrial system modeling,
- feature engineering strategies for exhaust aftertreatment prediction,
- model evaluation and validation methods for machine learning models trained on engine test data.

The collected literature was used to support the selection of input variables, preprocessing methods, feature engineering, model structure, and evaluation methods used in the thesis. It also provided the theoretical basis for Chapter 2 and supported the interpretation of the final modeling results.

3.2 Overall Modeling Workflow

The complete modelling workflow used in this thesis is summarised in Table 3.1. The workflow includes data extraction, preprocessing, physics-informed feature engineering, machine learning model development, validation, and SCR sensitivity evaluation.

The following sections describe each step of the workflow in more detail.

Table 3.1: Overview of the modelling workflow used in this thesis.

Workflow Step	Description
Raw test data	Volvo Penta engine and EATS measurement data stored in CAPE and Sakura XML files. The dataset included both transient M1 measurements and quasi-stationary LS measurements from the D5, D8, D11, and D13 engine platforms.
Data extraction and signal selection	Relevant measurement signals were extracted using the PySakura library and converted into structured pandas DataFrames. Alias mapping was used to standardise signal names across engine platforms and test configurations.
Preprocessing pipeline	EWMA filtering was applied to reduce measurement noise while preserving operating trends. Transient M1 datasets were reduced using k-means clustering in operating space, while LS measurements were retained at original resolution after filtering.
Unified dataset creation	Processed transient and stationary operating points were combined into one unified machine learning dataset where each row represented one representative operating condition.
Model inputs and targets	Engine-out variables such as exhaust temperature, exhaust flow, engine-out NO_x, urea dosing, exhaust pressure, and catalyst volume were used as model inputs. System-out NO_x, NH_3, HC, and CO concentrations were used as prediction targets.
Physics-informed feature engineering	Additional SCR-related features were generated before model training, including ANR, GHSV, NO_2 fraction, kinetic proxies, catalyst thermal state estimation, NH_3 storage proxy, and residence-time-related features.
Training and validation split	Separate training and external validation datasets were created at file and subtest level to avoid data leakage caused by highly autocorrelated time-series samples.
Machine learning model development	XGBoost regression models were trained separately for NO_x, NH_3, HC, and CO emissions. Hyperparameter optimisation was performed using RandomizedSearchCV together with ShuffleSplit cross-validation.
Model evaluation	Model performance was evaluated using R^2, MAE, RMSE, and RPIQ together with residual analysis, time-series comparison, and feature importance analysis on unseen validation datasets.
SCR sensitivity study	The trained physics-informed models were used to evaluate the effect of varying SCR catalyst volume on predicted NO_x reduction and NH_3 slip by recalculating volume-dependent features such as GHSV and ANR.

3.3 Data Access and Initial Data Assessment

This thesis is based exclusively on existing engine and EATS measurement data generated within Volvo Penta. No new experimental measurements were performed as part of the project. Instead, the work focused on identifying, extracting, filtering, and structuring historical test data suitable for machine learning-based prediction of downstream emissions, with primary focus on SCR-related NO_x conversion and ammonia slip behaviour.

The measurement data were accessed through Volvo Penta's internal data platform CAPE, where engine and EATS test campaigns are stored together with Sakura XML measurement files. CAPE was used to identify relevant tests, locate measurement files, and select suitable subtests for further analysis. The Sakura environment was also used to visualise and plot measurement signals, which made it possible to evaluate the behaviour of different tests before selecting them for model development. This early screening was carried out in close collaboration with the EATS group at Volvo Penta, whose experience with previous test campaigns helped identify the most relevant and useful datasets for the study.

The selected files were then processed in Python using the internal PySakura library, which allowed direct access to the measurement data without manual export. For each selected measurement file, subtest, and operating mode, the relevant signal channels were extracted and converted into pandas DataFrames for further preprocessing. Since the same physical parameter could appear under different channel names depending on engine platform, ECU configuration, or test setup, an alias-based parameter mapping was used to standardise variable names and ensure consistency across all datasets. This made it possible to combine data from multiple engine families into one unified machine learning dataset. The complete signal mapping used in this work is presented in Appendix A.1.

During the early phase of the project, the modelling work focused mainly on transient M1 measurement data. These datasets often contained very large numbers of raw time-series samples, where a single subtest could include several hundred thousand logged datapoints. At first, this appeared to provide a large and suitable dataset for model development. However, further analysis showed that a high number of samples did not necessarily mean a large number of physically different operating points.

Many measurements were recorded under nearly identical load conditions, resulting in limited variation in important SCR-related variables such as exhaust flow, exhaust temperature, engine-out NO_x, and urea demand. In practice, a dataset with several hundred thousand samples could still represent only a small number of repeated operating regions with similar catalyst conditions.

For machine learning, operating point diversity is often more important than raw sample count. A model trained mainly on repeated measurements from a few oper-

ating regions may learn local repetition instead of the broader physical behaviour of the catalyst system. This can lead to good performance near dominant operating points but weak generalisation for other temperatures, flow conditions, dosing levels, or engine loads.

This initial assessment showed that transient M1 data alone was not sufficient to provide robust coverage of the full SCR operating window. Since M1 measurements contain continuous transient time-series data, additional quasi-stationary LS measurements were included to improve dataset diversity and better represent stable operating conditions.

3.4 Data Selection and Inclusion Criteria

The final dataset was constructed from selected Volvo Penta engine and EATS test campaigns using measurements from the D5, D8, D11, and D13 engine platforms. These engines represent different displacement classes, catalyst volumes, and operating ranges, providing variation across multiple engine and EATS configurations. This improved the robustness of the model and supported the later evaluation of how catalyst volume influences predicted downstream emissions.

The objective of the data selection was to create a dataset with enough variation in important operating conditions such as exhaust temperature, exhaust flow, engine-out emissions, urea dosing, catalyst size, and downstream emissions. Since SCR performance depends on the interaction between these variables, one single test type was not enough to represent the full catalyst behaviour. Different test campaigns were therefore included to improve operating point coverage and create a more representative dataset for machine learning model development.

The included test categories are presented below, where each test type contributed different information relevant to SCR modelling and helped improve operating point coverage across different catalyst conditions.

- **Urea sweep tests** were included to study the relationship between urea dosing, NO_x conversion, and NH_3 slip. These tests were useful for understanding dosing sensitivity and the balance between low NO_x emissions and ammonia slip.

- **Part Load Mapping (PLM)** tests provided stable operating points across a wide range of loads, temperatures, and exhaust flow conditions. These tests were important because they represent normal engine operation and are well suited for steady-state modelling.

- **Full-load curves** were included to represent high-temperature and high-flow operating conditions close to maximum catalyst loading. These points were important for evaluating model performance under demanding conditions and for future catalyst sizing studies.

- **Transient emission cycles** introduced time-varying operating conditions where load, temperature, emissions, and dosing changed continuously. Although the final modelling approach was formulated as a static regression problem, selected transient measurements were still included after filtering and clustering in order to preserve operating range diversity without explicitly modelling full system dynamics.

Test files were included only if all required input and output signals were available with sufficient signal quality and clear operating descriptions. Since the machine learning model depends on both engine-out and system-out measurements, priority was given to tests containing the selected input parameters presented in Section 3.5, together with reliable target emission measurements.

Tests with missing critical parameters, unclear operating conditions, unstable sensor behaviour, or unreliable measurements were excluded from the final dataset. This was necessary to ensure that the final dataset was suitable for robust machine learning model development and reliable prediction of downstream NO_x, NH_3, HC, and CO emissions across multiple engine platforms and EATS configurations.

3.5 Model Inputs and Outputs

The modelling approach was based on predicting system-out emissions from engine-out operating conditions. Engine-out variables describe the conditions entering the EATS and were used as the main model inputs, while downstream emissions after the EATS were used as model outputs.

This structure is important for future virtual testing, where engine-out conditions can be generated from engine simulation models and used to estimate EATS performance without physical engine testing. The model was therefore designed not only for prediction of measured test data, but also as a possible support tool for future catalyst development and SCR-related studies.

The selected input variables are presented below. These variables were chosen because they represent the main factors affecting SCR behaviour and have a strong influence on NO_x conversion, ammonia storage, and the risk of NH_3 slip.

- **Exhaust gas temperature** (`t_exhaust`) was included because temperature strongly affects catalyst activity, urea decomposition, and SCR reaction efficiency.

- **Exhaust mass flow** (`flow_exhaust`) was included because it affects gas residence time inside the catalyst and therefore influences conversion efficiency.

- **Engine-out NO_x concentration** (`conc_nox_eo`), together with NO and NO_2 concentrations (`conc_no_eo`, `conc_no2_eo`), was included to represent

the incoming NO_x load and NO_2 fraction entering the SCR system.

- **Urea dosing** (`urea_demand`) was included because it controls ammonia availability and strongly affects both NO_x reduction and NH_3 slip.
- **Exhaust pressure** (`p_exhaust`) was used mainly for density estimation and flow-related calculations.
- **Catalyst volume** (`volym_kat`) was included to represent differences in SCR size between engine platforms and to support future catalyst sizing studies.
- **Traceability metadata**, such as test number, subtest index, engine family, and measurement mode, were stored for validation and result analysis, but were not used directly for model training.

The selected system-out emission variables consisted of downstream emissions measured after the EATS. These variables represent the final emissions after catalyst conversion and were used as the main prediction targets of the model. The selected target variables were:

- **System-out NO_x concentration** (`conc2_nox_so`), used to evaluate overall SCR conversion performance and the main target variable of this thesis.
- **System-out NH_3 concentration** (`conc_nh3_so`), used to represent ammonia slip and the balance between NO_x reduction and overdosing risk.
- **System-out HC concentration** (`conc2_hc_so`) and **system-out CO concentration** (`conc2_co_so`), included to support broader EATS emission modelling and overall model evaluation.

Additional physics-informed variables were created during the feature engineering step and added before model training. These derived features are described in Section 3.7.

The final selection of input and output variables was chosen to balance physical relevance, model robustness, and industrial applicability, creating a suitable foundation for both emission prediction and future virtual EATS development.

3.6 Preprocessing Pipeline

Before model training, the extracted measurement data required preprocessing to reduce measurement noise and remove redundant operating points from large transient time-series datasets. The dataset included both transient M1 measurements and quasi-stationary LS measurements. Since the raw Sakura measurements were stored as continuous time-series, a single transient subtest could contain several hundred thousand samples. Using all raw samples directly would introduce strong

autocorrelation and overrepresentation of repeated operating conditions.

The preprocessing pipeline therefore focused on signal smoothing using exponentially weighted moving average (EWMA) filtering together with operating-point reduction using k-means clustering for transient M1 measurements. LS measurements already represent stable operating points and were therefore retained at their original resolution after filtering without clustering.

This preprocessing strategy made it possible to combine transient and stationary datasets into one unified machine learning dataset while preserving the main operating range and reducing unnecessary redundancy.

3.6.1 EWMA Filtering

Before further preprocessing, all non-metadata numeric signals were smoothed using an EMWA filter. The purpose of this step was to reduce short-term sensor noise and signal fluctuations while preserving the overall physical behaviour of the system. This was especially important for exhaust temperature, emission concentrations, and exhaust flow signals, where short spikes or unstable measurements could otherwise affect model training.

EWMA was chosen because it gives more weight to recent observations and gradually less weight to older samples. Compared to a simple moving average, this makes the filter more responsive to real operating changes while still reducing high-frequency measurement noise. This is suitable for EATS data, where real changes in catalyst behaviour should be preserved while short non-physical fluctuations should be removed.

Different span values were used depending on the signal type. A default span was applied to most variables, while selected high-priority signals such as SCR inlet temperature (`t_b_scr`), exhaust flow (`flow_exhaust`), and engine-out NO_x concentration (`conc_nox_eo`) used manually adjusted spans based on signal behaviour and engineering judgement. This allowed stronger filtering for noisy signals while keeping sufficient sensitivity for important transient changes.

3.6.2 K-means Clustering for M1 Data

After filtering, transient M1 measurements were reduced using k-means clustering in operating space. Since many samples represented nearly identical operating conditions, clustering was used to reduce redundancy and create a more balanced dataset for machine learning. This also reduced computational cost and prevented repeated operating regions from dominating model training.

K-means groups samples with similar operating conditions and replaces them with representative mean operating points. In this way, the main physical operating range could be preserved while unnecessary repeated time-series samples were removed.

Clustering was performed using three key variables: exhaust mass flow (`flow_exhaust`), SCR inlet temperature (`t_b_scr`), and engine-out NO_x concentration (`conc_nox_eo`). These were selected because they strongly influence SCR conversion behaviour and together describe the main catalyst operating conditions.

Before clustering, the selected variables were standardised using z-score normalisation so that all variables had equal influence. K-means was then applied with approximately 300 clusters per subtest when the number of samples exceeded this limit. Each cluster was represented by the mean value of its numeric signals, while metadata such as test number and subtest index were kept for traceability.

This reduced each transient subtest to a smaller set of representative operating points, making model training more efficient while preserving the main operating range. It also improved the balance between transient and steady-state data and reduced the risk of overfitting repeated operating regions.

To verify that important operating regions were not removed, the data distribution was compared before and after clustering using histogram analysis and scatter plot inspection. The results of this comparison are presented in Section 4.1.

3.7 Physics-Informed Feature Engineering

In addition to the directly measured input variables, several physics-informed features were created before model training to better represent SCR chemistry and catalyst behaviour. The purpose was to improve prediction robustness, increase physical interpretability, and reduce the risk of the model learning only statistical correlations instead of the main physical mechanisms behind NO_x conversion and NH_3 slip.

Since SCR performance depends not only on individual sensor values but also on derived relationships between exhaust flow, temperature, catalyst size, and urea dosing conditions, purely data-driven inputs were considered insufficient for robust modelling across different engine platforms and operating conditions. The feature engineering strategy therefore aimed to combine machine learning with simplified physical understanding of catalyst behaviour while maintaining computational efficiency and practical industrial applicability.

The derived features were based on common engineering relationships such as catalyst residence time, ammonia-to-NO_x balance, thermal behaviour, and simplified reaction kinetics. These features were calculated from measured signals and added before model training.

The following physics-informed features were included in the final model:

- **NO_x mass flow proxy** (`flw_nox_gs`), calculated from exhaust flow and engine-out NO_x concentration to better represent the incoming NO_x load.

- **Gas Hourly Space Velocity (GHSV)** (`GHSV`), calculated from exhaust flow and catalyst volume. This represents catalyst loading and residence time.

- **Ammonia-to-NO_x Ratio (ANR)** (`ANR`), calculated from urea demand and estimated NO_x mass flow. This describes the balance between ammonia supply and NO_x reduction demand.

- **NO_2 fraction** (`no2_fraction`), calculated from the ratio between NO_2 and total NO_x. This is important for fast SCR reactions.

- **Catalyst thermal state proxy** (`t_stone_est`), estimated using an EWMA of SCR inlet temperature to represent catalyst thermal history.

- **NH_3 storage proxy** (`nh3_storage_proxy`), calculated from rolling urea demand to represent ammonia storage and delayed release.

- **Temperature-dependent kinetic term** (`kinetics_term`), calculated using an Arrhenius-inspired function of SCR inlet temperature to represent catalyst activity.

- **Damköhler number proxy** (`Da_number`), calculated from the ratio between the kinetic term and GHSV to describe the balance between reaction rate and residence time.

- **Inverse exhaust flow** (`inv_flow`), used as an additional residence-time-related feature, especially at low flow conditions.

- **Crystallisation risk indicator** (`crys_risk`), a binary feature representing low-temperature conditions with active urea dosing where deposit formation may occur.

To evaluate the effect of physics-informed feature engineering, two model configurations were developed. The baseline model used only directly measured engine and EATS variables, while the physics-informed model used the same XGBoost framework but included the additional derived features described above.

This made it possible to evaluate whether physics-informed features improved prediction accuracy, robustness, and model interpretability compared to a purely data-driven approach. The implementation was performed in Python before model training and the full feature definitions are presented in Appendix A.2.

3.7.1 Dataset Composition and Validation Strategy

The final dataset was divided into separate training and validation sets for each engine family. Individual CSV files were created for D5, D8, D11, and D13, with one training dataset and one validation dataset for each engine. This made it possible to train one combined model using data from multiple engine platforms while evaluating the results separately for each engine family. The dataset generation, preprocessing, training-validation split, and model training workflow were implemented in Python. A simplified overview of the implementation is provided in Appendix A.3 and Appendix A.4.

The split between training and validation data was done manually at test file and subtest level. Complete files or complete subtests were assigned either to training or validation, never to both. The validation data therefore consisted of separate measurement files that were not used during training. This was important to avoid data leakage, since neighbouring time-series samples are highly autocorrelated and should not be treated as independent observations. A random row-wise split would therefore give unrealistically optimistic validation results.

All engine-specific training files were then loaded and combined into one shared multi-engine training dataset. This gave the model access to a wider operating range and improved the possibility of generalisation across different engines and catalyst sizes.

Model tuning was performed using ShuffleSplit cross-validation together with RandomizedSearchCV during XGBoost hyperparameter optimisation. This internal cross-validation was only applied within the combined training dataset. The external validation datasets remained completely separate and were only used for final model evaluation.

This two-level validation strategy reduced the risk of overfitting and provided a more realistic evaluation of model performance on unseen data. The goal was not only high training accuracy, but also a robust model with good performance across different operating conditions and validation tests.

3.8 Machine Learning Model Development

After data extraction, preprocessing, and physics-informed feature engineering, the final machine learning workflow was developed to predict system-out emissions from engine-out operating conditions. The goal was to evaluate whether machine learning could accurately describe the relationship between exhaust flow, temperature, NO_x load, urea dosing, catalyst volume, and downstream emissions, with enough robustness for future use in SCR sizing, catalyst scaling, and virtual testing.

Since the target variables were continuous emission values, such as system-out NO_x, NH_3, HC, and CO concentrations, the problem was formulated as a supervised re-

gression task. Separate regression models were trained for each target variable, which made it possible to apply target-specific transformations, optimisation, and postprocessing.

To evaluate the effect of physics-informed feature engineering, two model configurations were compared. The first model used only directly measured variables, while the second model included additional physics-informed features derived from SCR-related relationships. The full feature definitions are presented in Section 3.7. This comparison was used to evaluate whether the added physics-informed features improved prediction accuracy, robustness, and model interpretability compared to using only measured signals.

3.8.1 Model Selection

Several regression methods were considered during the early model development. Since the dataset contains strong nonlinear relationships between exhaust temperature, exhaust flow, engine-out NO_x, catalyst volume, and urea dosing, simple linear regression models were not considered sufficient for describing SCR behaviour. Small changes in temperature or dosing can lead to large changes in NO_x conversion and NH_3 slip, which makes nonlinear models more suitable.

Initial model testing was also performed in MATLAB using its built-in regression and machine learning tools. MATLAB was useful for early analysis and understanding of the available datasets. However, the final workflow required more advanced preprocessing, including alias mapping, EWMA filtering, clustering of transient M1 data, physics-informed feature generation, and merging of data from multiple engine families. In addition, the validation strategy required flexible train-validation separation together with repeated cross-validation and hyperparameter optimisation. This became easier to manage in Python.

For this reason, the final model development was performed in Python using Scikit-learn for validation workflows and XGBoost for regression modelling. XGBoost was selected because it performed well on structured engineering datasets and could capture complex nonlinear relationships without requiring a full physical catalyst model.

Compared to neural networks, XGBoost required less preprocessing, handled tabular data well, and was less sensitive to moderate dataset size limitations. It also includes regularisation methods that reduce overfitting and improve robustness on unseen operating conditions.

Another advantage was the possibility to analyse feature importance after training. This improved model interpretability and made it possible to compare the learned behaviour with known SCR physics, which was important when evaluating the effect of physics-informed feature engineering.

The final model should therefore be seen as a supervised regression model for pre-

dicting steady-state operating points, not as a full dynamic catalyst model. The purpose was not to describe the full catalyst behaviour over time, but to provide reliable prediction of system-out emissions based on engine-out conditions and physically meaningful catalyst features.

3.8.2 Training Strategy

Separate XGBoost regression models were trained for the main target variables: system-out NO_x (`conc2_nox_so`), NH_3 (`conc_nh3_so`), HC (`conc2_hc_so`), and CO (`conc2_co_so`). NO_x and NH_3 were the main focus of the thesis, since they are directly linked to SCR conversion efficiency, ammonia dosing, and catalyst sizing. HC and CO were included mainly to support broader EATS evaluation and to verify that the modelling approach remained physically reasonable across multiple emissions.

The same XGBoost framework, preprocessing pipeline, and validation strategy were used for all target variables and for both model configurations. Only the input feature set was changed between the baseline and physics-informed models, which made it possible to isolate the effect of the added physics-informed features.

Before training, all target values were clipped to non-negative values since negative emissions caused by sensor noise are physically unrealistic. For NH_3, HC, and CO, a logarithmic transformation was applied using,

$$y_{\text{train}} = \log(1 + y)$$

to reduce the effect of large outliers and improve stability for skewed target distributions. NO_x was trained without transformation since its values were already suitable for regression, and keeping the original scale made the results easier to interpret for engineering evaluation.

Sample weighting was also applied during training to increase the influence of high engine-out NO_x operating points, since these are often most important for catalyst sizing and emission compliance. The weighting factor was defined as,

$$w_i = \frac{\text{conc\_nox\_eo} + 1}{\overline{\text{conc\_nox\_eo}}}$$

which gave higher importance to high-load operating conditions during model optimisation.

3.8.3 Hyperparameter Optimisation

Hyperparameter tuning was performed to find stable XGBoost settings for each target variable and to reduce the risk of both underfitting and overfitting. Since model performance depends strongly on the balance between model complexity and generalisation, this step was important for achieving robust predictions across different

engine families and operating conditions.

The optimisation was performed using `RandomizedSearchCV` together with `ShuffleSplit` cross-validation in Scikit-learn. This made it possible to test several parameter combinations efficiently while keeping the computational cost at a reasonable level.

The tuning was applied only within the combined training dataset. `ShuffleSplit` created three repeated train-validation splits, where 30% of the training data was used for internal validation in each split. The external validation datasets for D5, D8, D11, and D13 were kept completely separate and were only used for final blind evaluation.

The main tuned XGBoost parameters were the number of trees (`n_estimators`), maximum tree depth (`max_depth`), and learning rate (`learning_rate`), since these strongly affect model complexity and training stability. The final search space was:

- `n_estimators`: 500, 800
- `max_depth`: 6, 8
- `learning_rate`: 0.03, 0.05

These ranges were selected based on preliminary testing during model development to balance model flexibility and stable generalisation. Larger models increased the risk of overfitting, while smaller models reduced the ability to capture the nonlinear SCR behaviour. Final model selection was based on internal cross-validation during tuning together with blind validation on unseen engine datasets.

3.8.4 Prediction Constraints and Postprocessing

After model training, the raw predictions from the XGBoost models were postprocessed to ensure physically reasonable results and to reverse the target transformations used during training. Since the model works as a regression function without built-in physical limits, additional postprocessing was applied to avoid unrealistic predictions such as negative emissions or system-out values higher than engine-out values.

For NH_3, HC, and CO, the logarithmic transformation used during training was reversed using

$$y = \exp(\hat{y}) - 1$$

to return the predictions to the original concentration scale in ppm.

All predicted concentrations were clipped to non-negative values since negative emissions are physically impossible and usually come from regression noise or sensor uncertainty.

For NH_3, an additional upper limit of 20 ppm was applied after inverse transformation. Since ammonia slip values are normally low, this reduced occasional unrealistic overprediction and improved stability during validation.

For NO_x, the predicted system-out concentration was constrained to remain lower than the corresponding engine-out NO_x concentration, since the SCR system is designed to reduce NO_x:

$$NO_x^{out} \leq NO_x^{in}$$

The same rule was also applied to HC and CO, so that predicted system-out values could not exceed the corresponding engine-out values.

These postprocessing steps improved model robustness and ensured that the final predictions remained physically meaningful for engineering evaluation and future use in SCR sizing and virtual testing.

3.8.5 Evaluation Metrics

Model performance was evaluated using both internal cross-validation during training and external blind validation on unseen measurement files. The goal was not only to achieve good statistical accuracy, but also to evaluate whether the model could provide reliable prediction of EATS emissions across different engine families and operating conditions.

The main evaluation metrics were R^2, MAE, RMSE, and RPIQ, as described in Section 2.7. These metrics were chosen to give different views of model quality and robustness.

R^2 was used as the main metric during hyperparameter tuning and shows how well the model explains the variation in the measured emission data. Values closer to 1 indicate better predictive performance.

MAE was used to measure the average prediction error in physical units (ppm), which makes the results easier to interpret from an engineering perspective. RMSE was also included since it gives more weight to larger prediction errors and helps identify cases where the model produces occasional large deviations.

RPIQ was used as an additional robustness metric and was calculated as,

$$\text{RPIQ} = \frac{\text{IQR}}{\text{RMSE}}$$

where IQR is the interquartile range of the measured target values. Higher RPIQ values indicate stronger prediction quality relative to the spread of the dataset.

Validation was performed separately for each engine family using dedicated external validation datasets for D5, D8, D11, and D13 that were not included during train-

ing or hyperparameter tuning. This blind validation was important for evaluating model performance on unseen operating conditions rather than only repeated training patterns.

In addition to numerical metrics, time-series comparisons between measured and predicted emissions were used to visually evaluate model behaviour during transient and quasi-stationary operating regions. This helped verify that the model captured physically meaningful trends and not only acceptable average error values.

Feature importance analysis from the trained XGBoost models was also included to evaluate which variables contributed most strongly to each target prediction. This helped assess whether the model learned physically meaningful SCR relationships rather than only statistical correlations.

The complete machine learning workflow, including model training, hyperparameter optimisation, postprocessing, and feature importance extraction, is summarised in Appendix A.4.

3.8.6 SCR Scaling Feasibility Evaluation

In addition to emission prediction, a model-based feasibility study was performed to evaluate how changes in SCR catalyst volume influenced predicted emissions. Since catalyst size is strongly related to NO_x conversion efficiency, residence time, and NH_3 slip, this was relevant for investigating whether the trained physics-informed XGBoost model could support early-stage SCR sizing analysis.

The study was performed using the external validation datasets for the D5, D8, D11, and D13 engine platforms. For each dataset, the original operating conditions were kept unchanged while the catalyst volume parameter was scaled using relative volume factors between 70% and 130% of the original catalyst volume:

$$V_{\text{cat,scaled}} = f_V \cdot V_{\text{cat,original}}$$

where f_V is the catalyst volume scaling factor. The tested values were 0.70, 0.80, 0.90, 1.00, 1.10, 1.20, and 1.30.

For each scaled catalyst volume, the physics-informed features were recalculated using the same feature engineering pipeline as during model training. This was important because several derived variables depend directly on catalyst volume, especially GHSV, which describes the relationship between exhaust flow and catalyst size. A larger catalyst volume gives lower GHSV, which means longer residence time and potentially improved NO_x conversion.

The trained models were then used to predict system-out NO_x and NH_3 for each scaling factor, and the results were used to evaluate how catalyst volume influenced SCR performance.

For each engine platform and scaling factor, statistical summaries were calculated, including mean, median, 90th percentile, 95th percentile, maximum value, and number of valid prediction points. The mean values were used to evaluate how predicted NO_x and NH_3 changed with catalyst volume. Relative changes compared to the original catalyst volume were also calculated as

$$\Delta y_{\text{rel}} = \frac{\bar{y}_{f_V} - \bar{y}_{1.00}}{\bar{y}_{1.00}} \cdot 100$$

where $\bar{y}_{f_V}$ is the mean predicted emission for a given scaling factor and $\bar{y}_{1.00}$ is the mean prediction at the original catalyst volume.

This evaluation should be interpreted as a sensitivity study rather than a final catalyst sizing method, since the approach does not include all physical effects that would occur if the hardware were actually changed, such as thermal behaviour, flow distribution, catalyst aging, or recalibrated urea dosing. The predictions are also only reliable within the operating range represented by the training data.

Even so, the study provides an initial indication of how a physics-informed machine learning model could support early SCR sizing before physical prototype testing.

4
Results and Discussion

This chapter presents and discusses the results from the developed machine learning models for system-out emission prediction in EATS.

First, the preprocessing strategy and dataset characteristics are evaluated. The prediction performance for NO_x, NH_3, HC, and CO is then analysed using validation metrics and graphical evaluation. A comparison between the baseline and physics-informed models is also presented, followed by an assessment of the model's potential for virtual testing and SCR scaling applications.

4.1 Dataset Characteristics and Preprocessing Validation

To evaluate the preprocessing strategy, the raw selected measurements were compared with the final processed dataset after filtering and clustering..

For D5, the selected raw dataset initially contained 126,268 samples, while the final processed dataset used for model training contained 3,494 samples, corresponding to a reduction of 97.2%.

Most removed samples originated from highly repetitive transient regions where many consecutive points represented nearly identical operating conditions. The preprocessing therefore mainly reduced redundancy while preserving the physically relevant operating space required for model training.

Figure 4.1 compares the distributions before and after preprocessing for SCR inlet temperature (`t_b_scr`), exhaust flow (`flow_exhaust`), engine-out NO_x concentration (`conc_nox_eo`), and urea demand (`urea_demand`). The main operating ranges and extreme regions are preserved, while strongly overrepresented areas become more balanced. This indicates that preprocessing mainly removed redundant observations rather than physically important operating conditions.

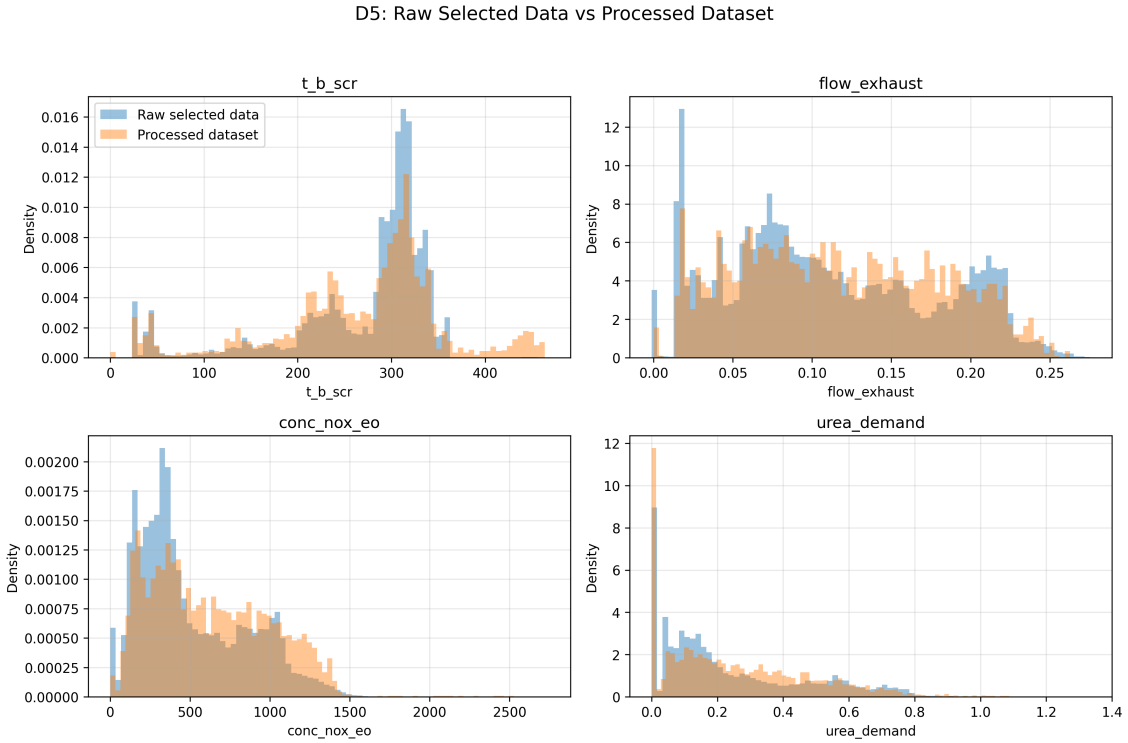


Figure 4.1: Distribution comparison between raw selected measurements and the final processed dataset for D5 after EWMA filtering and clustering reduction.

Figure 4.2 shows the operating space defined by exhaust flow and engine-out NO_x concentration before and after preprocessing. The main relationship between the variables remains unchanged, while the number of repeated points is greatly reduced. The processed dataset therefore provides a more balanced representation of the same operating space while reducing overrepresentation of dominant operating regions during training.

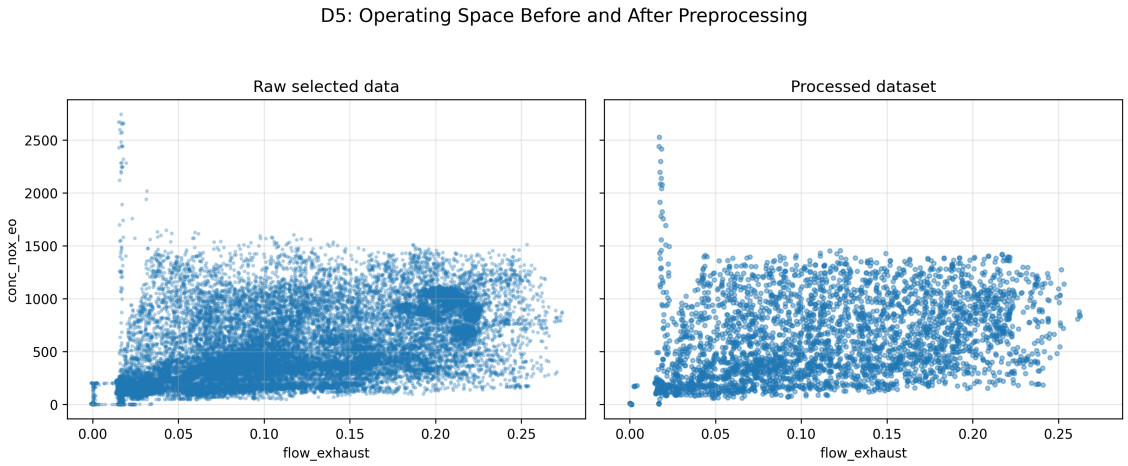


Figure 4.2: Operating space comparison before and after preprocessing for D5 using exhaust flow and engine-out NO_x concentration.

Similar results were observed for D8, D11, and D13, indicating that the preprocessing strategy consistently reduced redundancy while preserving the relevant operating space across all engine platforms.

4.2 Comparison Between Data-Driven and Physics-Informed Approaches

The baseline and physics-informed models were compared to evaluate whether the additional physics-informed features improved prediction accuracy and robustness. The largest improvements were expected for NO_x and NH_3, since these emissions are directly linked to SCR behaviour such as ammonia dosing, catalyst residence time, temperature-dependent conversion, and ammonia storage effects. Smaller improvements were expected for HC and CO, which are more strongly influenced by combustion behaviour and DOC oxidation.

Table 4.1: Cross-validation comparison between the baseline and physics-informed models across all emission targets.

Target	**Baseline Model** R^2	**Physics-Informed Model** R^2	**Difference** ΔR^2
NO_x	0.8738	0.8988	+0.0250
HC	0.9027	0.8961	-0.0066
CO	0.8379	0.8253	-0.0126
NH_3	0.7390	0.7995	+0.0605

As shown in Table 4.1, the largest improvements were observed for NO_x and NH_3, which are most directly linked to SCR behaviour. The strongest improvement was obtained for NH_3, where the R^2 value increased by more than 0.06. In contrast, HC and CO showed slightly lower cross-validation performance with the physics-informed model. This suggests that the added SCR-focused features provided limited additional information for DOC-dominated emissions and may in some cases have introduced unnecessary model complexity for these targets. The results therefore indicate that physics-informed feature engineering is most effective when the added features represent the dominant physical mechanisms of the target variable.

Table 4.2 shows the validation performance for the baseline and physics-informed models using unseen datasets from the D5, D8, D11, and D13 engine platforms.

Table 4.2: Validation comparison between the baseline and physics-informed models for NO_x and NH_3.

Engine	Target	Baseline	Physics-Informed	ΔR^2
D5	NO_x	0.7518	0.8016	+0.0498
D8	NO_x	0.7855	0.8896	+0.1041
D11	NO_x	0.9584	0.9688	+0.0104
D13	NO_x	0.9206	0.9379	+0.0173
D5	NH_3	0.8311	0.8571	+0.0260
D8	NH_3	0.0660	0.6626	+0.5966
D11	NH_3	0.7812	0.7240	-0.0572
D13	NH_3	-1.0491	0.5783	+1.6274

The physics-informed model improved NO_x prediction for all engine platforms. The largest improvement was observed for D8, where the validation R^2 increased by more than 0.10. This suggests that the physics-informed features helped the model better represent SCR conversion across different catalyst sizes and operating conditions.

The improvements were even larger for NH_3, particularly for D8 and D13. The baseline model showed poor generalisation for these engines, while the physics-informed model produced substantially more stable and physically reasonable predictions. The strongest improvement was observed for D13, where the validation performance improved from a negative to a clearly positive R^2 value. This suggests that the added features improved robustness for operating conditions strongly affected by ammonia storage and delayed release behaviour.

This behaviour is physically reasonable since NH_3 slip depends strongly on catalyst storage effects, delayed ammonia release, and storage saturation, which are difficult to observe directly from standard measurement signals. The added physics-informed features therefore helped the model better represent these hidden catalyst dynamics.

The weaker NH_3 result for D11 indicates that some operating regions remain difficult to model even with physics-informed features. Possible reasons include sensor uncertainty, limited representation of catalyst storage dynamics, or ASC-related behaviour that is not fully captured by the available signals. NH_3 prediction is therefore naturally more difficult than NO_x prediction.

4.3 Emission Prediction Performance

Based on the improved performance observed for the physics-informed model, the remaining prediction analysis focuses on the physics-informed XGBoost model.

Model performance was evaluated using measured-versus-predicted scatter plots, residual analysis, time-series comparisons, and numerical validation metrics. Together, these evaluations were used to assess prediction accuracy, robustness, and the model's ability to represent the main trends in the validation data. NO_x and NH_3

are presented separately, followed by a shorter evaluation of HC and CO prediction performance.

4.3.1 NO_x Prediction Results

Figure 4.3 shows measured and predicted system-out NO_x concentrations for the four engine platforms using the physics-informed XGBoost model. The results show strong overall agreement across all validation datasets, indicating that the model captures the dominant NO_x conversion behaviour of the EATS.

The strongest agreement was obtained for D11, followed by D13, with validation R^2 values of 0.969 and 0.938, respectively. These results indicate that the model captured the dominant NO_x conversion behaviour consistently across the evaluated operating conditions. D8 also showed strong performance with an R^2 value of 0.890. D5 showed lower but still acceptable performance, with an R^2 value of 0.802 and a mean absolute error of 10.5 ppm. The lower performance for D5 likely reflects stronger transient variation, larger temperature fluctuations, and greater sensitivity to urea dosing compared to the larger engine platforms.

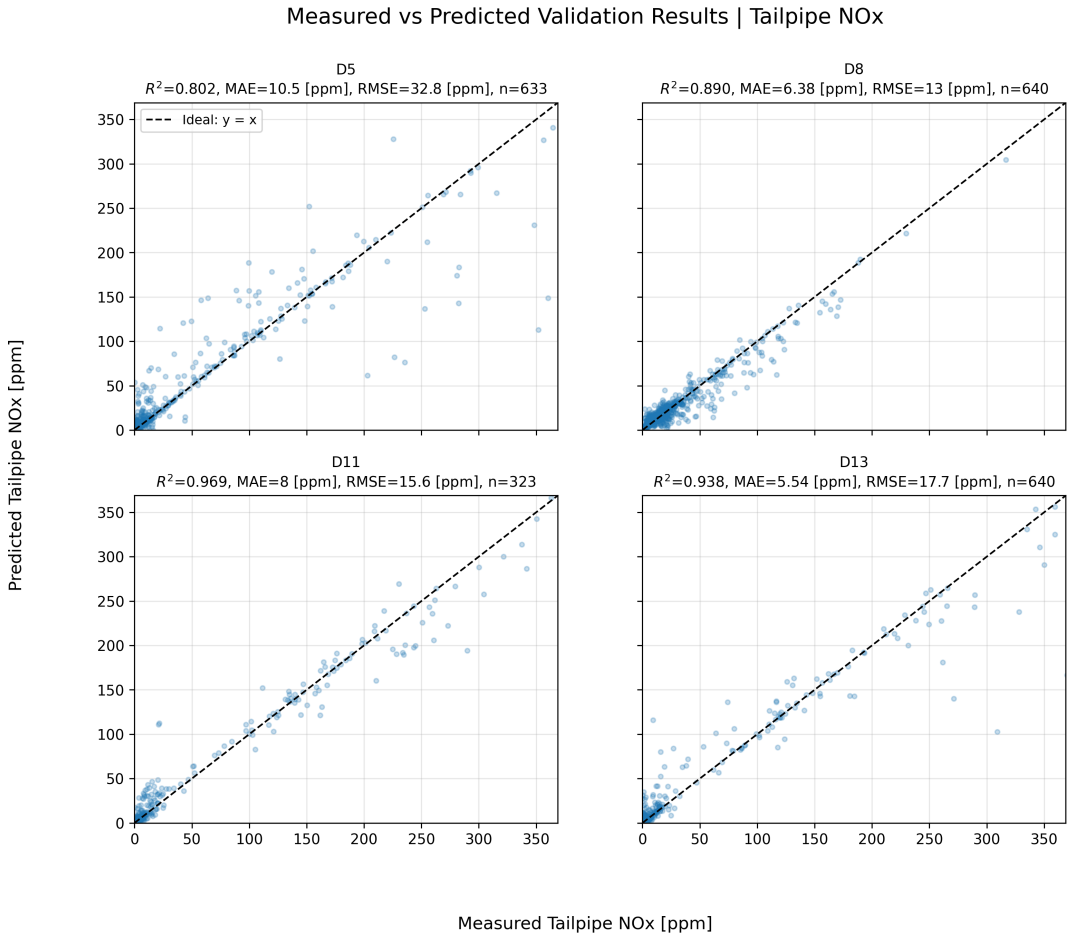


Figure 4.3: Measured versus predicted system-out NO_x concentrations for the physics-informed model across the four engine platforms.

The scatter plots show that most operating points are distributed close to the ideal prediction line, particularly for D11 and D13, where the spread remains small across both low and high concentration regions. D8 also shows strong agreement, while D5 has a larger spread, especially at higher NO_x concentrations. This indicates that rapid transient events and high-emission peaks, which are associated with fast-changing catalyst conditions, remain more difficult for the static regression model to reproduce accurately.

The residuals for NO_x prediction are shown in Figure 4.4. For most operating points, the residuals are concentrated close to zero, indicating good overall prediction accuracy without major systematic errors in the low and medium emission regions. However, for higher measured NO_x values, particularly for D8, D11, and D13, the residuals become increasingly negative, indicating that the model tends to underestimate the largest NO_x peaks.

This behaviour is physically reasonable since high-emission events are often associated with transient catalyst dynamics that are difficult to fully represent using a static operating-point model. Another contributing factor may be the imbalance toward low-emission operating points in the dataset, causing the model to become more strongly optimized for the dominant low-NO_x regions than for rare high-emission peaks. D5 also shows a larger residual spread across the full range, reflecting the greater difficulty of predicting smaller engines with stronger transient variation.

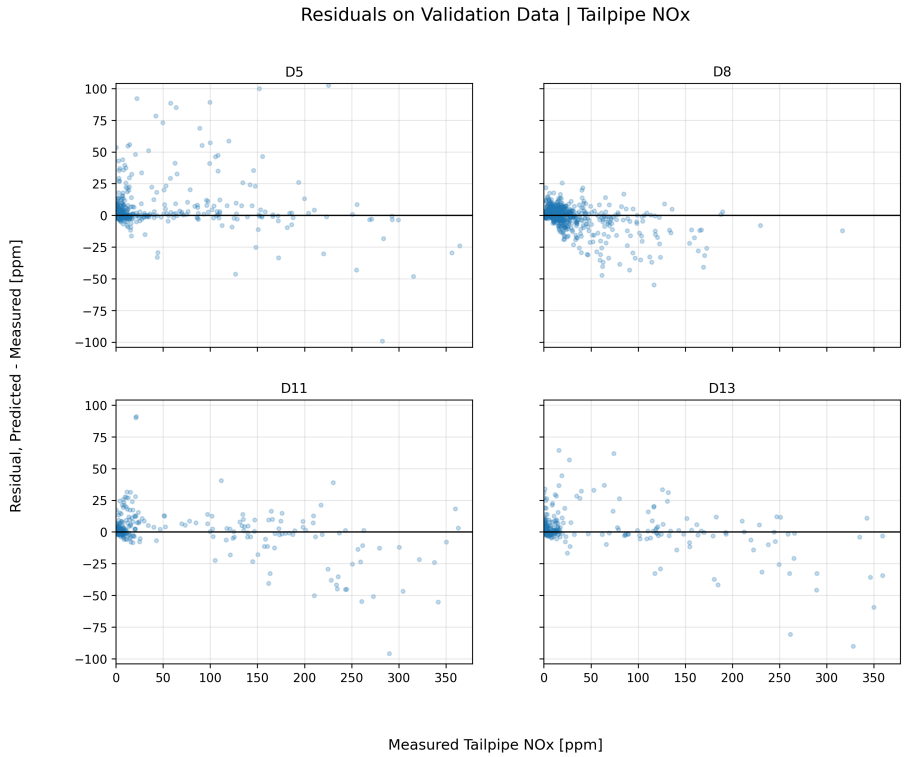


Figure 4.4: Residuals for system-out NO_x prediction using the physics-informed model. Residuals are defined as predicted minus measured concentration.

Figure 4.5 presents a time-series comparison between measured and predicted NO_x concentrations. The model follows the overall transient behaviour well and reproduces most of the major emission events across the validation cycles.

The general trend is captured well, particularly for D11 and D13, although the largest NO_x peaks are often underestimated, especially for D5 and D8. This is consistent with the residual patterns in Figure 4.4 and may indicate that rapid transient events remain the most challenging operating conditions for the model.

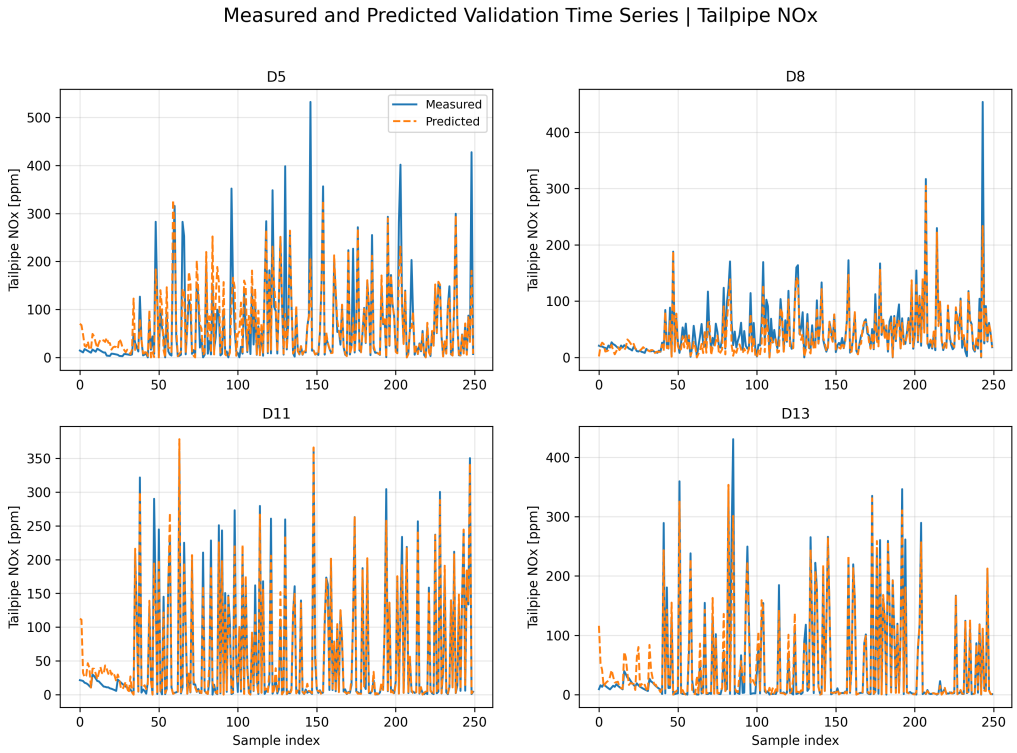


Figure 4.5: Time-series comparison between measured and predicted system-out NO_x concentrations for the physics-informed model.

The stronger prediction performance for D11 and D13 compared to D5 may indicate that larger engine platforms operate under more thermally stable conditions where SCR conversion becomes easier to model. Smaller engines such as D5 often experience faster load transitions and larger relative temperature fluctuations, which increase prediction difficulty and make the model more sensitive to local catalyst state variations.

Overall, the NO_x results show that the physics-informed model provides strong and physically consistent prediction performance across all evaluated engine platforms. The main remaining limitation is the underestimation of the largest transient NO_x peaks, indicating that future improvements should focus on dynamic catalyst state representation.

4.3.2 NH_3 Prediction Results

Figure 4.6 shows the measured and predicted NH_3 slip concentrations for the four engine platforms using the physics-informed XGBoost model. Compared to NO_x prediction, the NH_3 results show larger variation between engine platforms and operating conditions, reflecting the increased difficulty of modelling NH_3 slip behaviour using static operating-point prediction.

The best prediction performance was obtained for D5, where the model reached a validation R^2 of 0.857 and a mean absolute error of 0.10 ppm. D8 also showed good performance with an R^2 of 0.663. Lower accuracy was observed for D11 and D13, where larger deviations appeared during high-slip operating conditions. Even so, the physics-informed model maintained stable and physically reasonable predictions for D13 despite the increased difficulty of predicting transient NH_3 slip behaviour.

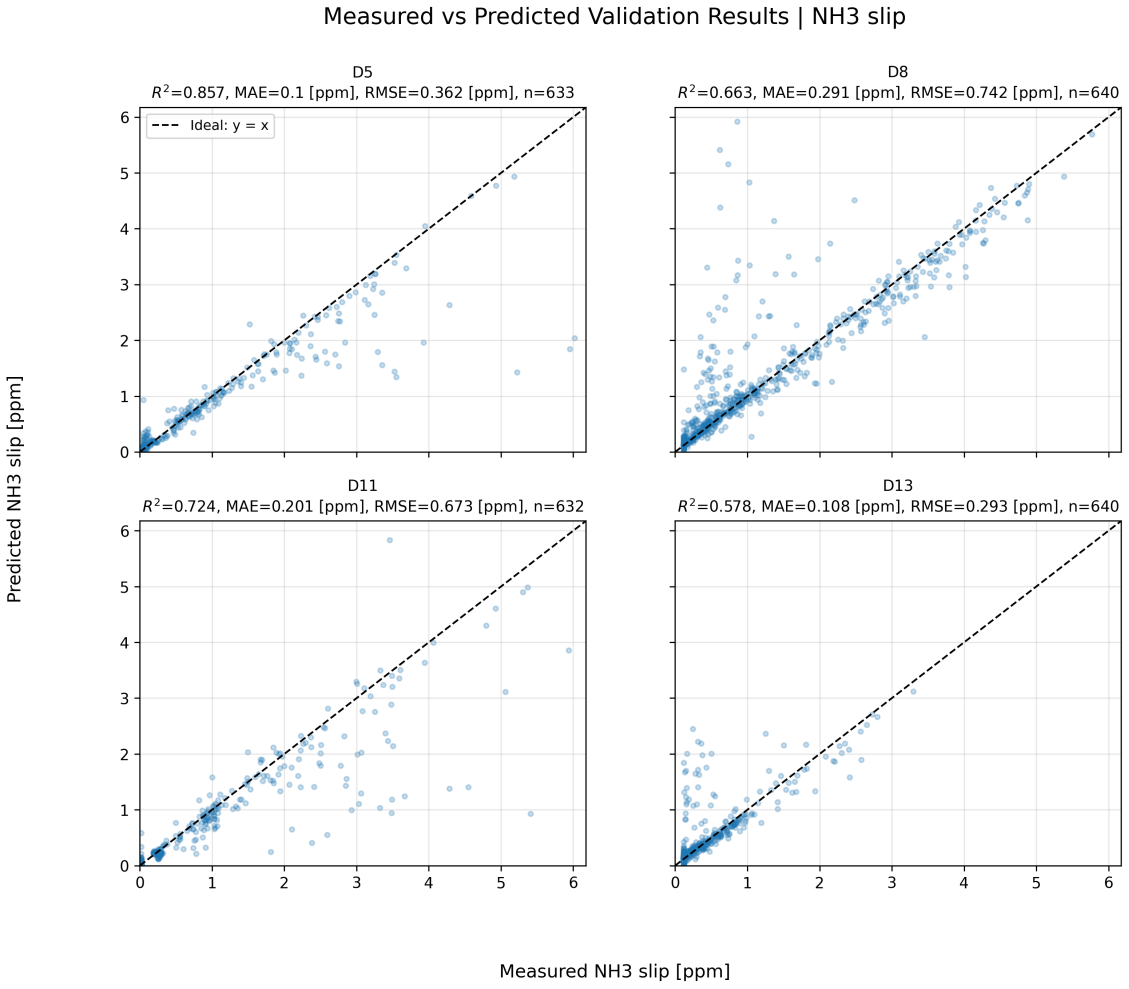


Figure 4.6: Measured versus predicted NH_3 slip concentrations for the physics-informed model across the four engine platforms.

The scatter plots show that most low-slip operating points are distributed close to the ideal prediction line, particularly for D5 and partly D8, indicating stable prediction performance during normal low-slip operating conditions. Larger deviations appear during high-slip events, especially for D8, D11, and D13, where both underprediction and overprediction can be observed. This indicates that transient NH_3 storage and release behaviour remains difficult to reproduce accurately using a static regression model.

Figure 4.7 shows the residuals for NH_3 prediction. Most residuals remain close to zero at low NH_3 levels, indicating stable prediction performance during normal operation. As measured NH_3 increases, the spread becomes larger and the model often underpredicts the highest slip values, especially for D11 and D13 where several large negative residuals appear. At the same time, some positive residuals can be observed at low measured values, particularly for D8 and D13, indicating occasional overprediction of slip events. This suggests that the model captures the overall NH_3 slip behaviour but remains sensitive to transient catalyst storage and release dynamics.

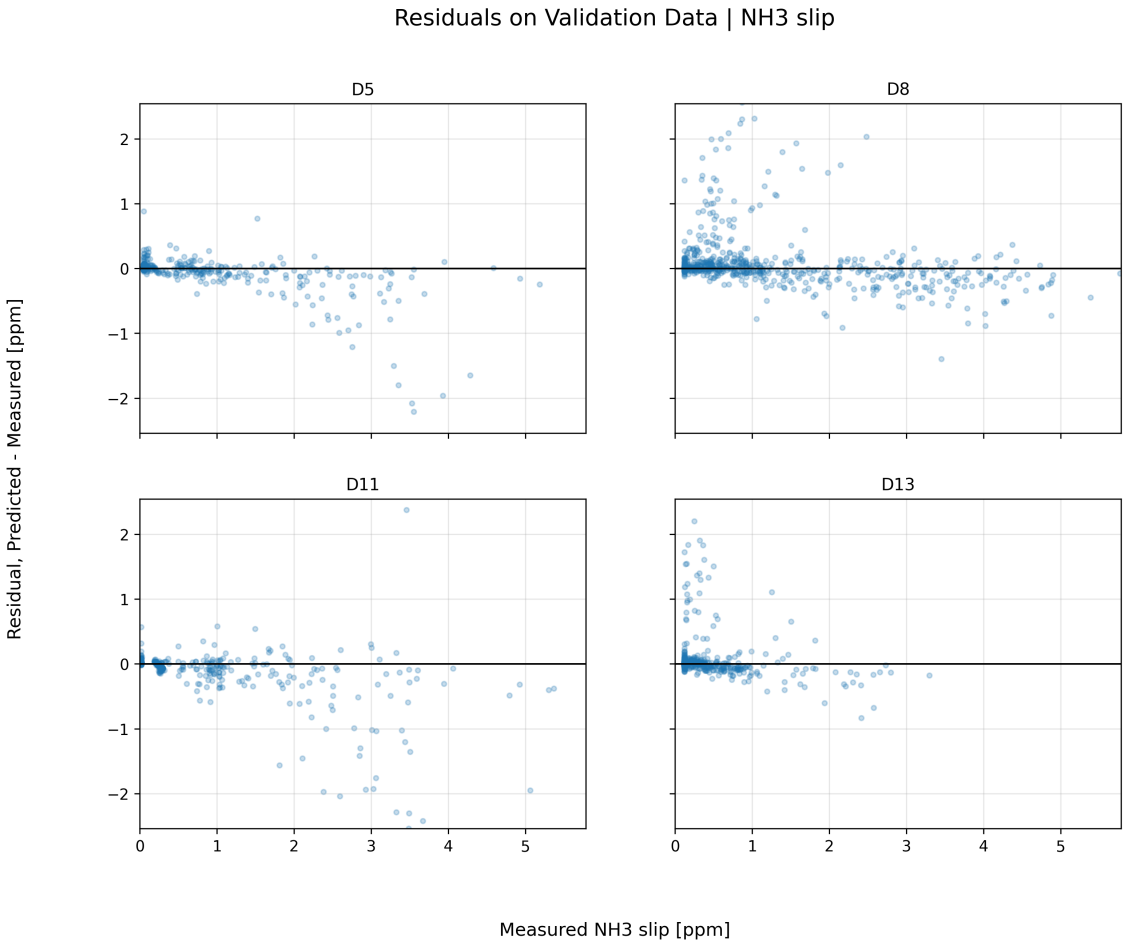


Figure 4.7: Residuals for NH_3 slip prediction using the physics-informed model. Residuals are defined as predicted minus measured concentration.

Figure 4.8 presents the time-series comparison between measured and predicted NH_3 slip. The model follows the overall transient behaviour and captures several major slip events across all engine platforms. D5 shows the strongest agreement, while larger deviations appear for D8, D11, and D13 during transient high-slip events. Both underprediction and overprediction of individual peaks can be observed, indicating that rapid NH_3 storage and release dynamics remain difficult to fully reproduce using static operating-point prediction.

These behaviours are consistent with the residual analysis and indicate that NH_3 prediction remains strongly influenced by catalyst history effects that are not directly represented in the available measurement signals. This makes NH_3 prediction inherently more difficult than NO_x prediction, where the relationship to the current operating point is generally stronger.

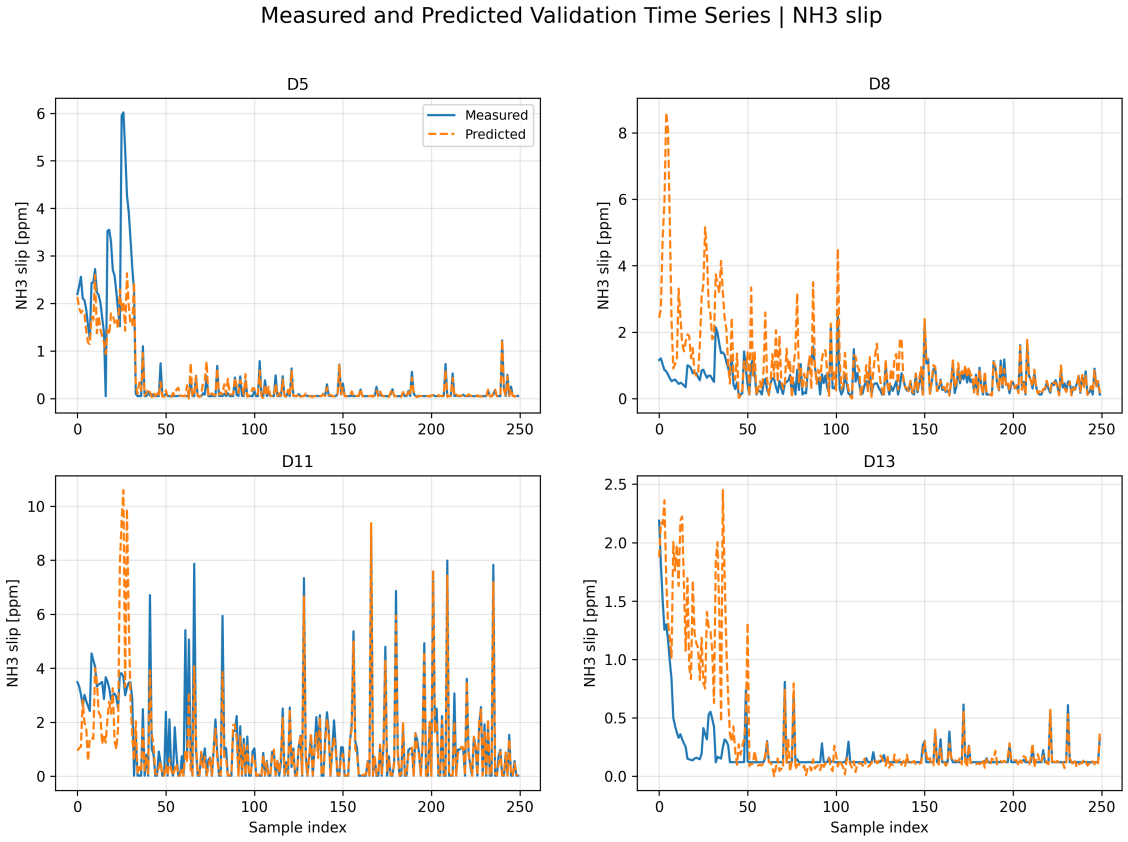


Figure 4.8: Time-series comparison between measured and predicted NH_3 slip concentrations for the physics-informed model.

Overall, the NH_3 results show that the physics-informed model captures the dominant NH_3 slip behaviour across the evaluated engine platforms, although the prediction accuracy remains lower than for NO_x. The main remaining limitation is the prediction of transient high-slip events, indicating that future improvements should focus on dynamic catalyst storage representation.

4.3.3 HC and CO Prediction Results

HC and CO showed smaller improvements from the physics-informed model compared to NO_x and NH_3, which was expected since the added engineered features were mainly focused on SCR-related behaviour. Even so, both emissions showed good overall prediction performance across the four engine platforms, indicating that the modelling framework remained robust also for emissions more strongly influenced by combustion conditions and DOC oxidation. HC prediction was generally more stable and consistent, while CO showed larger variation depending on engine platform and operating conditions.

Figure 4.9 shows measured and predicted system-out HC concentrations for the four engine platforms. The results show strong agreement, with validation R^2 values ranging from 0.818 for D5 to 0.972 for D13. The best performance was obtained for D11 and D13, where most operating points remain close to the ideal prediction line and the prediction spread stays small throughout the operating range. These results suggest that HC was comparatively easier to predict across the evaluated engine platforms. One likely reason is that HC oxidation depends more directly on combustion conditions and DOC conversion behaviour, which are well represented by the available measured variables. HC prediction is also less affected by hidden catalyst storage dynamics and delayed transient behaviour, reducing prediction uncertainty.

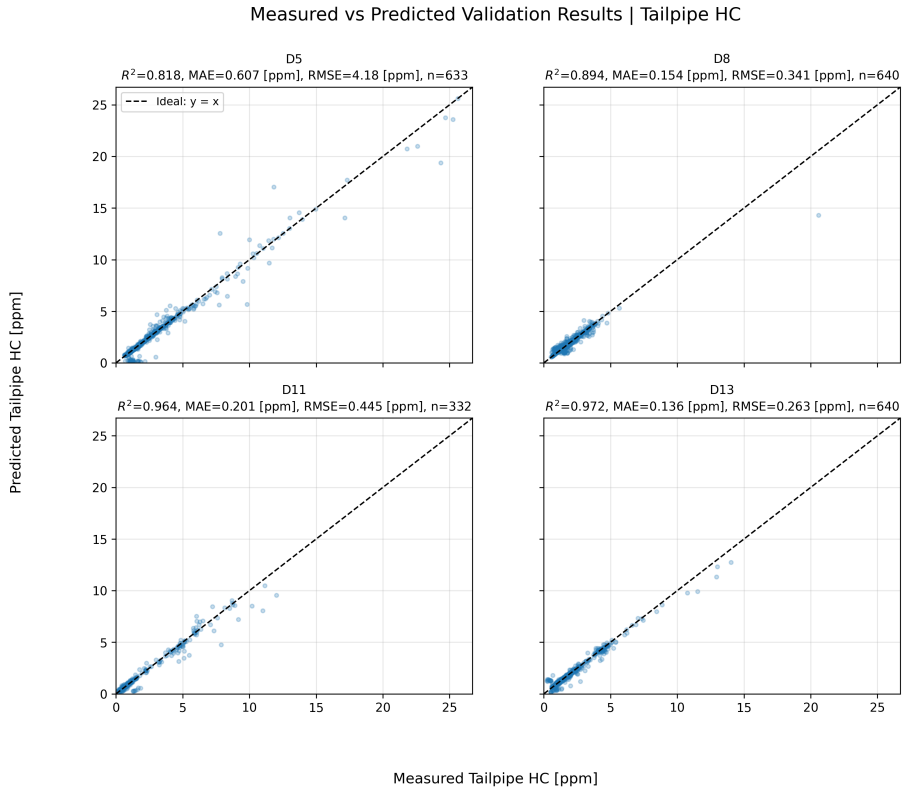


Figure 4.9: Measured versus predicted system-out HC concentrations for the physics-informed model across the four engine platforms.

Figure 4.10 shows the corresponding results for system-out CO concentrations. Compared to HC, the CO prediction results show larger variation between engine platforms and a wider spread around the ideal prediction line. The best performance was obtained for D8, where the model reached an R^2 value of 0.966, while D11 showed the lowest performance with an R^2 value of 0.696.

The larger spread for CO likely reflects the strong temperature dependence of DOC oxidation, particularly during transient and low-temperature operation where small temperature variations can produce large changes in system-out CO concentration. These results suggest that CO prediction is more sensitive to transient operating conditions and short-term combustion variations, which are difficult to fully represent using static regression features.

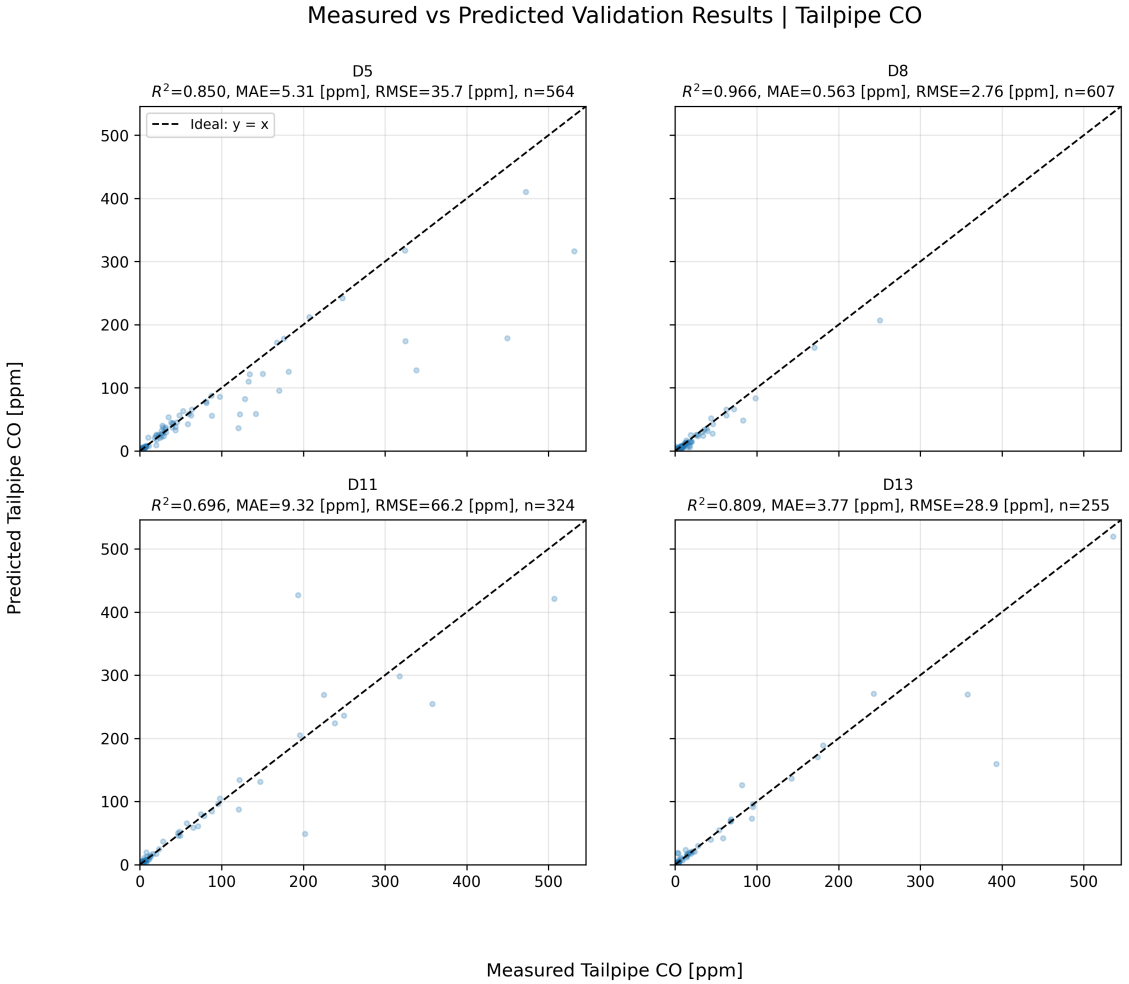


Figure 4.10: Measured versus predicted system-out CO concentrations for the physics-informed model across the four engine platforms.

Overall, the more limited improvement for HC and CO indicates that the current physics-informed feature engineering strategy was more effective for SCR-related emissions than for DOC- and combustion-dominated behaviour. This likely reflects the stronger focus on SCR chemistry in the added engineered features, while HC and CO prediction may require additional physical representation of DOC oxidation, catalyst light-off behaviour, and combustion-related dynamics. Future improvements for these targets may therefore benefit from additional physics-informed variables related to oxidation efficiency, catalyst ageing, and combustion quality.

4.4 Feature Importance and Model Interpretation

To better understand how the physics-informed XGBoost model generated its predictions, feature importance was analysed for all four emission targets. The purpose was to evaluate whether the model relied on physically meaningful relationships rather than mainly statistical correlations.

Figure 4.11 shows the feature importance results for the different system-out emission targets. For NO_x, the most important features were ANR, GHSV, temperature-related variables, and catalyst volume. This is physically reasonable since SCR conversion strongly depends on ammonia availability, residence time, and catalyst temperature. Similar behaviour was observed for NH_3, where temperature-related variables, catalyst volume, and flow-dependent features showed high importance. This supports the earlier observation that NH_3 prediction is strongly influenced by catalyst storage dynamics and transient operating conditions.

For HC, the temperature kinetics term clearly dominated the feature importance results, indicating that HC prediction mainly depended on temperature-driven oxidation behaviour. This is physically expected since HC oxidation in the DOC is highly temperature dependent and comparatively less influenced by hidden catalyst storage dynamics.

For CO, catalyst volume and temperature-related variables showed the highest importance. Compared to NO_x and NH_3, the interpretation is less directly linked to SCR chemistry and more related to DOC oxidation behaviour, particularly during transient and low-temperature operation. This is consistent with the larger variation observed in CO prediction across the engine platforms.

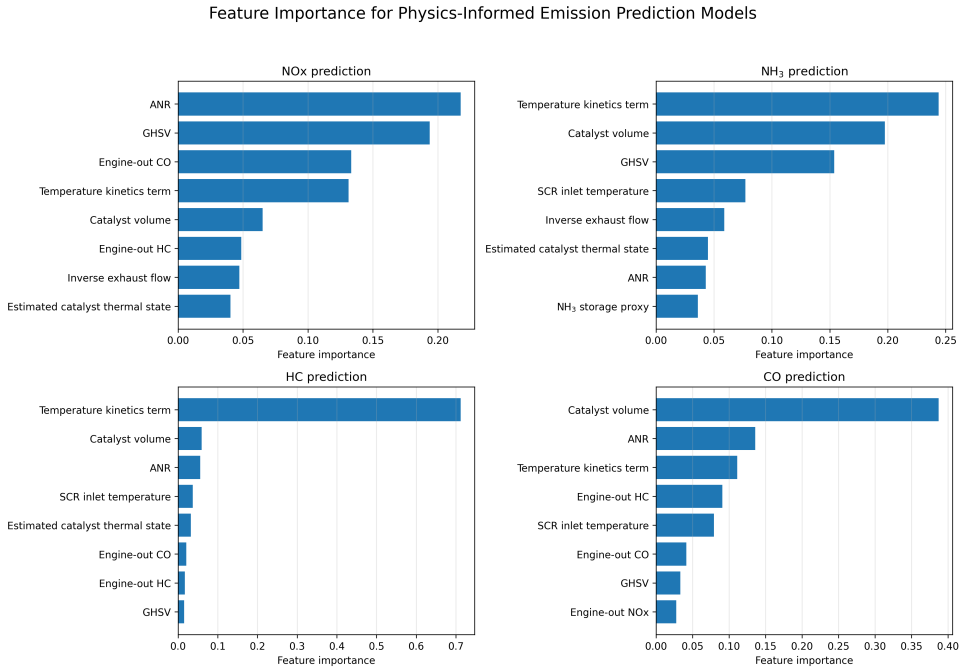


Figure 4.11: Feature importance for the physics-informed models for NO_x, NH_3, HC, and CO prediction.

Overall, the feature importance analysis indicates that the model relied primarily on variables with clear physical meaning rather than hidden statistical correlations. The results also show that the added physics-informed variables were actively used by the model across multiple targets, particularly for emissions related to SCR conversion behaviour. This improves confidence that the learned relationships are physically consistent and therefore more suitable for engineering applications such as virtual testing and aftertreatment analysis.

4.5 Potential for Virtual Testing and SCR Scaling

To evaluate whether the developed physics-informed framework could support early-stage SCR sizing analysis, a catalyst volume sensitivity study was performed using the trained model. The analysis investigated how predicted system-out NO_x and NH_3 changed when catalyst volume was varied for the different engine platforms.

The purpose was not to perform a validated SCR optimisation, but to evaluate whether the model responded with physically reasonable trends when catalyst volume was changed. The results should therefore be interpreted as a model-based sensitivity analysis rather than a final sizing methodology.

Even with these limitations, such an approach could potentially support early-stage virtual testing by identifying promising catalyst design regions before expensive physical testing is performed. The resulting catalyst volume sensitivity behaviour is presented in the following subsections.

4.5.1 Catalyst Volume Sensitivity for NO_x Prediction

Figure 4.12 shows the predicted system-out NO_x concentration as a function of relative catalyst volume for the four engine platforms. For all engines, larger catalyst volume generally resulted in lower predicted NO_x emissions, which is physically reasonable since increased SCR volume provides longer residence time and improved conversion conditions.

The strongest sensitivity was observed for D8 and D13, where predicted NO_x decreased rapidly as catalyst volume approached the baseline region. For D13, the predicted NO_x level decreased from approximately 66 ppm at 70% catalyst volume to around 29 ppm at the baseline size. A similar trend was observed for D8, indicating that these engines operate closer to a volume-sensitive SCR region where catalyst downsizing strongly affects emission performance.

For D5 and D11, the decrease in predicted NO_x was more gradual, suggesting a larger design margin or operating conditions where SCR conversion is less strongly limited by catalyst volume.

Another clear observation is that the NO_x curves become flatter above the baseline region, indicating diminishing returns from further catalyst volume increases. This suggests that the evaluated systems already operate near a region where additional catalyst volume provides only limited extra NO_x reduction under the tested operating conditions.

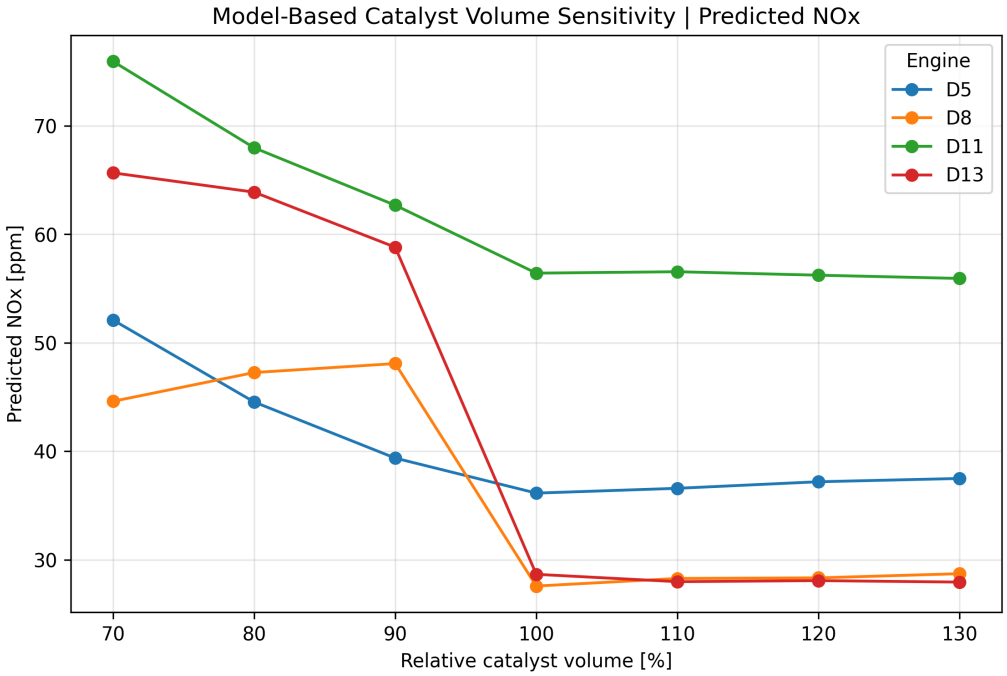


Figure 4.12: Predicted system-out NO_x concentration as a function of relative catalyst volume for the four engine platforms.

Overall, the NO_x sensitivity results show that the model responds with physically consistent trends when catalyst volume is varied, indicating potential usefulness for early-stage virtual screening of SCR sizing sensitivity.

4.5.2 Relative Change from Baseline Configuration

Figure 4.13 presents the relative change in predicted NO_x and NH_3 slip compared to the baseline catalyst volume. The results show that reducing catalyst volume below the baseline causes a clear increase in emissions, particularly for NO_x. The strongest sensitivity was observed for D13 and D8, where catalyst downsizing produced large relative increases in predicted NO_x emissions.

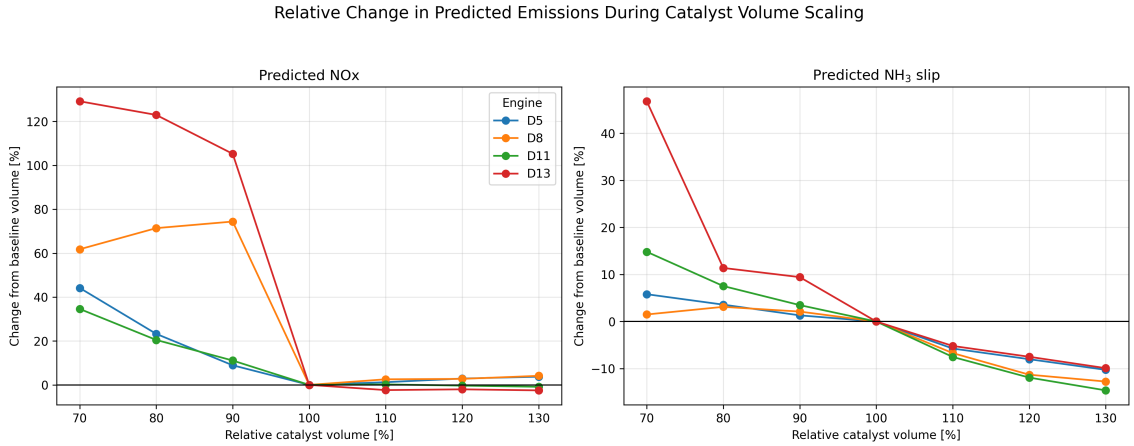


Figure 4.13: Relative change in predicted NO_x and NH_3 slip compared to the baseline catalyst volume (100%).

For D5 and D11, the relative increase in NO_x was smaller when catalyst volume was reduced, suggesting a larger SCR sizing margin under the evaluated operating conditions.

When catalyst volume increased above the baseline level, the improvement became substantially smaller for all engines, indicating diminishing returns from further catalyst oversizing. This behaviour suggests that the baseline configurations already operate near a region where SCR conversion efficiency approaches its practical limit under the evaluated conditions.

It should also be noted that the urea dosing strategy was kept unchanged during the analysis. In a real SCR system, catalyst resizing would likely require recalibrated dosing control, meaning that the predicted trends should be interpreted as relative sensitivity indications rather than final optimized performance levels.

The NH_3 results showed the same overall behaviour, although the sensitivity to catalyst volume was weaker than for NO_x. Reducing catalyst volume increased predicted NH_3 slip, while larger catalyst volumes mainly improved conversion stability

rather than producing large direct reductions in slip.

Another important observation is that the sensitivity differed substantially between engine platforms. This indicates that SCR sizing margins are highly platform dependent and influenced by factors such as engine-out NO_x, exhaust temperature, duty cycle, and dosing strategy.

Overall, the relative scaling analysis shows that the developed model responds with physically reasonable trends when catalyst volume is varied. The results indicate clear potential for future model-based virtual testing and early-stage SCR sizing support before detailed experimental validation.

5

Conclusions and Future Work

This chapter summarizes the main findings of the thesis and discusses future work related to physics-informed emission prediction and SCR sizing applications.

5.1 Conclusions

The results showed that physics-informed feature engineering improved prediction performance mainly for SCR-related emissions such as NO_x and NH_3. The clearest improvements were observed for NH_3 prediction, where the physics-informed model provided substantially improved robustness on unseen datasets, particularly for D8 and D13.

For HC and CO, the improvements were smaller, indicating that the current feature engineering strategy was more effective for SCR-related emissions than for DOC- and combustion-dominated behaviour.

Feature importance analysis further showed that the model relied primarily on variables with clear physical meaning, increasing confidence that the learned relationships were physically consistent rather than purely statistical.

The catalyst volume sensitivity study also showed that the model produced physically consistent trends when SCR volume was varied. Increasing catalyst volume reduced predicted NO_x and NH_3 slip, while diminishing returns were observed above the baseline design region. However, the study should be interpreted as a sensitivity analysis since the urea dosing strategy was kept unchanged during catalyst scaling.

Overall, the study demonstrates that physics-informed machine learning can improve both prediction robustness and physical interpretability for exhaust aftertreatment modelling. The developed framework also shows clear potential for supporting early-stage virtual testing and model-based SCR sizing within industrial engine development.

5.2 Future Work

Although the developed framework showed strong performance and clear industrial relevance, several opportunities remain for further improvement. One important limitation of the present study was the available dataset, since the operating conditions, catalyst configurations, and transient behaviours were limited by existing test campaigns rather than specifically designed for machine learning model development. Future work should therefore focus on obtaining larger and more diverse datasets covering a wider range of operating conditions, engine loads, catalyst volumes, and aftertreatment configurations. In particular, additional transient data would likely improve prediction robustness for difficult operating regions such as NH_3 slip peaks and rapid NO_x changes.

An important methodological improvement would be the use of more robust validation strategies such as GroupKFold or leave-one-engine-out validation instead of ShuffleSplit cross-validation. Stronger separation between engine platforms during validation would provide a more realistic estimate of generalisation performance for unseen applications.

The hyperparameter optimisation could also be extended further. Future work may include broader optimisation of subsampling, regularisation, and tree-structure parameters, as well as Bayesian optimisation methods to improve both efficiency and model quality.

An even stronger improvement would be the design of dedicated experimental test campaigns specifically intended for model development. Instead of relying only on existing production test data, controlled experiments could be performed where parameters such as urea dosing, catalyst volume, exhaust temperature, and engine-out NO_x are systematically varied. This would provide significantly better training data for capturing catalyst dynamics, ammonia storage effects, and SCR conversion limits. Dedicated measurements before and after the ammonia slip catalyst would also improve NH_3 model interpretation and reduce uncertainty related to slip catalyst behaviour.

Future work should also include a stronger connection to emission legislation and certification requirements. A highly relevant next step would be to integrate regulatory constraints such as EU Stage V directly into the SCR sizing framework. By combining engine simulation data with the developed emission prediction model, it would be possible to estimate the minimum SCR volume required to satisfy legal NO_x and NH_3 limits under representative operating cycles.

Another important extension would be the development of more advanced hybrid modelling approaches. A multi-stage structure where NO_x conversion efficiency is predicted before final tailpipe emissions could improve physical interpretability compared to direct tailpipe prediction and provide clearer engineering understanding of how catalyst size and dosing strategy affect conversion behaviour.

Since NH_3 prediction remains strongly affected by hidden catalyst states and delayed storage-release dynamics, future work should also investigate dynamic hybrid models that combine machine learning with reduced-order physical catalyst models or state-estimation methods. Such approaches could improve prediction accuracy during highly transient operation and better represent internal catalyst storage behaviour that cannot be fully captured using static feature engineering alone.

Additional future work could include catalyst aging effects, uncertainty estimation, and alternative fuel operation such as HVO or hybrid fuel strategies. Prediction uncertainty is particularly important for industrial SCR sizing applications, where engineering decisions depend not only on mean predictions but also on confidence in emission compliance margins.

Overall, the developed framework provides a strong foundation for future work toward physics-informed, model-based SCR sizing and virtual exhaust aftertreatment optimisation within industrial engine development. With improved validation, richer datasets, and stronger integration of physical catalyst behaviour and emission legislation, the approach has strong potential to become a practical engineering tool for future engine platform development.

Overall, the developed framework provides a strong foundation for future physics-informed and model-based SCR sizing methods within industrial exhaust aftertreatment development. With improved validation strategies, richer datasets, and stronger integration of catalyst dynamics and emission legislation, the approach shows strong potential as a practical engineering support tool for future engine platform development.

Bibliography

[1] W. A. Majewski, "EU: Nonroad Engine Emission Standards," DieselNet Technology Guide, DieselNet. [Online]. Available: https://dieselnet.com/standards/eu/nonroad.php. Accessed: May 18, 2026.

[2] S. S. Ravi, S. Osipov, and J. W. G. Turner, "Impact of Modern Vehicular Technologies and Emission Regulations on Improving Global Air Quality," *Atmosphere*, vol. 14, no. 7, p. 1164, Jul. 2023. [Online]. Available: https://www.mdpi.com/2073-4433/14/7/1164. Accessed: May 18, 2026.

[3] W. A. Majewski, "What Are Engine Emissions?" DieselNet Technology Guide, DieselNet. [Online]. Available: https://dieselnet.com/tech/emissions.php. Accessed: May 18, 2026.

[4] "Sustainability – We Want to Lead the Way," Volvo Penta. [Online]. Available: https://www.volvopenta.com/en-us/about-us/sustainability/. Accessed: May 18, 2026.

[5] W. A. Majewski, "Selective Catalytic Reduction," DieselNet Technology Guide, DieselNet. [Online]. Available: `https://dieselnet.com/tech/cat_scr.php`. Accessed: May 18, 2026.

[6] W. A. Majewski, "Emission control catalysts," DieselNet Technology Guide, DieselNet. [Online]. Available: https://dieselnet.com/tech/catalysts.php. Accessed: Jan. 28, 2026.

[7] K. Jiang, F. Yan, and H. Zhang, "Data-driven control of automotive diesel engines and after-treatment systems: State of the art and future challenges," *Proc. Inst. Mech. Eng., Part D: J. Automobile Engineering*, vol. 237, no. 9, pp. 2083–2098, Aug. 2023. [Online]. Available: https://journals.sagepub.com/doi/10.1177/09544070221104893. Accessed: May 18, 2026.

[8] C. Meng, S. Seo, D. Cao, S. Griesemer, and Y. Liu, "When physics meets machine learning: A survey of physics-informed machine learning," arXiv:2203.16797 [cs.LG], Apr. 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2203.16797. Accessed: May 18, 2026.

[9] "What Are Physics-Informed Neural Networks (PINNs)?," MathWorks. [Online]. Available: `https://se.mathworks.com/discovery/physics-informed-neural-networks.html`. Accessed: May 18, 2026.

[10] European Parliament and Council, "Regulation (EU) 2016/1628 of 14 September 2016 on requirements relating to gaseous and particulate pollutant emission limits and type-approval for internal combustion engines for non-road mobile machinery," Official Journal of the European Union. [Online]. Available: `https://eur-lex.europa.eu/eli/reg/2016/1628/oj`. Accessed: May 18, 2026.

[11] W. A. Majewski, "Diesel Oxidation Catalysts," DieselNet Technology Guide, DieselNet. [Online]. Available: https://dieselnet.com/tech/cat_doc.php. Accessed: May 18, 2026.

[12] F. Posada, B. Sanchez, and J. Miller, "Costs of Emission Reduction Technologies for Diesel Engines Used in Non-Road Vehicles and Equipment," International Council on Clean Transportation (ICCT), Jul. 2018. [Online]. Available: https://theicct.org/wp-content/uploads/2021/06/Non_Road_Emission_Control_20180711.pdf. Accessed: May 18, 2026.

[13] F. Posada, "EU Stage V Emission Standards for Non-Road Mobile Machinery," International Council on Clean Transportation (ICCT), Nov. 2016. [Online]. Available: https://theicct.org/sites/default/files/publications/EU-Stage-V_policy%20update_ICCT_nov2016.pdf. Accessed: May 18, 2026.

[14] W. A. Majewski, "Engine Emission Control," DieselNet Technology Guide, DieselNet. [Online]. Available: https://dieselnet.com/tech/engine_emission-control.php. Accessed: May 18, 2026.

[15] T. Feng, J. Wang, S. Shuai, J. Wang, and X. Li, "The characteristics of ammonia storage and the ammonia storage control strategy in the SCR system of diesel engine," *International Journal of Automotive Technology*, vol. 16, no. 4, pp. 593–600, Aug. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1226086X15000404. Accessed: May 18, 2026.

[16] W. A. Majewski, "Engine Fundamentals," DieselNet Technology Guide, DieselNet. [Online]. Available: https://dieselnet.com/tech/diesel_fundamentals.php. Accessed: May 18, 2026.

.

[17] W. A. Majewski, "Diesel Catalysts," DieselNet Technology Guide, DieselNet. [Online]. Available: https://dieselnet.com/tech/cat_diesel.php. Accessed: May 18, 2026.

[18] W. A. Majewski, "Diesel Particulate Filters," DieselNet Technology Guide, DieselNet. [Online]. Available: https://dieselnet.com/tech/dpf.php. Accessed: May 18, 2026.

[19] W. A. Majewski, "SCR Systems for Diesel Engines," DieselNet Technology Guide, DieselNet. [Online]. Available: https://dieselnet.com/tech/cat_scr_diesel.php. Accessed: May 18, 2026.

[20] W. A. Majewski, "Urea Dosing and Injection Systems," DieselNet Technology Guide, DieselNet. [Online]. Available: https://dieselnet.com/tech/cat_scr_diesel_urea_dosing.php. Accessed: May 18, 2026.

[21] "Industrial Engines," Volvo Penta. [Online]. Available: https://www.volvopenta.com/industrial/industrial-engines/. Accessed: May 18, 2026.

[22] "Steady-State vs. Transient Modeling in CFD: Choosing the Right Approach," SimuLarge. [Online]. Available: https://www.simularge.com/blog/steady-state-vs-transient-modeling-in-cfd-choosing-the-right-approach. Accessed: May 18, 2026.

[23] A. Lindén, "Modeling and Control of SCR Aftertreatment Systems," Master's Thesis, Chalmers University of Technology, 2023. [Online]. Avail-

able: https://research.chalmers.se/publication/541251/file/541251_Fulltext.pdf. Accessed: May 18, 2026.

[24] M. Alhassan and A. B. M. Said, "Imbalanced Data Problem in Machine Learning: A Review," ResearchGate preprint, 2024. [Online]. Available: https://www.researchgate.net/publication/388208416_Imbalanced_Data_problem_in_Machine_Learning_A_review. Accessed: May 18, 2026.

[25] M. Karniadakis et al., "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, pp. 422–440, 2021. [Online]. Available: https://www.nature.com/articles/s42254-021-00314-5. Accessed: May 18, 2026.

[26] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD '16)*, San Francisco, CA, USA, Aug. 2016, pp. 785–794. [Online]. Available: https://doi.org/10.1145/2939672.2939785. Accessed: May 18, 2026.

[27] "XGBoost Documentation," XGBoost. [Online]. Available: https://xgboost.readthedocs.io/en/release_3.2.0/. Accessed: May 18, 2026.

[28] A. Paul, A. Mishra, and S. Choudhary, "A Physics-Informed Feature Engineering Approach to Use Machine Learning with Limited Amounts of Data for Alloy Design: Shape Memory Alloy Demonstration," ResearchGate preprint, 2020. [Online]. Available: https://www.researchgate.net/publication/339710705_A_physics-informed_feature_engineering_approach_to_use_machine_learning_with_limited_amounts_of_data_for_alloy_design_shape_memory_alloy_demonstration. Accessed: May 18, 2026.

[29] "Hybrid Modeling in Chemical Processes," RISE Research Institutes of Sweden. [Online]. Available: https://www.ri.se/en/hybrid-modeling-in-chemical-processes. Accessed: May 18, 2026.

[30] "How Aftertreatment Systems Work," Cummins Inc. [Online]. Available: https://www.cummins.com/en-na/components/aftertreatment/how-it-works. Accessed: May 18, 2026.

[31] J. Girard, R. Snow, G. Cavaiao, and C. Lambert, "The Influence of Ammonia to NOx Ratio on SCR Performance," SAE Technical Paper 2007-01-1581, 2007. [Online]. Available: https://saemobilus.sae.org/papers/influence-ammonia-nox-ratio-scr-performance-2007-01-1581. Accessed: May 18, 2026.

[32] T. Srivastava, "12 Important Model Evaluation Metrics for Machine Learning Everyone Should Know," Analytics Vidhya, 2026. [Online]. Available: https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/. Accessed: May 18, 2026.

A

Data Processing and Model Implementation

A.1 Signal Alias Mapping

Since identical physical quantities can appear under different channel names depending on engine platform, ECU configuration, and test setup, an alias-based parameter mapping was applied during data extraction. This ensured that the same physical variable was consistently represented by a single canonical variable name across all datasets. The alias mapping was implemented during export from PySakura to pandas DataFrames and allowed direct comparison across multiple engine families and test campaigns. Table A.1 presents the main canonical variables used in this work together with the corresponding alternative signal names found in the original Sakura XML files.

Table A.1: Signal alias mapping used during data extraction and preprocessing.

Canonical variable	**Alternative signal names**
conc2_nox_so	Conc2_NOx
conc2_hc_so	Conc2_HC
conc2_co_so	Conc2_CO
conc_nh3_so	Conc_NH3
conc2_n2o_so	Conc2_N2O
flow_exhaust	Flw_Exhaust_M, Flw_Exhaust, ECU059
t_exhaust	T_Exhaust, T_Exhaust_pipe, T_A_Turb, ECU064
t_b_dpf	T_B_DPF, asm_SensValHcHeatTemp, ECU063
t_a_dpf	T_A_DPF, asm_SensValDPF, ECU065
t_b_scr	T_B_SCR, T_SCR_In, ECU063
t_a_scr	T_A_SCR
p_exhaust	P_Exhaust
conc_no_eo	Conc_NO
conc_no2_eo	Conc_NO2
conc_nox_eo	Conc_NOx
conc_co_eo	Conc_CO
conc_hc_eo	Conc_HC
conc_o2_eo	Conc_O2
urea_demand	atns_UreaMassFlowDemand, ECU017

A.2 Physics-Informed Feature Definitions

The physics-informed features were calculated during preprocessing before model training. These features were derived from measured engine and EATS signals in order to represent important SCR-related relationships such as NO_x load, catalyst residence time, ammonia availability, thermal behaviour, and simplified reaction activity.

Table A.2 summarises the derived features used in the physics-informed model. Small numerical constants were added in some expressions to avoid division by zero and improve numerical stability.

Table A.2: Physics-informed features used in the final model.

Feature	Definition
`flw_nox_gs`	$\dot{m}_{exh} \cdot 1000 \cdot c_{NOx,eo} \cdot 10^{-6} \cdot \frac{M_{NO_2}}{M_{air}}$
`rho_exhaust`	$\rho = \frac{p_{exh}}{RT_{SCR}}$
`GHSV`	$\frac{\dot{m}_{exh}/\rho \cdot 3600}{V_{cat}/1000}$
`ANR`	$\frac{\texttt{urea\_demand}}{\texttt{flw\_nox\_gs}+10^{-7}}$
`no2_fraction`	$\frac{c_{NO2,eo}}{c_{NOx,eo}+10^{-7}}$
`t_stone_est`	EWMA of `t_b_scr` with span 60
`nh3_storage_proxy`	Rolling sum of `urea_demand` with window 100
`kinetics_term`	$\exp\left(-\frac{5000}{T_{SCR}}\right) \cdot \mathbb{1}(t_{b,scr} \geq 200°C)$
`Da_number`	$\frac{\texttt{kinetics\_term}}{\texttt{GHSV}+10^{-7}}$
`inv_flow`	$\frac{1}{\dot{m}_{exh}+10}$
`crys_risk`	$\mathbb{1}(t_{b,scr} < 250°C$ and `urea_demand` $> 0.5)$

In the equations, $\dot{m}_{exh}$ represents exhaust mass flow, $c_{NOx,eo}$ represents engine-out NO_x concentration in ppm, $c_{NO2,eo}$ represents engine-out NO_2 concentration in ppm, V_{cat} represents catalyst volume in litres, and T_{SCR} represents SCR inlet temperature in Kelvin. The constants used were $M_{NO_2} = 46.01$ g/mol, $M_{air} = 28.97$ g/mol, and $R = 287.05$ J/(kg K).

If `t_b_scr` was not available, `t_exhaust` was used as a fallback temperature signal. If `conc_no2_eo` was not available, it was estimated as the difference between total NO_x and NO concentration when possible. Negative sensor values were clipped where physically required before feature calculation.

A.3 Dataset Processing Workflow

The following simplified code structure shows how selected Sakura measurements were extracted, filtered, reduced, and exported to CSV files. Internal file paths and complete test lists are omitted for confidentiality and readability.

Listing A.1: Simplified dataset processing workflow.

```
# Extract selected subtests, apply alias mapping, filter signals, reduce M1 data,
    and export one processed CSV file.

all_subtests = []

for test in selected_tests:

    df = load_subtest_dataframe(
        file_path=test.file_path,
        subtest_index=test.subtest_index,
        test_number=test.test_number,
        mode=test.mode,
        alias_map=ALIAS_MAP,
        catalyst_volume=test.catalyst_volume
    )

    if df is None:
        continue

    # Smooth numeric signals, but keep metadata unchanged
    df = apply_ewma_filter(df, span_map, default_span=10)

    # Reduce transient M1 data only
    if test.mode == "M1":
        df = cluster_m1_data(
            df,
            features=["flow_exhaust", "t_b_scr", "conc_nox_eo"],
            n_clusters=300
        )

    # LS data already represents quasi-stationary points and is therefore kept
        without clustering.
    df["description"] = test.description
    all_subtests.append(df)

df_processed = pd.concat(all_subtests, ignore_index=True)
df_processed.to_csv(output_path, index=False)
```

A.4 Model Training and Validation Workflow

The following simplified code structure shows how the processed training files were combined, how physics-informed features were added, and how the XGBoost models were trained and evaluated. Internal file paths are omitted.

Listing A.2: Simplified model training and validation workflow.

```
# Load processed training data from all engine platforms

def load_and_prepare(paths):
    dataframes = []

    for path in paths:
        df = pd.read_csv(path)
        df = add_physics_features(df)
        dataframes.append(df)

    return pd.concat(dataframes, ignore_index=True)


df_train = load_and_prepare(training_paths)

features = [
    "flow_exhaust", "inv_flow", "t_b_scr", "t_stone_est",
    "conc_nox_eo", "GHSV", "ANR", "kinetics_term",
    "Da_number", "no2_fraction", "volym_kat_val",
    "nh3_storage_proxy", "conc_co_eo", "conc_hc_eo"
]

targets = [
    "conc2_nox_so",
    "conc2_hc_so",
    "conc2_co_so",
    "conc_nh3_so"
]

param_dist = {
    "n_estimators": [500, 800],
    "max_depth": [6, 8],
    "learning_rate": [0.03, 0.05]
}

trained_models = {}

for target in targets:

    df_clean = df_train.dropna(subset=features + [target])
    X = df_clean[features]
    y = df_clean[target].clip(lower=0)

    # Target-specific transformations
    if target in ["conc2_hc_so", "conc2_co_so", "conc_nh3_so"]:
        y_train = np.log1p(y)
    else:
        y_train = y

    sample_weights = (
        df_clean["conc_nox_eo"] + 1.0
    ) / df_clean["conc_nox_eo"].mean()

    search = RandomizedSearchCV(
        estimator=XGBRegressor(n_jobs=-1, random_state=42),
        param_distributions=param_dist,
        n_iter=5,
        cv=ShuffleSplit(n_splits=3, test_size=0.3, random_state=42),
        scoring="r2",
        random_state=42
```

```
    )

    search.fit(X, y_train, sample_weight=sample_weights)
    trained_models[target] = search.best_estimator_


# External validation on separated validation files

for validation_path in validation_paths:

    df_val = pd.read_csv(validation_path)
    df_val = add_physics_features(df_val)

    predictions = {}

    for target, model in trained_models.items():

        y_pred = model.predict(df_val[features])

        if target in ["conc2_hc_so", "conc2_co_so", "conc_nh3_so"]:
            y_pred = np.expm1(y_pred)

        y_pred = np.maximum(0, y_pred)
        predictions[target] = y_pred

    evaluate_predictions(df_val, predictions)
```

DEPARTMENT OF MECHANICS AND MARITIME SCIENCE
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY