Journal Title
XX(X):1–18

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

A Correlation Based Model Order Reduction Approach for a Diesel Engine NOx and BMEP Dynamic Model Using Machine Learning

Armin Norouzi[1], Masoud Aliramezani[1], and Charles Robert Koch[1]

Abstract

A correlation based Model Order Reduction (MOR) algorithm is developed using Support Vector Machine (SVM) to model $\mathrm{NO_x}$ emission and Break Mean Effective Pressure (BMEP) of a medium-duty diesel engine. The SVM-based MOR algorithm is used to reduce the number of features of a 34-feature Full-Order Model (FOM) by evaluating the regression performance of the SVM-based model. Then, the SVM-based MOR algorithm is used to reduce the number of features of the FOM. Two models for $\mathrm{NO_x}$ emission and BMEP are developed via MOR, one complex model with high-accuracy, called High-Order Model (HOM), and the other with an acceptable accuracy and a simple structure, called Low-Order Model (LOM). The HOM has 29 features for $\mathrm{NO_x}$ and 20 features for BMEP, while the LOM has 9 features for $\mathrm{NO_x}$ and 6 features for BMEP. Then, the steady-state LOM and HOM are implemented in a Nonlinear Control-Oriented Model (NCOM). To verify the accuracy of NCOM, a fast response electrochemical $\mathrm{NO_x}$ sensor is used to experimentally study the engine transient $\mathrm{NO_x}$ emissions. The HOM and LOM SVM models of $\mathrm{NO_x}$ and BMEP are compared to a conventional Artificial Neural Network (ANN) with one hidden layer. The results illustrate that the developed SVM model has shorter training times (5 to 14 times faster) and higher accuracy especially for test data compared to the ANN model. A control-oriented model (COM) is then developed to predict the dynamic behavior of the system. Finally, the performance of the LOM and HOM are evaluated for different raising and falling input transients at four different engine speeds. The transient test results validate the high accuracy of the HOM and the acceptable accuracy of LOM for both $\mathrm{NO_x}$ and BMEP. The HOM is proposed as an accurate virtual plant while the LOM is suitable for model-based controller design.

Keywords

Machine Learning, Support Vector Machine, Model Order Reduction, Diesel Engine, Emissions, Control Oriented Model

Introduction

Diesel engine $\mathrm{NO_x}$ emission

The long lifespan, high efficiency, fuel economy advantages and the wide range of operating conditions have made Direct Injection (DI) Diesel engines interesting for both stationary power generation and the transportation industry[1,2]. The high combustion temperature and the lean air-fuel mixture of Diesel engines lead to a relatively high $\mathrm{NO_x}$ emission. Reducing diesel engine $\mathrm{NO_x}$ and particulate matters (PM) emission to meet the stringent emission regulations has always been a challenge for the automotive industry as typically reducing one of them results in increasing the other one[3,4]. This is mainly due to the inverse effect of in-cylinder peak temperature and air fuel ratio on $\mathrm{NO_x}$ and PM[5–7]. Therefore, to develop an effective and practical $\mathrm{NO_x}$ control strategy, all other gaseous and PM emissions should also be considered.

The $\mathrm{NO_x}$ components of diesel engine exhaust gas typically contains approximately $1/7$ to $1/3$ $\mathrm{NO_2/NO}$ fraction[8] while $\mathrm{NO_2}$/NO ratio increases after the Diesel Oxidation Catalyst (DOC) to approximately one[9]. Different methods have been used by the automotive industry to reduce diesel engine $\mathrm{NO_x}$ emission including Exhaust Gas Recirculation (EGR)[10], Low Temperature Combustion (LTC)[11,12] and most effectively, urea-based Selective Catalytic Reduction (SCR)[13–15]. Although all these methods can help reducing the engine tailpipe $\mathrm{NO_x}$ emissions, it has become more difficult to keep pace with increasingly stringent emission regulations by only using the conventional $\mathrm{NO_x}$ reduction approaches[16,17]. To address this issue, more complex engine control strategies and after treatment systems are needed[18–20]. This requires a complex and flexible engine dynamic model that captures the nonlinear relation between engine operating parameters, engine performance and engine emissions.

Use of machine learning for modelling internal combustion engines

Data-driven models have become of especial interest to many of the researchers in the last two decades

[1] Mechanical Engineering Department, University of Alberta, Edmonton, Canada

Corresponding author:
Armin Norouzi, Mechanical Engineering Department, University of Alberta, Edmonton, T6G 2R3, AB, Canada.
Email: norouziy@ualberta.ca

as an efficient way to predict, optimize or diagnose the performance of internal combustion engines (ICEs) mostly using Artificial Neural Network (ANN)[21–23].Machine learning is increasingly used for ICEs optimization and calibration. Machine learning methods, along with the state-of-the-art optimization algorithms such as Genetic Algorithm (ML-GGA) were also used for ICE optimization. For instance, a ML-GGA algorithm was used for optimizing the operating conditions of a heavy-duty engine, and the results showed an improved performance compared to Computational fluid dynamics (CFD) based methods[24,25]. Ensemble methods, using multiple learning algorithms to obtain a more accurate prediction performance, were also used for ICEs. A novel active learning method, Active-O, was developed for ICE optimization and and has shown a better performance compared with the GA, micro GA, and particle swarm optimization (PSO)[26].

Support Vector Machine (SVM) is another popular data-driven method that is now being increasingly used for modeling internal combustion engines mostly for steady-state prediction of engine performance or emissions. SVM is a machine learning technique which is capable of modeling complex and non-linear input-output relations based on a sufficiently large training data set[27–29]. This approach provides a black box model without directly involving physical understanding of the system but can be accurately trained if the model features are selected appropriately[30–34]. The simple structure of SVM, especially by using linear kernel SVM when compared to ANN, allows for improved learning performance which, when combined with the rapid in-cycle Field Programmable Gate Arrays (FPGAs) based calculations, could allow for online learning of in-cycle combustion metrics[35,36].

A NO_x prediction model was developed for a hydrogen-enriched compressed natural gas engine using an optimal SVM method where particle swarm optimization (PSO) was used to find the regulatory parameters of SVM[37]. Also, the effect of SVM model parameters such as penalty factor kernel, insensitive band loss function, and the training sample size was studied in[37]. An optimal SVM for diesel engine NO_x prediction was developed by Liu et al[38], where the Genetic Algorithm (GA) was used to find a regulatory parameter of SVM. Principal Component Analysis (PCA) was used for dimension reduction and less than 5% of information was lost during the data extrusion process[38].

An ANN and an SVM model are developed by Niu et al[39] to predict the performance and emission of a marine diesel engine. The developed ANN and SVM were compared and is was shown that for a limited number of experimental data, SVM has a better performance in finding the global optimum solution compared to the ANN[39]. The prediction of a spark-ignition engine fueled with butanol-gasoline blends was developed using SVM, and the high performance of the SVM method to predict engine parameters was illustrated in[40]. Predicting NO_x emission using SVM method is not limited to internal combustion engines. SVM has also been used to predict NO_x emissions of gas turbines[41] and coal-fired utility boilers[42]. The Least-Square type SVM (LS-SVM) has also been used for NO_x emission prediction[43–45]. In the LS version of SVM, a set of linear optimization methods is solved instead of a quadratic optimization problem. Also, instead of including inequality constraints, equality constraints are considered in the optimization problems of LS-SVMs[46].

Model order reduction (MOR) and Feature Selection (FS)

The main disadvantage of using such a black-box model is the risk of overfitting the model, particularly when a large number of features are used[47]. To address this issue, different methods have been used to reduce the number of negligible features through Model Order Reduction (MOR), also called Feature Selection (FS). The MOR and FS use the same concept to select more significant features; however, different approaches are developed to enhance their performance. In FS, the more significant features with a high score of a so called merit function are selected. Similarly, in MOR, the less significant features are removed based on a pre-defined criteria. Typical feature selection steps are: 1) calculating the overall effects of features on the target and considering it as a score, 2) selecting the feature with the largest score and adding it to the set of the selected features, 3) recalculating the score, 4) repeat 2 and 3 until certain number of features are selected[48].

Different methods have been developed for FS problem such as minimum-redundancy-maximum-relevance (mRMR)[49], conditional mutual information[50], joint mutual information[51], and correlation feature selection[52]. In a correlation-based feature selection approach, typically, a merit function is defined based on the features-output correlation, and a criteria for feature selection is defined to maximize the merit function. By maximizing the merit function, the feature with a high score (high value of merit function) is selected as a feature[52]. A feature selection for the SVM algorithm in time series forecasting was proposed in[53], where hyperparameters were optimized based on the Genetic Algorithm (GA). Then, by defining a scaling factor for system features, the features with smaller scaling factor were removed. In other words, as the features with small scaling factors have less information about the system, they can be removed without any significant accuracy loss[53]. An SVM-based feature selection method was proposed for electric load forecasting in[54] where the mean absolute percentage error (MAPE) was used for the algorithm stoppage criteria[54].

Similarly, the MOR technique is used for dimensionality reduction but from a slightly different perspective. The difference between MOR and FS techniques is that in MOR techniques, new features are created based on the original features but in a lower dimension set of information with a small (but acceptable) loss of the total information. In other words, the MOR techniques are equivalent to so called feature extraction methods[55]. The principal component analysis (PCA) is one of most well-known feature extraction algorithms in which the co-variance matrix and its eigenvalues and eigenvectors are used to find the "principal components." In this method, only the projection of data into principal component will be considered as a new data set (lower than raw data set)[55,56]. Generally, MOR approaches can be classified into three main groups. The first MOR approach uses frequency domain. Some of the methods used in this approach are eigenmode analysis, moment matching category, singular value decomposition (SVD), proper

orthogonal decomposition (POD), balanced truncation, and Hankel approximation[57]. The other popular MOR class includes time-domain approaches, such as Chebyshev and Wavelet[58,59]. Also, different Machine Learning Techniques were used for MOR, including Neural Network, Genetic Algorithm, Fuzzy Logic, Particle swarm optimization (PSO), simulated annealing, and Ant-Colony[57,60,61].

A wide range of methods have been used in the literature for feature selection; however, SVM-based model order reduction or feature selection has rarely been studied. In general, the SVM algorithm for regression is an optimization problem that finds a model for a given data set by solving a trade-off optimization problem between model smoothness and tolerated outlier data[62]. By adjusting the regulatory parameter of this trade-off, model features containing less information compared to the other features can be removed. Hence, the SVM regulatory parameter for smoothness and tolerated error can be used to develop a MOR algorithm.

Motivation and objectives

Several phenomenological and physical control-oriented models of diesel engine performance and emission have been developed for feedback control of emissions[63–65]. Physical and chemical sub-models such as a spray model, mixture formation, ignition delay, combustion characteristics and emission formation can be included in a physics based model providing insight into the combustion phenomena that affects gaseous and particulate emissions. However, despite the ongoing advancements in the physics based combustion and emission models, they can not capture all the complex phenomena that take place inside the combustion chamber to exactly predict the emission levels. In addition, developing such models requires high effort and computing power. This makes complex physics-based models less desirable or unsuitable for real time applications[66].

While the physics-based combustion and emission modeling approach provides physical insight, an accurate detailed 3D combustion simulation model is too computationally expensive for model-based calibration. Using simpler types of physics based models can result in lower emission prediction accuracy[66,67]. Machine Learning based model that is trained with appropriate data and sufficiently large data size, can provide an accurate, flexible and fast data-driven prediction[68]. These data-driven models are typically faster than physics based approaches and can be easily trained using experimental data[66]. Developing machine learning based engine models also requires a deep understanding of the system and the modeling approach. Most of the state-of-the-art machine learning techniques have a large number of design parameters such as the model features, kernel types, and the values of hyperparameters. All of these design parameters must be thoroughly examined to optimize the performance of the model for engine applications. This paper, provides comprehensive examination of the design parameters of a SVM model to be used for internal combustion engine emissions and performance prediction.

Studies carried out in the literature have shown that SVM is a promising method for engine emission and performance prediction due to its high capability of converging to global optimums with relatively smaller size of training datasets compared to other data-driven approaches used for internal combustion engines. First, most of the data-driven models developed for internal combustion engines have used non-linear kernels without exclusively investigating feature interactions. Second, none of these models propose a control-oriented model to be used for engine control or observer design. In addition, use of a correlation-based feature selection technique provides more insight into selecting the most important features without manipulating any of the primary features or their interactions.

To improve the performance of an engine control strategy using emission feedback in the controller is used. Due to the advancements in the sensor industry, fast response production electrochemical $\mathrm{NO_x}$ sensors can now provide accurate input for engine feedback control. Therefore, developing accurate control-oriented model (COM) and virtual plant (VP) to be used as a model observer is essential. An accurate COM of the engine that includes emissions can be used to design a model-based controller or model-based observer. A model-based observer design can play an essential role in engine feedback control systems, especially to detect any fault or malfunction in the emission sensors[69].

A full-order model (FOM) SVM is first developed to predict engine $\mathrm{NO_x}$ emission and brake mean effective pressure (BMEP) at steady-state conditions. The steady-state model is trained using a sufficiently large number of engine data points and then validated at different engine test points. To minimize the risk of overfitting the model, a new correlation-based feature selection approach is proposed and used to reduce the order of the FOM. Next, two high order models and two low order models were derived for $\mathrm{NO_x}$ and BMEP using the proposed model order reduction approach and the results are compared with a two-layer Artificial Neural Network (ANN) model (one hidden layer and one output layer). Finally, a control-oriented model was developed by adding a first-order lag that represents the system dynamics for BMEP and $\mathrm{NO_x}$. The results show that the high order and low order control-oriented models were both capable of accurately tracking the transient $\mathrm{NO_x}$ emission and engine BMEP at different engine speeds and loads for experimental data that was not used to develop the model.

This paper is organized in sections. The "**Experimental Setup**" section provides information about experimental setup, data collection, and collected data map. The section "**Support Vector Machine**" presents background information about SVM and problem formulating. In the section "**Full-order Model (FOM)**" the full order model for $\mathrm{NO_x}$ and BMEP is developed. The "**Model Order Reduction (MOR)**" section introduces a novel correlation-based model order reduction algorithm and the corresponding results for the MOR are provided for $\mathrm{NO_x}$ and BMEP. A detailed discussion about the different orders of the models and comparisons between the developed models and a conventional ANN model is provided. Section "**BMEP and $\mathrm{NO_x}$ Non-linear Control Oriented Model**" presents a nonlinear control-oriented model along with the dynamic model validation with experimental data. The main conclusions of the paper are detailed in the "**Conclusions**" section.

Experimental Setup

Diesel Engine

A 4-cylinder medium duty diesel engine (Cummins QSB4.5 160 - Tier 3/Stage IIIA) is used in this work. The engine characteristics are listed in Table 1.

Table 1. Diesel engine characteristics[70]

Parameter	value
Engine type	In-Line, 4-Cylinder
Displacement	4.5 L
Bore × Stroke	102 mm × 120 mm
Peak Torque	624 N.m @ 1500 rpm
Peak power	123kW @ 2000 rpm
Aspiration	Turbocharged and Charge Air Cooled
Certification Level	Tier 3 / Stage IIIA

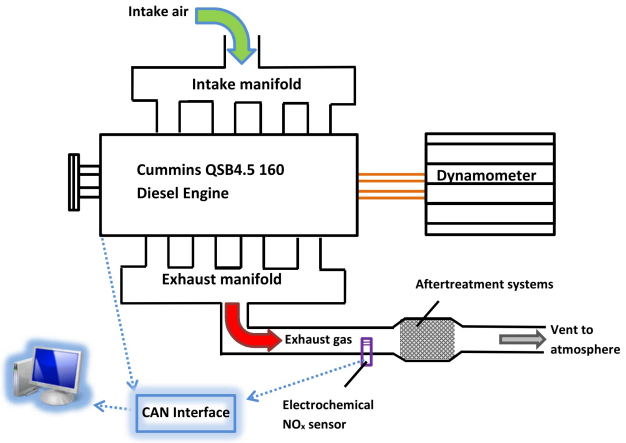


Figure 1. Experiment setup - Internal Combustion Engines

Electrochemical NO_x sensor

A production amperometric NO_x sensor (ECM-06-05) was used in the experiments. All the sensor working parameters were set using the sensor control module (*ECM-NOxCANt* P/N: 02-07). The sensor control module was connected to a computer via a CAN interface (*Kvaser Light HS*) to monitor and log the measurements.

Fourier-Transform Infrared Spectroscopy (FTIR)

A FTIR analyser (*MultiGas 2030*) was used to validate the ECM NO_x sensor measurement and to measure the concentration of other species in the exhaust gas. The FTIR analyser was connected to the diesel engine exhaust pipe, upstream of the catalysts to measure the engine raw emissions. The sample exhaust gas passes through two Heated Filters (*Flexotherm Flex*) connected with Heated Sample Lines (*Flexotherm*) heated to 191^{o}C to avoid water vapor condensation in the sample gas as schematically shown in Fig. 1. The engine operating conditions and corresponding NO_x concentrations are shown is Fig. 2

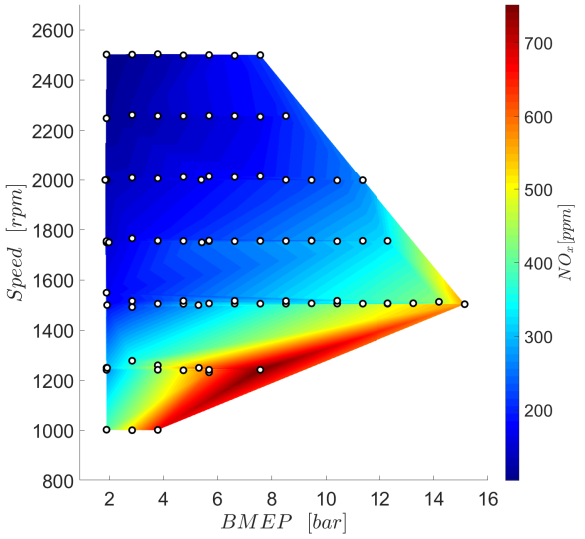


Figure 2. Experimental engine operating points used for model training

Support Vector Machine

Convex Optimization Problem

Support Vector Machine (SVM), introduced by Vapnik[71,72], is a supervised machine learning approach. SVM is typically used for classification of labeled data by creating a set of hyperplanes in an infinite-dimensional space[73]. SVM is also used for regression and function approximation, also called Support Vector Regression (SVR), which was introduced by Vapnik[74]. The main idea of SVM is to find an optimal hyperplane, $\mathbf{y}(\mathbf{u_i})$, to describe a set of labeled training data, $\{\mathbf{u_i}, \mathbf{z_i}\}$, where $\{\mathbf{u_i}\}$ is the feature (input) vector and $\{\mathbf{z_i}\}$ is the target (output) vector of training data. The function $\mathbf{y}(\mathbf{u_i})$ has two main characteristics:

1. $\mathbf{y}(\mathbf{u_i})$ must be as flat as possible,
2. $\mathbf{y}(\mathbf{u_i})$ has at most ϵ deviation for all training data.

In other words, the optimization problem is to find the flattest function for which the acceptable deviation from training data is at most ϵ. The optimal hyperplane to which describes the training data, $\{\mathbf{u_i}, \mathbf{z_i}\}$, can be defined as:

$$\mathbf{y}(\mathbf{u_i}) = \mathbf{w}^T\mathbf{u_i} + \mathbf{b} \tag{1}$$

where $\mathbf{w}$ and $\mathbf{b}$ are found by solving the SVM algorithm for regression problems. Flatness of $\mathbf{y}(\mathbf{u_i})$ in the Eq. (1) is achieved by minimizing the second norm of $\mathbf{w}$. Therefore, the main objective of the SVM algorithm is to find a function which minimizes $||\mathbf{w}||_2^2$ subject to the training error tolerance of ϵ. Then, the optimization problem to find the optimum $\mathbf{y}(\mathbf{u_i})$ is defined as:

$$\textbf{Minimize: } \frac{1}{2}||\mathbf{w}||_2^2$$

$$\textbf{Subject to: } \begin{cases} \mathbf{z_i} - \mathbf{w}^T\mathbf{u_i} - \mathbf{b} \leq \epsilon \\ \mathbf{w}^T\mathbf{u_i} + \mathbf{b} - \mathbf{z_i} \leq -\epsilon \end{cases} \qquad i = 1, ..., n \tag{2}$$

The convex optimization problem, Eq. (2), is feasible when such a $\mathbf{y}(\mathbf{u_i})$ exists which is as flat as possible and

approximates all training data with at most ϵ deviation. In other words, the convex optimization problem is feasible when:

$$-\epsilon \leq \mathbf{z_i} - \mathbf{y_i} \leq \epsilon \tag{3}$$

So, the ϵ-insensitive linear loss function is defined as[74]:

$$L_\epsilon(\mathbf{z_i}, \mathbf{y_i}) = \begin{cases} 0 & |\mathbf{z_i} - \mathbf{y_i}| \leq \epsilon \\ |\mathbf{z_i} - \mathbf{y_i}| - \epsilon & otherwise \end{cases} \tag{4}$$

where the loss function would be zero if training error is less than ϵ. Also, the empirical risk function, R_{emp}, is defined based on the loss function as[75]:

$$R_{emp}(\mathbf{w}, \mathbf{b}) = \frac{1}{N} \sum_{i=1}^{N} L_\epsilon(\mathbf{z_i}, \mathbf{y_i}) \tag{5}$$

where $R_{emp}(\mathbf{w}, \mathbf{b})$ is used in the optimization problem to minimize the defined loss. If this function does not exist, the convex optimization problem is infeasible. In this case, slack variables are added to Eq. (3) to overcome the above optimization problem infeasibility as:

$$-\epsilon - \zeta_i^- \leq \mathbf{z_i} - \mathbf{y_i} \leq \epsilon + \zeta_i^+ \tag{6}$$

where the slack variables are introduced as penalty variables to overcome this infeasibility of the convex optimization problem. The empirical risk function can then be rewritten based on the slack variables Using Eq. (6) as:

$$R_{emp}(\mathbf{w}, \mathbf{b}) = \frac{1}{N} \sum_{i=1}^{N} \left(\zeta_i^- + \zeta_i^+\right) \tag{7}$$

Then, the convex optimization problem is modified by adding the minimizing empirical risk function term to Eq. (2)

$$\begin{aligned} &\textbf{Minimize: } \frac{1}{2}||\mathbf{w}||_2^2 + C \sum_{i=1}^{n} (\zeta_\mathbf{i}^+ + \zeta_\mathbf{i}^-) \\ &\textbf{Subject to: } \begin{cases} \mathbf{z_i} - \mathbf{w}^T\mathbf{u_i} - \mathbf{b} \leq \epsilon + \zeta_\mathbf{i}^+ \\ \mathbf{w}^T\mathbf{u_i} + \mathbf{b} - \mathbf{z_i} \leq \epsilon + \zeta_\mathbf{i}^- \\ \zeta_\mathbf{i}^-, \zeta_\mathbf{i}^+ \geq 0 \end{cases} \end{aligned} \tag{8}$$

where C is a positive regulatory parameter defined as a trade-off factor between flatness of the model and minimizing the training error tolerance. A model with tolerated error and slack variables for a single feature-single target system is schematically shown in Fig. 3. The ϵ-insensitive linear loss function is schematically shown in Fig. 4.

Dual Optimization Problem and computing $\mathbf{w}$

To consider constraints of the convex optimization problem in Eq. (8), the Lagrangian function is calculated to change the convex optimization problem to a dual optimization problem (also called primal problem) as[62]:

$$\begin{aligned} L =& \frac{1}{2}||\mathbf{w}||_2^2 + C \sum_{i=1}^{N} \left(\zeta_i^- + \zeta_i^+\right) \\ &- \sum_{i=1}^{N} \alpha_i^+ (-\mathbf{z_i} + \mathbf{y_i} + \epsilon + \zeta_i^+) - \sum_{i=1}^{N} \mu_i^+ \zeta_i^+ \\ &- \sum_{i=1}^{N} \alpha_i^- (\mathbf{z_i} - \mathbf{y_i} + \epsilon + \zeta_i^-) - \sum_{i=1}^{N} \mu_i^- \zeta_i^- \end{aligned} \tag{9}$$

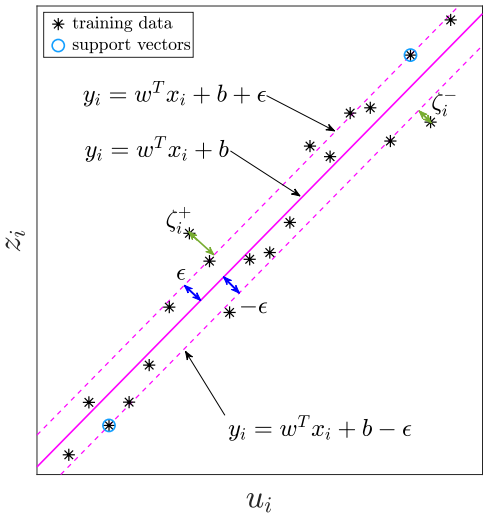


Figure 3. SVM regression and support vectors example

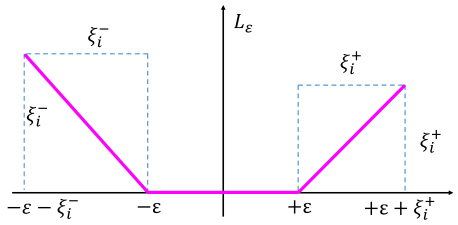


Figure 4. $\epsilon-$sensitive Loss function with slack variable based on[75]

where α_i^+, α_i^-, μ_i^+, and μ_i^- are Lagrangian Multipliers and $\alpha_i^+, \alpha_i^-, \mu_i^+, \mu_i^- \geq 0$. Based on the Saddle points condition, the partial differential of the Lagrangian function with respect to the optimization variables $(\mathbf{w}, \mathbf{b}, \zeta_i^+, and \zeta_i^-)$ must be equal to zero as[62]:

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \rightarrow w = \sum_{i=1}^{N} (\alpha_i^+ - \alpha_i^-)\mathbf{u_i} \tag{10a}$$

$$\frac{\partial L}{\partial \mathbf{b}} = 0 \rightarrow \sum_{i=1}^{N} (\alpha_i^+ - \alpha_i^-) = 0 \tag{10b}$$

$$\frac{\partial L}{\partial \zeta_i^+} = 0 \rightarrow \alpha_i^+ + \mu_i^+ = C \tag{10c}$$

$$\frac{\partial L}{\partial \zeta_i^-} = 0 \rightarrow \alpha_i^- + \mu_i^- = C \tag{10d}$$

where Eq. (10a) is the support vector expansion, Eq. (10b) is the bias constraints, Eq. (10c), and Eq. (10d) are the box constraint. Based on the support vector expansion, Eq. (1), the prediction function (model) can be rewritten using Eq. (10a) as

$$\mathbf{y(u)} = \sum_{i=1}^{N} (\alpha_i^+ - \alpha_i^-)\mathbf{u_i}\mathbf{u} + \mathbf{b} \tag{11}$$

The dual optimization problem is obtained by substituting Eqs. (10a)-(10d) into Eq. (9) as

$$\begin{aligned}\textbf{Minimize: } L =& \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-)\mathbf{u_i}^T\mathbf{u_j} \\ &- \sum_{i=1}^{N}(\alpha_i^+ - \alpha_i^-)\mathbf{z_i} + \epsilon\sum_{i=1}^{N}(\alpha_i^+ + \alpha_i^-) \\ \textbf{Subject to: } & \begin{cases} \sum_{i=1}^{N}(\alpha_i^+ - \alpha_i^-) = 0 \\ 0 \le \alpha_i^+ \le C \\ 0 \le \alpha_i^- \le C \end{cases}\end{aligned} \tag{12}$$

Eq. (12) can be rewritten in a standard Quadratic Programming form (QP)[76]:

$$\begin{aligned}\textbf{Minimize: } & \frac{1}{2}\alpha^T\mathcal{H}\alpha + f^T\alpha \\ \textbf{Subject to: } & A_{eq}\alpha = B_{eq}\end{aligned} \tag{13}$$

where

$$\alpha = \begin{bmatrix} \alpha^+ \\ \alpha^- \end{bmatrix},\ \mathcal{H} = \begin{bmatrix} H & -H \\ -H & H \end{bmatrix},\ \mathbf{f} = \begin{bmatrix} -\mathbf{z_i} + \epsilon \\ \mathbf{z_i} + \epsilon \end{bmatrix},$$
$$H = \begin{bmatrix}\mathbf{u_i}^T\ \mathbf{u_j}\end{bmatrix},\quad \mathbf{A_{eq}} = [1...1 \quad -1...-1],\quad \mathbf{B_{eq}} = [0] \tag{14}$$

where $\mathbf{w}$ can be calculated by finding α (Solving Eq. (14)) and substituting it into Eq. (10a). This fact shows that matrix $\mathbf{w}$ is calculated based on the linear combination of α and the training data.

Karush-Kuhn-Tucker (KKT) conditions and computing $\mathbf{b}$

Based on KKT approach, the following equations must be fulfilled at the optimum point [77,78]:

$$\alpha_i^+(-\mathbf{z_i} + \mathbf{y_i} + \epsilon + \zeta_i^+) = 0 \tag{15a}$$
$$\alpha_i^-(\mathbf{z_i} - \mathbf{y_i} + \epsilon + \zeta_i^-) = 0 \tag{15b}$$
$$\mu_i^+\zeta_i^+ = (C - \alpha_i^+)\zeta_i^+ \tag{15c}$$
$$\mu_i^-\zeta_i^- = (C - \alpha_i^-)\zeta_i^- \tag{15d}$$

Considering Eq. (15), only the following five cases are possible:

$$\alpha_i^+ = \alpha_i^- = 0 \tag{16a}$$
$$0 < \alpha_i^+ < C,\ \alpha_i^- = 0 \tag{16b}$$
$$0 < \alpha_i^- < C,\ \alpha_i^+ = 0 \tag{16c}$$
$$\alpha_i^+ = C,\ \alpha_i^- = 0 \tag{16d}$$
$$\alpha_i^- = C,\ \alpha_i^+ = 0 \tag{16e}$$

For $|\mathbf{z_i} - \mathbf{y_i}|$ to be exactly equal to ϵ, the only Eqs. (16b) and (16c) are necessary. So, the points of the training data which have $|\mathbf{z_i} - \mathbf{y_i}| = \epsilon$ are called support vectors (circled data points in the Fig. 3). Hence, the support vectors domain, S, is calculated as:

$$S = \{\ i \mid 0 < \alpha_i^- + \alpha_i^+ < C\} \tag{17}$$

where S is the index of the training data which form the SVM training algorithm support vectors. Accordingly, for the set of support vectors, $\mathbf{z_i}$ equals to $\mathbf{y_i} + sign(\alpha_i^+ - \alpha_i^-)\epsilon \quad (i \in S)$. As a result, $\mathbf{b}$ is calculated as:

$$\mathbf{b} = \frac{1}{|S|}\sum_{i\in S}^{S}(\mathbf{z_i} - \mathbf{w}^T\mathbf{u_i} - sign(\alpha_i^+ - \alpha_i^-)\epsilon) \tag{18}$$

In summary, the convex problem (Eq. (8)) is changed to the dual problem (Eq. (12)). Then, by solving the quadratic programming, Eq. (13), and substituting it into the support vector expansion, Eq. (10a), $\mathbf{w}$ is calculated. Then, vector $\mathbf{b}$ is calculated using Eq. (18) (KKT conditions). Finally, by substituting $\mathbf{w}$ and $\mathbf{b}$ into Eq. (1), the prediction model of a given data set ($\{\mathbf{u_i}, \mathbf{z_i}\}$) is found as:

$$\begin{aligned}\mathbf{y(u)} = & \sum_{i=1}^{N}(\alpha_i^+ - \alpha_i^-)\mathbf{u_i}\mathbf{u} \\ & + \frac{1}{|S|}\sum_{i\in S}^{S}(\mathbf{z_i} - \mathbf{w}^T\mathbf{u_i} - sign(\alpha_i^+ - \alpha_i^-)\epsilon)\end{aligned} \tag{19}$$

In this study, $\mathbf{y(u)}$ is used to predict steady-state diesel engine NO_x emission and Brake Mean Effective Pressure (BMEP). This function is used to predict steady-state behaviour of engine and will be denoted as $\mathbf{y_{ss}(u)}$ in the subsequent sections.

Full-order Model (FOM)

The diesel engine model consists three inputs and two outputs. The model inputs are injected fuel amount m_f, engine speed n, and fuel rail pressure P_r. The model outputs are engine-out NO_x emission and BMEP. To provide the maximum model flexibility and to minimize the model bias, the interactions of the primary features should also be considered. The number of resulting features depends on the highest order of interactions considered for the model. The number of total features is calculated based on the r-combination with repetitions formula $\left(\frac{(n+r-1)!}{r!\ (n-1)!}\right)$, where n is the number of original features (in our case $n = 3$), and r is the order of interactions [79]. So, the total number of features in a model of order r is equal to the sum of all the features with orders from 1 to r. The number of features for each interaction order are listed in Table 2.

Table 2. Number of features in each order from 1 to 6 using r-combination with repetitions formula

Order(r)	**r-combination with repetitions**	**features number up to order r**
1	$\frac{(3+1-1)!}{1!\ 2!} = 3$	3
2	$\frac{(3+2-1)!}{2!\ 2!} = 6$	$3 + 6 = 9$
3	$\frac{(3+3-1)!}{3!\ 2!} = 10$	$10 + 9 = 19$
4	$\frac{(3+4-1)!}{4!\ 2!} = 15$	$15 + 19 = 34$
5	$\frac{(3+5-1)!}{5!\ 2!} = 21$	$34 + 21 = 55$
6	$\frac{(3+6-1)!}{6!\ 2!} = 28$	$55 + 28 = 83$

The total number of experimental points used for training is 62. To simultaneously minimize the model bias and to avoid over-fitting, orders 1 to 4 of the original inputs and their interactions are considered as the base FOM model (34 features) to predict the steady-state values of NO_x and BMEP. The FOM features are listed in Table 3. The feature

Table 3. Features of the Full-Order Model (FOM) of NO_x and BMEP

$U_1 = m_f$	$U_2 = n$	$U_3 = P_r$
$U_4 = m_f^2$	$U_5 = n^2$	$U_6 = P_r^2$
$U_7 = m_f n$	$U_8 = m_f P_r$	$U_9 = nP_r$
$U_{10} = m_f^3$	$U_{11} = n^3$	$U_{12} = P_r^3$
$U_{13} = m_f^2 n$	$U_{14} = m_f^2 P_r$	$U_{15} = (n^2)P_r$
$U_{16} = n^2 m_f$	$U_{17} = P_r^2 m_f$	$U_{18} = P_r^2 n$
$U_{19} = m_f n P_r$	$U_{20} = m_f^4$	$U_{21} = n^4$
$U_{22} = P_r^4$	$U_{23} = m_f^3 n$	$U_{24} = m_f^3 P_r$
$U_{25} = n^3 P_r$	$U_{26} = n^3 m_f$	$U_{27} = P_r^3 m_f$
$U_{28} = P_r^3 n$	$U_{29} = (m_f n)^2$	$U_{30} = (m_f P_r)^2$
$U_{31} = (nP_r)^2$	$U_{32} = P_r^2 n m_f$	$U_{33} = n^2 m_f P_r$
$U_{34} = m_f^2 P_r n$		

vector, $\mathbf{U_j}$, is defined using Table 3 as

$$\mathbf{U_j} = \{\mathbf{u_i}\}_\mathbf{j} \quad i = 1, ..., n, \quad j = 1, ..., 34 \tag{20}$$

where n is the number of data points and j is the index number of the features. As the dimensions and the range of features are quite different, all of the features must be normalized to improve the training performance[80]. Particularly, for SVMs, the training time can be significantly reduced by normalizing the features[81]. Here the rescaling or also called min-max normalization method is used to normalized feature for the SVM:

$$\bar{\mathbf{U}} = \frac{\mathbf{U} - min(\mathbf{U})}{max(\mathbf{U} - min(\mathbf{U})} \tag{21}$$

The system outputs vector is defined as

$$\mathbf{Z} = \{\mathbf{z_i}\} = [\mathbf{NO_{x,i}} \ \mathbf{BMEP_i}]^T \quad i = 1, ..., n \tag{22}$$

Then, the predicted steady-state NO_x and BMEP are:

$$\mathbf{y_{ss}} = [\mathbf{NO_{x,ss}} \ \mathbf{BMEP_{ss}}]^T \tag{23}$$

By solving the SVM algorithm for a given training data set, $\{\bar{\mathbf{U}}_j, \mathbf{Z}\}$, where $\bar{\mathbf{U}}_j$ and $\mathbf{Z}$ are calculated from Eq. (20) and Eq. (22), respectively, the approximate function, $\mathbf{y_{ss}}$ is obtained to predict the steady-state values of NO_x and BMEP. To cover a wide range of engine operating conditions, the diesel engine was run at 84 operating points, 62 data points (74 %) are used as the training data, and 22 data points (26 %) are used to test the SVM learning algorithm. To find hyperparameters of SVM (C), 15% of the training data set (9 points of 62 training points) are selected randomly and used for cross-validation. To find the best regulatory parameter C of the FOM for both NO_x and BMEP, the effect of varying C on the squared correlation coefficient (R^2), maximum error between prediction and actual data (E_{max}) and cost function ($J(E_{max}, R^2)$), for both training data and test data are analysed. The proposed cost function to find C is defined as

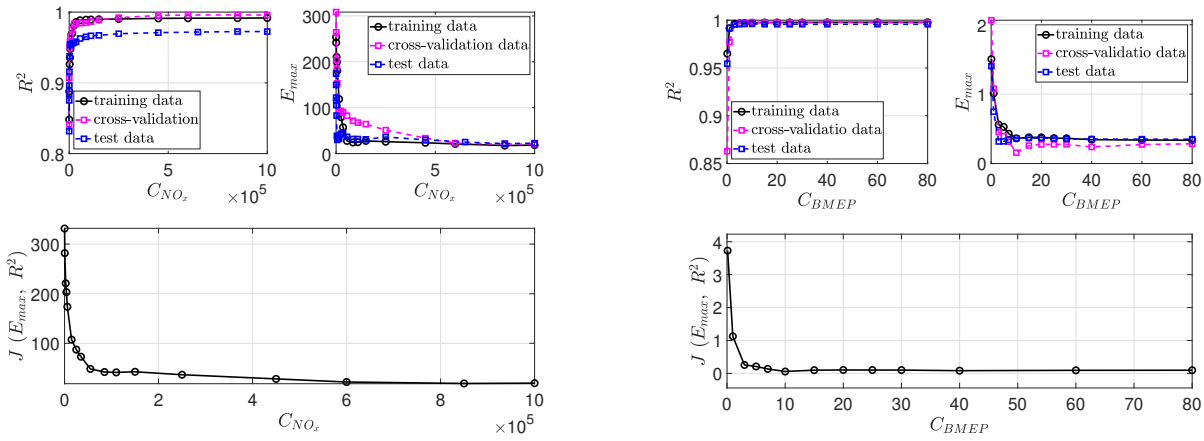
$$J(E_{max}, R^2) = \sqrt{\frac{E_{max,tr} \ E_{max,ts}}{R_{tr}^2 \ R_{ts}^2}} \tag{24}$$

where $E_{max,tr}$ and $E_{max,ts}$ are the maximum errors between the prediction and the actual data for training and test data set respectively. Also, R_{tr}^2 and R_{ts}^2 are squared correlation coefficient for training and test data set, respectively. The goal is to increase C and to minimize the maximum error and to maximize the squared correlation coefficients for both the training and test data. Therefore, the best C for modeling is obtained by minimizing $J(E_{max}, R^2)$. In this section, cross-validation data are used in Eq. 24 to find the regulatory parameter C. The squared correlation coefficient R^2, maximum error between prediction and actual data E_{max} and cost function, $J(E_{max}, R^2)$ with respect to the regulatory parameter C for training, cross-validation, and test data set of the FOM NO_x and BMEP model are shown in Fig. 5. The regulatory parameter, C, is a trade-off between the model flatness and the tolerated error. Based on the results shown in Fig. 5, the prediction error increases by decreasing C.

The squared correlation coefficient (R^2) is used to quantify the model accuracy. The maximum error between the prediction and the actual data for both of the training and the cross-validation data decrease as the regulatory parameter C increases resulting in a decrease in $J(E_{max}, R^2)$. After C reaches a certain value of C_o, the model performance enhancement levels off since the squared correlation coefficient and the maximum error for all data are saturated. By increasing C to more than C_o, the model performance remains unchanged, but the model flatness decreases, i.e., the over-fitting probability has increased. As a result, the model is less robust for new test data due to possible over-fitting. Therefore, by setting $C = C_o$, the model performance is maximized while over-fitting constraints are fulfilled. To ensure all the important features are considered when minimizing the slack variables in the optimization problem, a sufficiently large value of regulatory parameter C must be selected. Based on Fig. 5, C_o for NO_x and BMEP are selected to be $C_{o,NO_x} = 85000$ and $C_{o,bmep} = 60$, respectively. The prediction versus the actual value for FOM NO_x and BMEP are shown in Fig. 6. Here, the cross-validation portion of training data are shown for both NO_x and BMEP; however, to reduce the complexity of figures, for the rest of the paper, combined cross-validation and training data is illustrated as training data. It should be noted that the regulatory parameter remains constant throughout the MOR process.

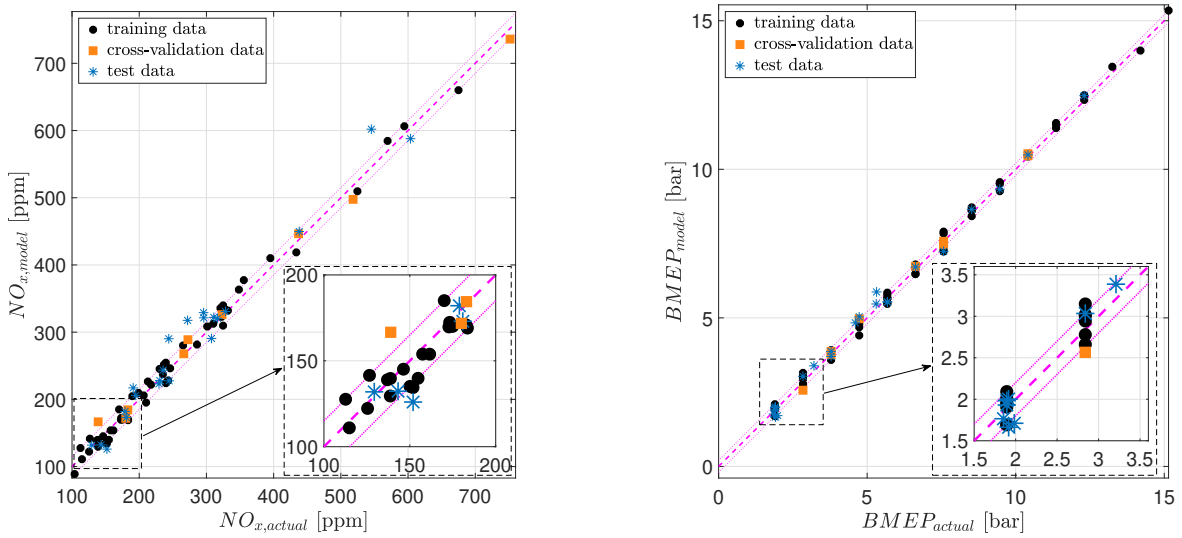
Model Order Reduction (MOR)

Next, using the proposed FOM and by solving the SVM algorithm for regression, matrices $\mathbf{w}$ and $\mathbf{b}$ are obtained. The best C value for the SVM problem, C_o, is found using a different criteria. In this section, the Model Order Reduction (MOR) algorithm is proposed to reduce the order of NO_x and BMEP steady-state FOM. MOR helps to achieve an appropriate model by removing redundant features and selecting the important ones. The Nonlinear Reduced Control Oriented Model (NRCOM) is found through flow chart in Fig. 7. For a given data set of (m_f, n, P_r) as the inputs and $(BMEP, NO_x)$ as the targets, and starting from FOM with 34 features, first, $\mathbf{w}$ and $\mathbf{b}$ are calculated. Then, the value of $\mathbf{w}$ is evaluated for each feature. Then, the feature for which the $\mathbf{w}$ array has the minimum value, is removed. Then, the SVM algorithm for regression is solved for a new set of features. As a result of MOR



(a) NO_x **(b)** BMEP

Figure 5. Maximum error (E_{max}), correlation coefficient (R^2), and cost function ($J(E_{max}, R^2)$) vs regulatory parameter C



(a) Prediction vs Actual for NO_x FOM **(b)** Prediction vs Actual for BMEP FOM

Figure 6. Prediction vs actual data for NO_x and BMEP Full Order Model (FOM)

algorithm for NO_x and BMEP, two types of models are proposed as:

1. High-Order Model (HOM): For the HOM, the model accuracy is the priority rather than the number of features and the computation time. Therefore, only the unnecessary features of FOM are removed. The HOM model can be used in applications that require high accuracy such as developing a NO_x sensor fault-detection algorithm or virtual plants to evaluate a controller in simulation.
2. Low-Order Mode (LOM): For the LOM, the number of features and the computation time are as important as the model accuracy. The objective is to find a simple model with fewer features and an acceptable accuracy. As LOM has a simple structure and acceptable accuracy, it is useful for designing a controller[82].

As shown in Fig. 7, the features of HOM, m_{HOM}, are selected in a way that $J_m(R^2, E_{max})$ is minimized since the main objective of model order reduction for HOM is maximizing the model accuracy by removing the redundant features, with no concern for reducing the size of the model. However, reducing the model size while keeping the accuracy acceptable, was the objective for model order reduction to the LOM. To avoid significant loss of the model accuracy, the least significant features are removed one by one until the relative deference between the cost functions J_m and J_{m-1} becomes more than the acceptable threshold. Then the corresponding feature number is the number of features of the LOM, m_{LOM}. The relative deference between J_m and J_{m-1} is defined as:

$$d_r(J_m, J_{m-1}) = \frac{|J_m - J_{m-1}|}{max(J_m, J_{m-1})} \tag{25}$$

In this study, the LOM is found by defining 25% threshold. In other words, starting MOR from initial features, as soon as relative deference between J_m and J_{m-1} is more than 25%, the corresponding m is considered as LOM features set, m_{LOM}.

NO_x *steady State Model*

The squared correlations coefficient (R^2) and the maximum error between prediction and actual data (E_{max}) for both the training and the test data and defined cost

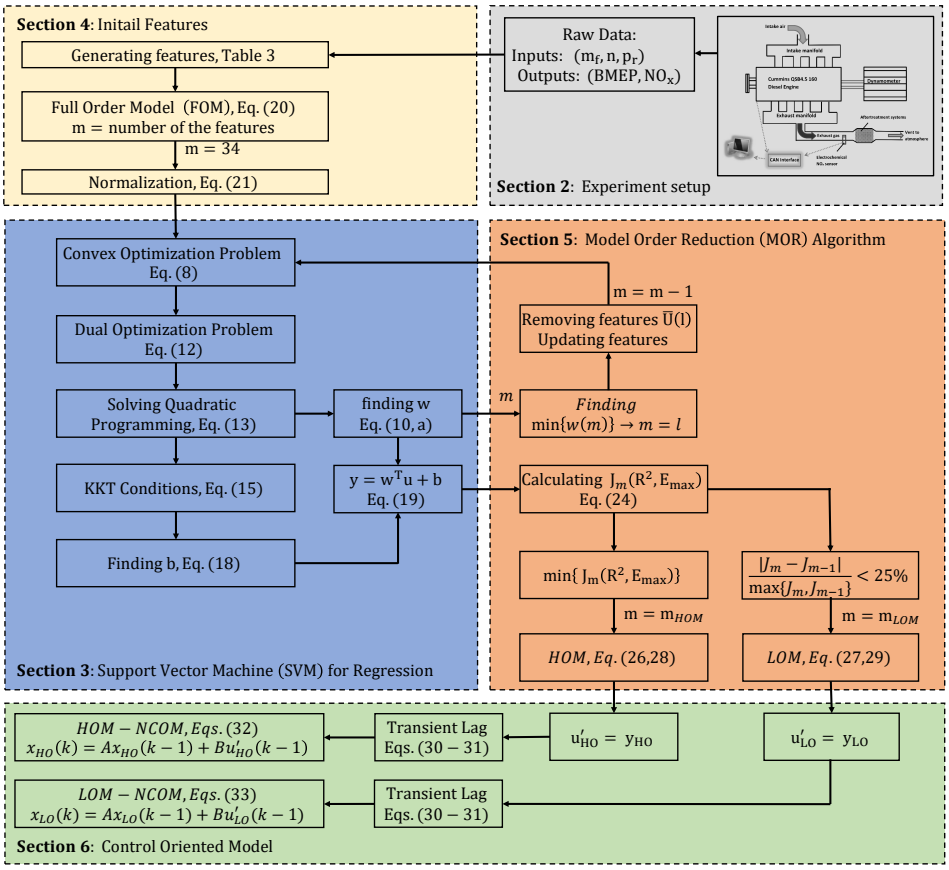


Figure 7. Control Oriented Model (COM) development and SVM-based MOR algorithm

function ($J(R^2, E_{max})$) with respect to the number of features are shown in Fig. 8.

Based on Fig. 8, since $min\{J(R^2, E_{max})\}$ is achieved for a 29-feature model ($m_{HOM,NO_x} = 29$). These models with these 29 features is chosen as the HOM NO_x. In other words, the 29-feature model is chosen as the HOM because it has the highest accuracy among all the models studied. Tracking $J_m - J_{m-1}$ as a function of m in Fig. 8 by starting from $m = 34$, the first relative difference larger than 25% occurs for a 9-feature model ($m_{LOM,NO_x} = 9$), the models with 9 features is chosen as the LOM for NO_x, i.e., the model with 9 features is chosen as LOM because by decreasing the model features to less than 9, a significant reduction in model performance (J_m) is occurred. As ANN is widely used for the engine performance and emission modeling, the SVM model for all of the developed models (FOM, HOM, and LOM) are compared with an ANN using the same set of features. This provide a standard to compare these results to a ANN. In this study, a two-layer (one hidden layer and one output layer) feed-forward backpropagation network with three neurons in the hidden layer is employed, and the Levenberg-Marquardt training method is used to train the model which has a relatively fast convergence [39].The selection of hidden layer and neurons number was based on similar ANN-based studies in the literature. To make sure that the number of neurons are compatible with the size of data set, three neurons are considered for the hidden layer as proposed by a similar study with a similar data size [22,39]. The same training, cross validation, and test data set are used for the SVM and the ANN models. Both algorithms use 15% of the training data set to find the model hyperparameters.

The maximum error between the prediction and the actual training data set, squared correlation coefficient, the defined cost function, and training time for both SVM and ANN training methods are listed in Table 4. The results reveal that, the SVM model has a shorter training time and a more accurate model (larger squared correlation coefficient and smaller maximum error between actual and model), especially for the test data. This is partly due to the fact that ANN uses gradient descent algorithm for training which increases the risk of converging to local minima. Additionally, the risk of overfitting is higher for ANN for the same size of training data [83]. This problem is also shown in the results where the squared correlation coefficient of test data for ANN is less than the SVM model. Since the training time of SVM is significantly less than ANN, it is more suitable for real-time applications. Another benefit of using this SVM is that the model is far simpler to explain mathematically in the form of an equation, especially when a linear kernel is used. When using the linear kernel, the SVM model is defined based on the vector $\mathbf{w}$ with a bias $\mathbf{b}$.

It should be noted that the performance of HOM is even better than FOM as a result of removing unnecessary features that affect the flatness of the SVM algorithm. Based on Table 4, the accuracy of LOM is acceptable where the error is above the defined threshold.

Thus, the HOM and LOM features are (see Table 3)

$$\bar{\mathbf{U}}_{\mathbf{NO_x,HO}} = \bar{\mathbf{U}}_{\mathbf{j}} \qquad j = 1-9, 11-15, 17, 19-27, 29, 31-34 \tag{26}$$

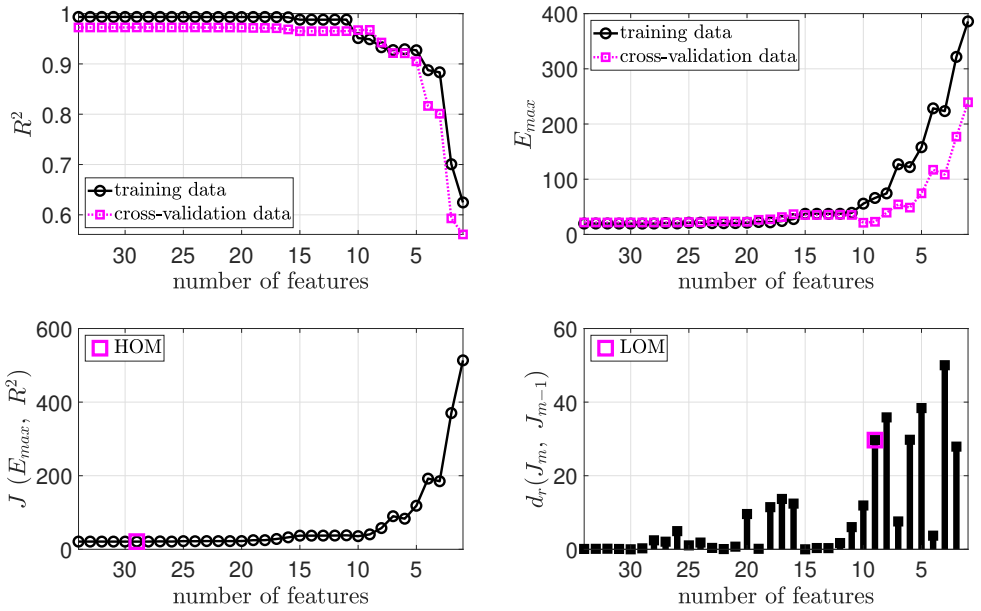


Figure 8. Maximum error (R^2), squared correlation coefficient (R^2), and cost function ($J(E_{max}, R^2)$) vs number of features of prediction function for steady-state NO_x prediction

Table 4. Performance of the NO_x Full-Order Model (FOM), High-Order Model (HOM), and Low-Order Model (LOM)

Model Type	**FOM**		**HOM**		**LOM**	
No. of Features	34		29		9	
Training Method	**SVM**	**ANN**	**SVM**	**ANN**	**SVM**	**ANN**
$E_{max,tr}$ $[ppm]$	19.6888	25.6473	19.5689	27.3405	66.02	57.9259
$E_{max,ts}$ $[ppm]$	21.660	60.7375	21.6665	47.7841	22.91	60.2836
R^2_{tr}	0.9934	0.9969	0.9934	0.9837	0.9490	0.9891
R^2_{ts}	0.9725	0.9775	0.9725	09664	0.9677	0.9760
$J(E_{max}, R^2)$ $[ppm]$	21.0106	39.9824	20.9490	37.0706	40.58	54.67
Training Time $[ms]$	9.47	240.6	11.07	202.0	13.10	194.5

$$\bar{\mathbf{U}}_{\mathbf{NO_x,LO}} = \bar{\mathbf{U}}_{\mathbf{j}} \quad j = 2, 5, 8, 15, 21, 22, 27, 32, 33 \tag{27}$$

where i is the data index and j is the feature index. By solving the SVM algorithm for NO_x, the features of HOM and LOM are obtained. The equation of $\mathbf{NO_{x,HO,ss}}$ and $\mathbf{NO_{x,LO,ss}}$ are listed in Appendix (), Eqs. (40) and (41). The predicted steady-state NO_x vs the actual value for both the high-order and the low-order steady-state NO_x model in shown in Fig. 9. Based on Fig. 9-(a), most of the test and the training data are within the defined tolerance ϵ for HOM of NO_x. However, as shown in Fig. 9-(b), the accuracy of LOM is not consistent throughout all data points for both training and test points and item number of outliers are greater than the HOM.

BMEP *steady state Model*

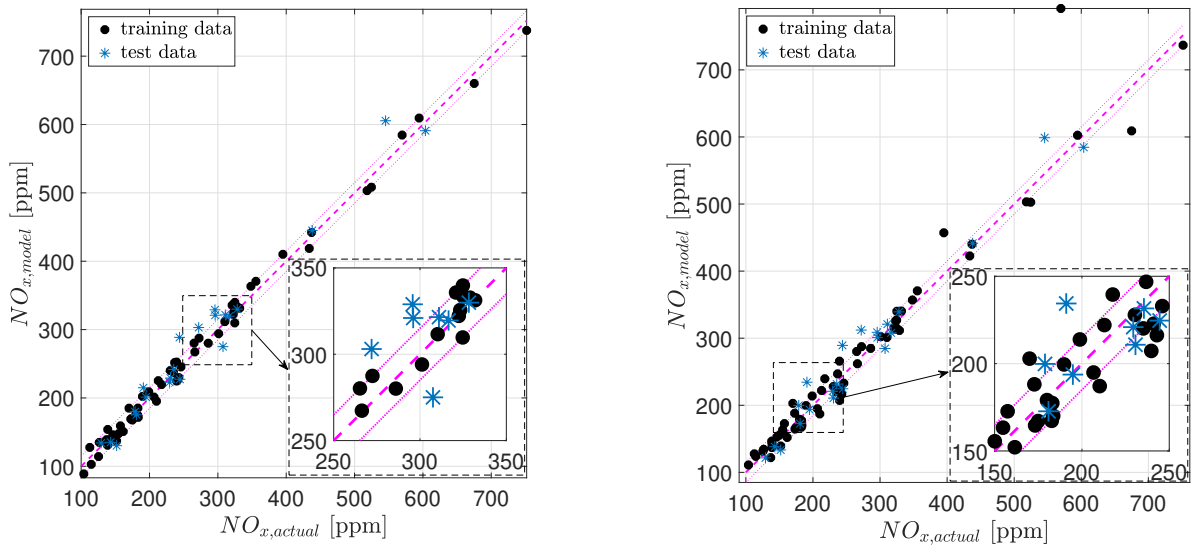
Similar to the NO_x steady-state model, the BMEP reduced steady-state model is obtained. The squared correlations coefficient (R^2) and the maximum error between prediction and actual data (E_{max}) for both the training and the test data and defined cost function ($J(R^2, E_{max})$) with respect to the number of features are shown in Fig. 10. Based on Fig. 10, a 20-feature model ($m_{HOM,BMEP} = 20$) and a 6-feature model ($m_{LOM,BMEP} = 6$) are chosen as the HOM and LOM of BMEP, respectively. The maximum error between the prediction and the actual data (E_{max}), the squared correlation coefficient (R^2), cost function ($J(E_{max}, R^2)$), and training time for ANN and SVM training methods are listed in Table 5. Similar to the NO_x model, for all the BMEP models, the SVM has faster training and more accurate response compared to the ANN especially for test data. The general performance of HOM is acceptable with respect to the FOM, while it has a simpler structure. Also, the 6-feature model is chosen as LOM. As shown in Table 5, the accuracy of the model is acceptable, and by reducing the model further, the model becomes inaccurate.

Thus, the HOM and LOM features are obtained as:

$$\bar{\mathbf{U}}_{\mathbf{BMEP,HO}} = \bar{\mathbf{U}}_{\mathbf{j}} \quad j = 1, 4, 7-10, 17-21, 24-27, 29-32, 34 \tag{28}$$

$$\bar{\mathbf{U}}_{\mathbf{BMEP,LO}} = \bar{\mathbf{U}}_{\mathbf{j}} \quad j = 1, 18, 21, 25, 27, 30 \tag{29}$$

where i is the is data index, and j is the feature index. By solving the SVM algorithm for BMEP, the HOM and the LOM are achieved. The equation of $\mathbf{BMEP_{HO,ss}}$ and $\mathbf{BMEP_{LO,ss}}$ are listed in Appendix () Eqs. (42) and (43). The predicted steady-state BMEP with respect to the actual value for both of the high-order and the low-order steady-state BMEP models are shown in Fig. 11. As shown in Table 5, the HOM and LOM have an acceptable accuracy



(a) HOM of $\mathrm{NO_x}$ (with 29 features)

(b) LOM of $\mathrm{NO_x}$ (with 9 features)

Figure 9. Prediction vs actual data for the Low-order Model (LOM) and the High-order Model (HOM) of $\mathrm{NO_x}$

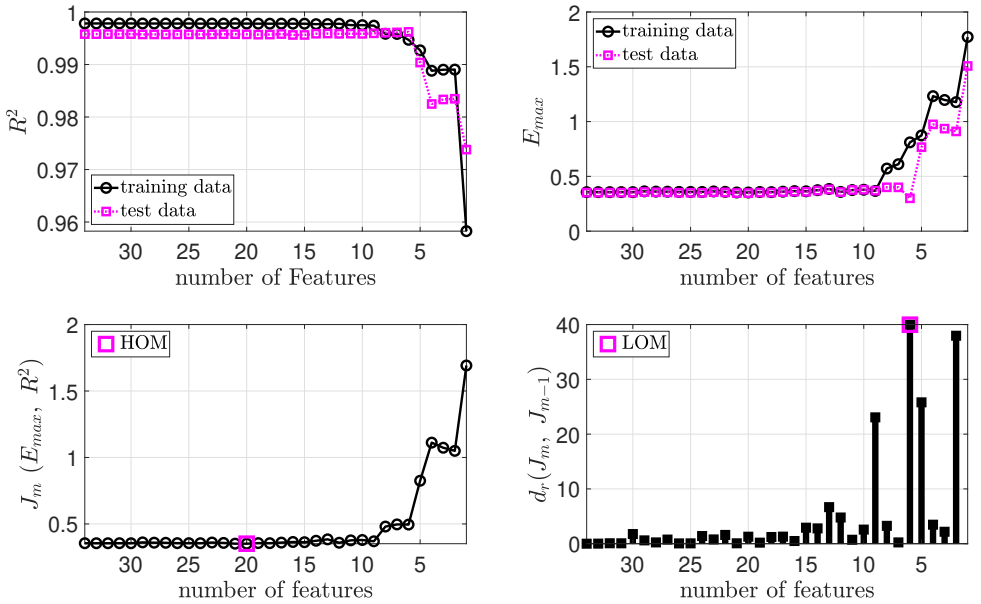


Figure 10. Maximum error (R^2), squared correlation coefficient (R^2), and cost function ($J(E_{max}, R^2)$) vs number of features of prediction function for steady-state BMEP prediction

Table 5. Performance of the BMEP Full-Order Model (FOM), High-Order Model (HOM), and Low-Order Model (LOM)

Model Type	**FOM**		**HOM**		**LOM**	
No. of Features	34		20		6	
Training Method	**SVM**	**ANN**	**SVM**	**ANN**	**SVM**	**ANN**
$E_{max,tr}$ $[ppm]$	0.3560	0.4006	0.3526	0.3859	0.810	0.5435
$E_{max,ts}$ $[ppm]$	0.3513	0.4484	0.3477	0.4151	0.2998	0.4732
R^2_{tr}	0.9978	0.9987	0.9978	0.9953	0.9947	0.9961
R^2_{ts}	0.9957	0.9959	0.9957	0.9961	0.9962	0.996
$J(E_{max}, R^2)$ $[ppm]$	0.3548	0.4250	0.3513	0.4020	0.4952	0.5091
Training Time $[ms]$	35.9	199.8	9.2	218.0	9.5	214.7

while the HOM has a higher accuracy than LOM. However, the LOM of BMEP has only 6 features, which makes it a simple model that requires a low computational effort. Most of the test and the training data for both HOM and LOM of BMEP are within the defined tolerance ϵ, as shown in Fig. 11. This means that the MOR improves in the accuracy of the FOM by removing its unnecessary features.

One important observation from the training time (Table 4 and Table 5) is that by increasing number of features in ANN, the training time is increased. However, this trend inverses in the SVM such that by decreasing the number of

the features, the training time increases. This behavior results in reducing the overall training time due to the time that is saved by reducing the number of training iterations[84]. This trend appears in the HOM and LOM models.

BMEP and NO_x Non-linear Control Oriented Model

The model described above is used to determine steady state NO_x and BMEP. Now a simple 1^{st} order dynamic model for transient operation will be defined. To derive the discrete-time dynamic control oriented model, NO_x concentration at step k for a sampling interval of T, is calculated as follows:

$$NO_x(k) = (1 - \frac{T}{\tau_{NOx} + T})NO_x(k-1) + \frac{T}{\tau_{NOx} + T}NO_{x,ss}(k-1) \quad (30)$$

and the BMEP at step k is calculated using the following equation:

$$BMEP(k) = (1 - \frac{T}{\tau_{BMEP} + T})BMEP(k-1) + \frac{T}{\tau_{BMEP} + T}BMEP_{ss}(k-1) \quad (31)$$

where $NO_{x,ss}(k-1)$ and $BMEP_{ss}(k-1)$ are the steady state NO_x and BMEP calculated using Eqs. (40) and (43). Hence τ is the sample interval and k is the sample time while τ_{NOx} and τ_{BMEP} are the time constant parameters for NO_x and BMEP respectively, which are estimated based on the experimental data and are found to be 1 and 0.2 seconds, for NO_x and BMEP respectively[70]. The state space of the control oriented model for both high-order and low-order models can be defined as:

$$\mathbf{x_{HO}}(k) = \mathbf{A}\mathbf{x_{HO}}(k-1) + \mathbf{B}\hat{\mathbf{u}}_{\mathbf{HO}}(\mathbf{k-1}) \quad (32)$$

$$\mathbf{x_{LO}}(k) = \mathbf{A}\mathbf{x_{LO}}(k-1) + \mathbf{B}\hat{\mathbf{u}}_{\mathbf{LO}}(\mathbf{k-1}) \quad (33)$$

where vector **x**(K) contains two model states:

$$\mathbf{x_{HO}}(\mathbf{k}) = \begin{bmatrix} NO_{x,HO}(k) & BMEP_{HO}(k) \end{bmatrix}^T \quad (34)$$

$$\mathbf{x_{LO}}(\mathbf{k}) = \begin{bmatrix} NO_{x,LO}(k) & BMEP_{LO}(k) \end{bmatrix}^T \quad (35)$$

and vector $\hat{\mathbf{u}}(\mathbf{k})$ is calculated as

$$\hat{\mathbf{u}}_{\mathbf{HO}}(\mathbf{k}) = \begin{bmatrix} \mathbf{NO_{x,HO,ss}} \\ \mathbf{BMEP_{HO,ss}} \end{bmatrix} \quad (36)$$

$$\hat{\mathbf{u}}_{\mathbf{LO}}(\mathbf{k}) = \begin{bmatrix} \mathbf{NO_{x,LO,ss}} \\ \mathbf{BMEP_{LO,ss}} \end{bmatrix} \quad (37)$$

where $\mathbf{NO_{x,HO,ss}}$, $\mathbf{BMEP_{HO,ss}}$, $\mathbf{NO_{x,LO,ss}}$, and $\mathbf{BMEP_{LO,ss}}$ are listed in Appendices A and B. The vector **y** contains two model outputs:

$$\mathbf{y}(\mathbf{k}) = \begin{bmatrix} x_1(k) & x_2(k) \end{bmatrix} \quad (38)$$

Matrices $\mathbf{A}$ and $\mathbf{B}$ are:

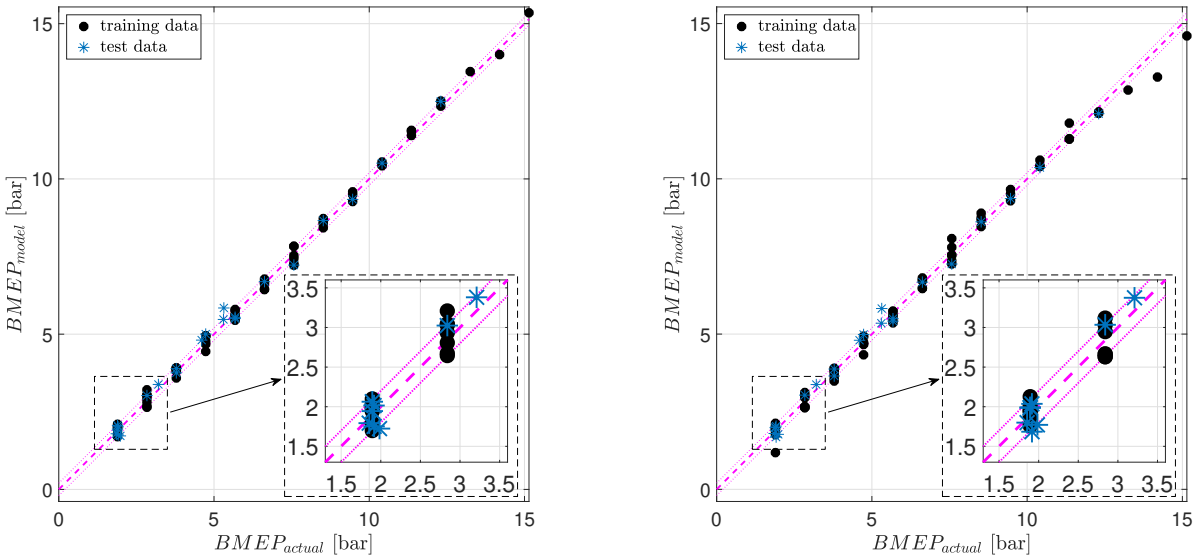
$$\mathbf{A} = \begin{bmatrix} 1 - \frac{T}{\tau_{NOx}+T} & 0 \\ 0 & 1 - \frac{T}{\tau_{BMEP}+T} \end{bmatrix} \quad (39)$$

$$\mathbf{B} = \begin{bmatrix} \frac{T}{\tau_{NOx}+T} & 0 \\ 0 & \frac{T}{\tau_{BMEP}+T} \end{bmatrix}$$

Therefore, the HOM and LOM Nonlinear Control Oriented Models (NCOM) for NO_x and BMEP are obtained as Eqs. (31) and (30). The open-loop response for HOM-NCOM and LOM-NCOM for NO_x and BMEP at four different engine speeds of 1250, 1500, 1750, and 2000 rpm are shown in Fig. 12, 13, 14, and 15. In all cases, P_r and m_f are the system inputs are applied to both HOM-NCOM and LOM-NCOM. In each transient test, the engine speeds remains constant (with a 10 rpm tolerance). In all of these plots the open-loop response of both HOM-NCOM and LOM-NCOM based on the model vs actual measurements are shown. For BMEP both the HOM-NCOM and LOM-NCOM follow the experimental closely as expected when looking at Fig. 11. However, NO_x response for LOM-NCOM has different accuracies at different engine speeds. For instance, in Fig. 15 and Fig. 13, the LOM-NCOM NO_x response is less accurate than the HOM model. Nonetheless, NO_x response for HOM-NCOM is accurate at the all speeds studied. As the HOM-NCOM model has an accurate response over a wide range of engine operating points and is an accurate model for possible use as a virtual plant to simulate the designed controller before implementation in a real-time system. Additionally, it can be used as an accurate model for a NO_x sensor fault detecting algorithm. As the LOM-NCOM has a simple structure, it is quite suitable for designing a model-based robust controller such as Sliding Mode Controller (SMC)[85] and is also capable of predicting samples ahead based on the current states and the inputs of the system. A robust controller can be used to overcome the model mismatch between LOM-NCOM and HOM-NCOM.

Conclusions

A Model Order Reduction (MOR) algorithm is developed using support vector machine (SVM) approach to predict the steady-state NO_x and BMEP of a medium-duty diesel engine. Based on the proposed SVM-based MOR algorithm and starting with a 34-feature Full-Order Model (FOM), a High-Order Model (HOM) and a Low-Order Model (LOM) are developed to predict the steady-state NO_x emission and BMEP. The features of the models are calculated based on orders 1 to 4 of the main model inputs and the interactions of them. In this study, 84 engine operating points are considered, 74% of which is used to train the steady-state NO_x and BMEP , and 26% is used as test data. The model inputs are engine speed, injected fuel amount, and fuel rail pressure. The results of the steady-state model show that the HOM model has an accurate prediction but a more complex structure with 29 features



(a) of BMEP (with 20 features)

(b) LOM of BMEP (with 6 features)

Figure 11. Prediction vs actual data for High-Order Model (HOM) and Low-Order Model (LOM) of BMEP

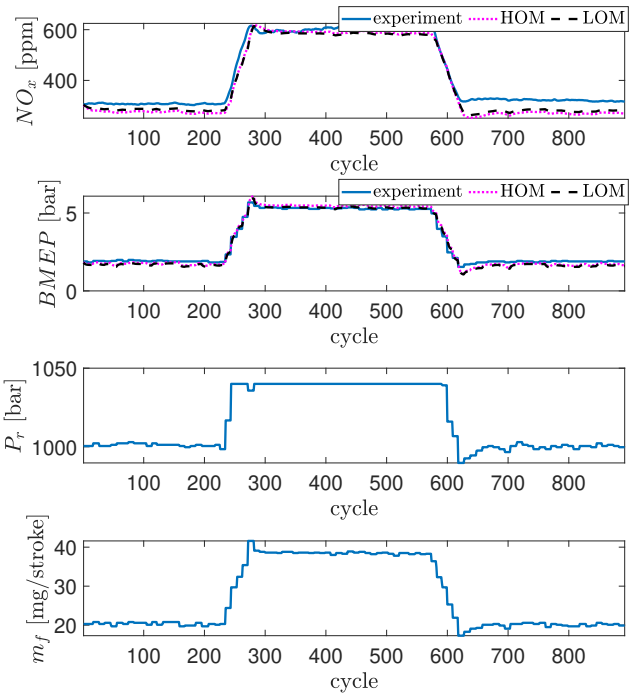


Figure 12. Transient response at engine speed $= 1250\ rpm$

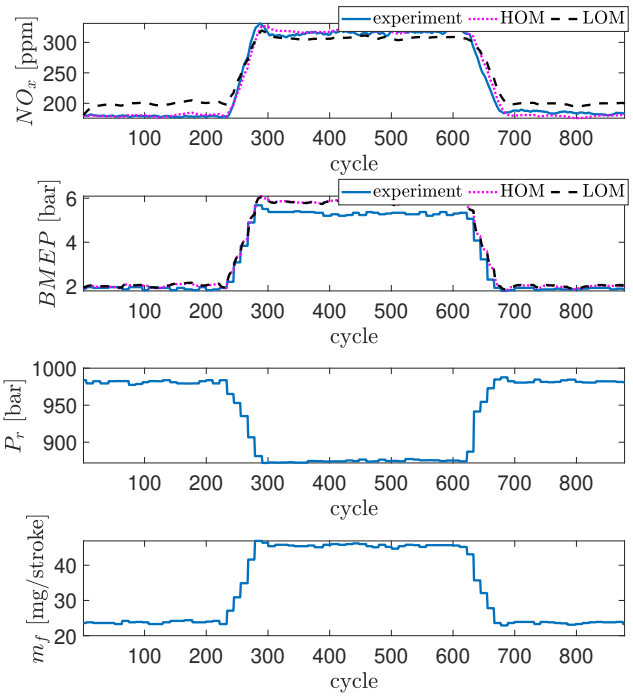


Figure 13. Transient response at engine speed $= 1500\ rpm$

for NO_x and 20 features for BMEP. For the steady-state NO_x model, the squared correlation coefficient of test (R^2_{ts}) is equal to 0.9724, 0.9725, and 0.9677 for FOM, HOM, and LOM, respectively. The R^2_{ts} value is equal to 0.9957, 0.9957, and 0.9962 for FOM, HOM, and LOM, respectively for the BMEP steady-state model. Consequently, by removing unnecessary features based on the SVM-based MOR algorithm, the performance of the HOM for both NO_x and BMEP is enhanced while the HOM complexity decreases 27.9 % with respect to the FOM. The LOM model has an acceptable accuracy with squared correlation coefficient of 0.9393 for NO_x and 0.9961 for BMEP while it has 77.9 % and 69.4 % fewer features with respect to the FOM and HOM, respectively. All FOM, HOM, and LOM SVM models of NO_x and BMEP are compared with an ANN, and results show shorter training time and more accurate results in the test data for the SVM models compared to the ANN. The SVM model training are at least 5 to 14 times faster than the corresponding ANN models with the same set of features, for NO_x and BMEP respectively. In addition, the use of a linear kernel in the SVM make it more suitable for real-time applications and for control-oriented models.

Then, a nonlinear control-oriented model (NCOM) is developed based on the developed SVM models to predict the transient behavior of the system. A fast response electrochemical NO_x sensor is used to verify the transient response of the NCOM. The transient results of HOM and LOM are compared to experimental data showing an accurate and robust prediction of engine BMEP at

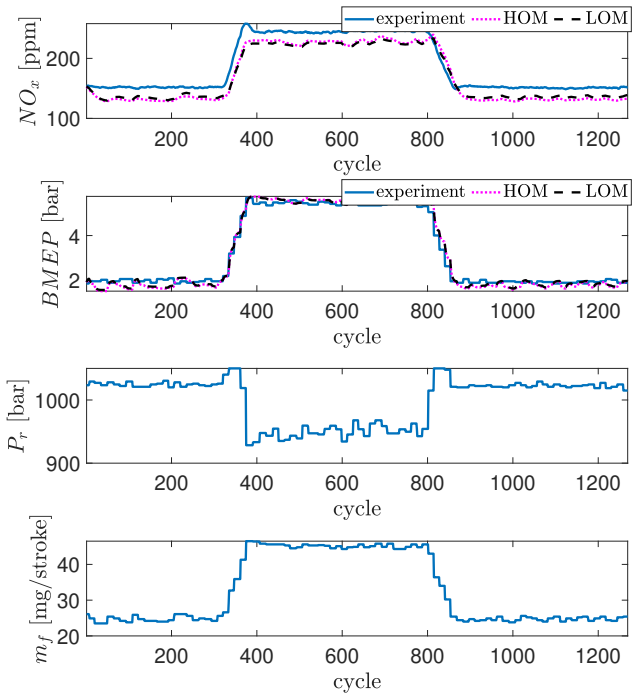


Figure 14. Transient response at engine speed $= 1750\ rpm$

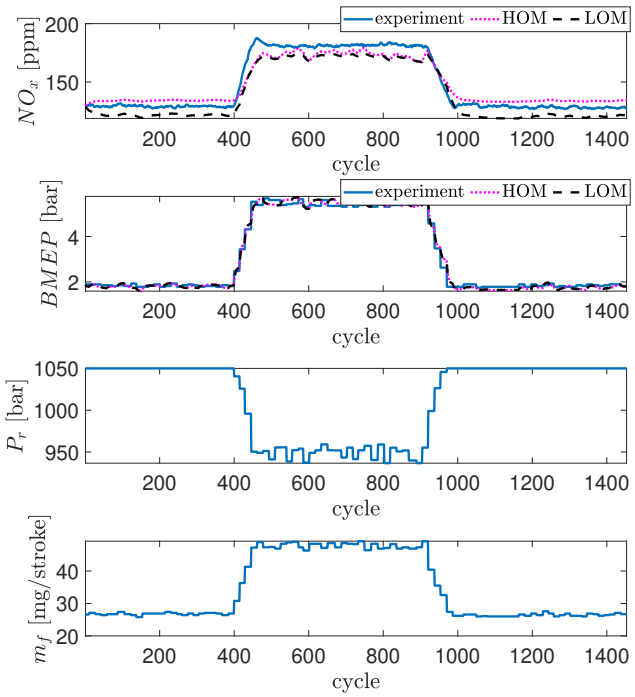


Figure 15. Transient response at engine speed $= 2000\ rpm$

different engine speeds for rising and falling step changes of the fuel rail pressure and the injected fuel amount for HOM. Additionally, the LOM model has an accurate response at different speeds for BMEP; however, the NO_x prediction with LOM has varying accuracy at different engine speeds. It can be concluded that the HOM can predict NO_x and BMEP over a wide range of operating points, which makes it ideal to be used as a virtual plant for fault detection purposes. The LOM has a simpler structure, and an acceptable accuracy which makes it useful in designing a model-based robust controller such as sliding mode or backstepping controllers. Extending the MOR algorithm to predict the steady-state values of other engine emissions and developing a NO_x sensor fault detection algorithm based on this model will be investigated in future work.

Acknowledgements

The authors gratefully acknowledge financial support from the Natural Sciences Research Council of Canada Grant 2016-04646 and from Canada First Research Excellence Fund through Future Energy Systems.

References

1. Vuorinen A. Planning of Optimal Power Systems. Vammalan Kirjapaino Oy, Vammala, Finland; 2009.
2. Tutak W, Jamrozik A, Bereczky Á, Lukacs K. Effects of injection timing of diesel fuel on performance and emission of dual fuel diesel engine powered by diesel/E85 fuels. Transport. 2018;33(3):633–646.
3. Chang YC, Lee WJ, Wu TS, Wu CY, Chen SJ. Use of water containing acetone–butanol–ethanol for NOx-PM (nitrogen oxide-particulate matter) trade-off in the diesel engine fueled with biodiesel. Energy. 2014;64:678 – 687.
4. Bohl T, Smallbone A, Tian G, Roskilly AP. Particulate number and NOx trade-off comparisons between HVO and mineral diesel in HD applications. Fuel. 2018;215:90 – 101.
5. Johnson TV. Review of Vehicular Emissions Trends. SAE Int J Engines. 2015 04;8.
6. Praveena V, Martin MLJ. A review on various after treatment techniques to reduce NOx emissions in a CI engine. Journal of the Energy Institute. 2017.
7. Geng P, Tan Q, Zhang C, Wei L, He X, Cao E, et al. Experimental investigation on NOx and green house gas emissions from a marine auxiliary diesel engine using ultralow sulfur light fuel. Science of The Total Environment. 2016;572(Supplement C):467 – 475.
8. Blanco-Rodriguez ID. Modelling and observation of exhaust gas concentrations for diesel engine control. Springer; 2014.
9. Koebel M, Elsener M, Kleemann M. Urea-SCR: a promising technique to reduce NOx emissions from automotive Diesel engines. Catalysis Today. 2000;59(34):335 – 345.
10. Maiboom A, Tauzia X, Hétet JF. Experimental study of various effects of Exhaust Gas Recirculation (EGR) on combustion and emissions of an automotive direct injection Diesel engine. Energy. 2008;33(1):22–34.
11. Ebrahimi K, Aliramezani M, Koch CR. An HCCI Control Oriented Model that Includes Combustion Efficiency. IFAC-PapersOnLine. 2016;49(11):327–332.
12. Gordon D, Wouters C, Wick M, Lehrheuer B, Andert J, Koch C, et al. Development and experimental validation of a field programmable gate array–based in-cycle direct water injection control strategy for homogeneous charge compression ignition combustion stability. International Journal of Engine Research. 0;0(0):1468087419841744.
13. Chavannavar P. Development and Implementation of a Mapless, Model Based SCR Control System. SAE Int J Engines. 2014 07;7:1113–1124.
14. Aliramezani M, Koch CR, Hayes RE. Estimating tailpipe NOx concentration using a dynamic NOx/ammonia cross

sensitivity model coupled to a three state control oriented SCR model. IFAC-PapersOnLine. 2016;49(11):8 – 13. 8th IFAC Symposium on Advances in Automotive Control AAC 2016.
15. Triantafyllopoulos G, Katsaounis D, Karamitros D, Ntziachristos L, Samaras Z. Experimental assessment of the potential to decrease diesel NOx emissions beyond minimum requirements for Euro 6 Real Drive Emissions (RDE) compliance. Science of the Total Environment. 2018;618:1400–1407.
16. Senda T, Harumi K. Prospects and Challenges for the Future of Marine Power Systems. Marine Engineering. 2018;53(3):279–284.
17. Tang W, Siani A, Chen F, Chen B. On Developing Advanced Catalysts Systems to Meet China New Regulations. SAE Technical Paper; 2019.
18. Wong PK, Wong HC, Vong CM. Online time-sequence incremental and decremental least squares support vector machines for engine air-ratio prediction. International Journal of Engine Research. 2012;13(1):28–40.
19. Aliramezani M, Koch CR, Patrick R. Phenomenological model of a solid electrolyte NOx and O2 sensor using temperature perturbation for on-board diagnostics. Solid State Ionics. 2018;321:62 – 68.
20. Ansari E, Menucci T, Shahbakhti M, Naber J. Experimental investigation into effects of high reactive fuel on combustion and emission characteristics of the Diesel-Natural gas Reactivity Controlled Compression Ignition engine. Applied Energy. 2019;239:948–956.
21. Potenza R, Dunne J, Vulli S, Richardson D, King P. Multicylinder engine pressure reconstruction using NARX neural networks and crank kinematics. International Journal of Engine Research. 2007;8(6):499–518.
22. He Y, Rutland C. Application of artificial neural networks in engine modelling. International Journal of Engine Research. 2004;5(4):281–296.
23. Rezaei J, Shahbakhti M, Bahri B, Aziz AA. Performance prediction of HCCI engines with oxygenated fuels using artificial neural networks. Applied Energy. 2015;138:460–473.
24. Moiz AA, Pal P, Probst D, Pei Y, Zhang Y, Som S, et al. A machine learning-genetic algorithm (ML-GA) approach for rapid optimization using high-performance computing. SAE International Journal of Commercial Vehicles. 2018;11(2018-01-0190):291–306.
25. Badra J, Khaled F, Tang M, Pei Y, Kodavasal J, Pal P, et al. Engine Combustion System Optimization Using CFD and Machine Learning: A Methodological Approach. In: Internal Combustion Engine Division Fall Technical Conference. vol. 59346. American Society of Mechanical Engineers; 2019. p. V001T06A007.
26. Owoyele O, Pal P. A Novel Active Optimization Approach for Rapid and Efficient Design Space Exploration Using Ensemble Machine Learning. In: ASME 2019 Internal Combustion Engine Division Fall Technical Conference. American Society of Mechanical Engineers Digital Collection;. .
27. Xu Y, Guo R, Wang L. A twin multi-class classification support vector machine. Cognitive computation. 2013;5(4):580–588.
28. Tanveer M. Robust and sparse linear programming twin support vector machines. Cognitive Computation. 2015;7(1):137–149.
29. Bertram AM, Kong SC. Computational Optimization of a Diesel Engine Calibration Using a Novel SVM-PSO Method. SAE Technical Paper; 2019.
30. Hanuschkin A, Schober S, Bode J, Schorr J, Böhm B, Krüger C, et al. Machine learning–based analysis of in-cylinder flow fields to predict combustion engine performance. International Journal of Engine Research. 0;0(0):1468087419833269.
31. Yusri IM, Majeed APPA, Mamat R, Ghazali MF, Awad OI, Azmi WH. A review on the application of response surface method and artificial neural network in engine performance and exhaust emissions characteristics in alternative fuel. Renewable and Sustainable Energy Reviews. 2018;90:665 – 686.
32. Silitonga AS, Masjuki HH, Ong HC, Sebayang AH, Dharma S, Kusumo F, et al. Evaluation of the engine performance and exhaust emissions of biodiesel-bioethanol-diesel blends using kernel-based extreme learning machine. Energy. 2018;159:1075 – 1087.
33. Wong PK, Gao XH, Wong KI, Vong CM. Online extreme learning machine based modeling and optimization for point-by-point engine calibration. Neurocomputing. 2018;277:187–197.
34. Aliramezani M, Norouzi A, Koch CR. Support vector machine for a diesel engine performance and NOx emission control-oriented model. In: 21st IFAC World Congress; 2020. .
35. Gordon D, Wouters C, Wick M, Lehrheuer B, Andert J, Koch C, et al. Development and experimental validation of a field programmable gate array–based in-cycle direct water injection control strategy for homogeneous charge compression ignition combustion stability. International Journal of Engine Research. 2019;20(10):1101–1113.
36. Gordon D, Wouters C, Wick M, Xia F, Lehrheuer B, Andert J, et al. Development and experimental validation of a real-time capable field programmable gate array–based gas exchange model for negative valve overlap. International Journal of Engine Research. 2020;21(3):421–436.
37. Duan H, Huang Y, Mehra RK, Song P, Ma F. Study on influencing factors of prediction accuracy of support vector machine (SVM) model for NOx emission of a hydrogen enriched compressed natural gas engine. Fuel. 2018;234:954–964.
38. Liu B, Hu J, Yan F, Turkson RF, Lin F. A novel optimal support vector machine ensemble model for NOx emissions prediction of a diesel engine. Measurement. 2016;92:183–192.
39. Niu X, Yang C, Wang H, Wang Y. Investigation of ANN and SVM based on limited samples for performance and emissions prediction of a CRDI-assisted marine diesel engine. Applied Thermal Engineering. 2017;111:1353–1364.
40. Zuo Q, Zhu X, Liu Z, Zhang J, Wu G, Li Y. Prediction of the performance and emissions of a spark ignition engine fueled with butanol-gasoline blends based on support vector regression. Environmental Progress & Sustainable Energy. 2019;38(3).
41. Azzam M, Awad M, Zeaiter J. Application of evolutionary neural networks and support vector machines to model NOx emissions from gas turbines. Journal of environmental chemical engineering. 2018;6(1):1044–1052.
42. Zhou H, Zhao JP, Zheng LG, Wang CL, Cen KF. Modeling NOx emissions from coal-fired utility boilers using support vector regression with ant colony optimization. Engineering Applications of Artificial Intelligence. 2012;25(1):147–158.
43. Lv Y, Liu J, Yang T, Zeng D. A novel least squares support vector machine ensemble model for NOx emission prediction of a coal-fired boiler. Energy. 2013;55:319–329.

44. Lv Y, Yang T, Liu J. An adaptive least squares support vector machine model with a novel update for NOx emission prediction. Chemometrics and Intelligent Laboratory Systems. 2015;145:103–113.
45. Wong KI, Wong PK, Cheung CS, Vong CM. Modeling and optimization of biodiesel engine performance using advanced machine learning methods. Energy. 2013;55:519–528.
46. Ye J, Xiong T. SVM versus least squares SVM. In: Artificial Intelligence and Statistics; 2007. p. 644–651.
47. Dietterich T. Overfitting and undercomputing in machine learning. ACM computing surveys. 1995;27(3):326–327.
48. Brown G, Pocock A, Zhao MJ, Luján M. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. Journal of machine learning research. 2012;13(Jan):27–66.
49. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis & Machine Intelligence. 2005;(8):1226–1238.
50. Nguyen XV, Chan J, Romano S, Bailey J. Effective global approaches for mutual information based feature selection. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM; 2014. p. 512–521.
51. Yang HH, Moody J. Data visualization and feature selection: New algorithms for nongaussian data. In: Advances in neural information processing systems; 2000. p. 687–693.
52. Hall MA. . Correlation-based Feature Selection for Machine Learning. PhD thesis. The University of Waikato (1999).
53. He W, Wang Z, Jiang H. Model optimizing and feature selecting for support vector regression in time series forecasting. Neurocomputing. 2008;72(1-3):600–611.
54. Tao X, Renmu H, Peng W, Dongjie X. Input dimension reduction for load forecasting based on support vector machines. In: 2004 IEEE International Conference on Electric Utility Deregulation, Restructuring and Power Technologies. Proceedings. vol. 2. IEEE; 2004. p. 510–514.
55. Hira ZM, Gillies DF. A review of feature selection and feature extraction methods applied on microarray data. Advances in bioinformatics. 2015;2015.
56. Schölkopf B, Smola A, Müller KR. Nonlinear component analysis as a kernel eigenvalue problem. Neural computation. 1998;10(5):1299–1319.
57. Salah Mohamed K. Machine learning for model order reduction. Springer; 2018.
58. Wang JM, Chu CC, Yu Q, Kuh ES. On projection-based algorithms for model-order reduction of interconnects. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications. 2002;49(11):1563–1585.
59. Alam M, Nieuwoudt A, Massoud Y. Wavelet-based passivity preserving model order reduction for wideband interconnect characterization. In: 8th International Symposium on Quality Electronic Design (ISQED'07). IEEE; 2007. p. 432–437.
60. Barrasso D, Tamrakar A, Ramachandran R. A reduced order PBM–ANN model of a multi-scale PBM–ANN description of a wet granulation process. Chemical Engineering Science. 2014;119:319–329.
61. Parmar G, Mukherjee S, Rasad R. REDUCED ORDER MODELLING OF LINEAR MIMO SYSTEMS USING GENETIC ALGORITHM. International Journal of Simulation Modelling (IJSIMM). 2007;6(3).
62. Smola AJ, Schölkopf B. A tutorial on support vector regression. Statistics and computing. 2004;14(3):199–222.
63. Kihas D, Pachner D, Baramov L, Uchanski M, Naik P, Khaled N. Concept analysis and initial results of engine-out NOx estimator suitable for on ECM implementation. SAE Technical Paper; 2016.
64. Quérel C, Grondin O, Letellier C. Semi-physical mean-value NOx model for diesel engine control. Control Engineering Practice. 2015;40:27–44.
65. Guardiola C, Martín J, Pla B, Bares P. Cycle by cycle NOx model for diesel engine control. Applied Thermal Engineering. 2017;110:1011–1020.
66. Rezaei R, Hayduk C, Alkan E, Kemski T, Delebinski T, Bertram C. Hybrid Phenomenological and Mathematical-Based Modeling Approach for Diesel Emission Prediction. SAE Technical Paper; 2020.
67. Irdmousa BK, Rizvi SZ, Veini JM, Nabert J, Shahbakhti M. Data-driven Modeling and Predictive Control of Combustion Phasing for RCCI Engines. In: 2019 American Control Conference (ACC). IEEE; 2019. p. 1617–1622.
68. Oğuz H, Sarıtas I, Baydan HE. Prediction of diesel engine performance using biofuels with artificial neural network. Expert Systems with Applications. 2010;37(9):6579–6586.
69. Tschanz F, Amstutz A, Onder CH, Guzzella L. Feedback control of particulate matter and nitrogen oxide emissions in diesel engines. Control engineering practice. 2013;21(12):1809–1820.
70. Aliramezani M, Norouzi A, Koch CR, Hayes RE. A control oriented diesel engine NOx emission model for on board diagnostics and engine control with sensor feedback. In: Proceedings of Combustion Institute - Canadian Section (CICS); 2019. p. 1–6.
71. Vapnik V, Lerner A. Generalized portrait method for pattern recognition. Automation and Remote Control. 1963;24(6):774–780.
72. Vapnik V, Chervonenkis A. A note on class of perceptron. Automation and Remote Control. 1964;24.
73. Drucker H, Burges CJ, Kaufman L, Smola AJ, Vapnik V. Support vector regression machines. In: Advances in neural information processing systems; 1997. p. 155–161.
74. Cortes C, Vapnik V. Support-vector networks. Machine learning. 1995;20(3):273–297.
75. Vapnik V. The nature of statistical learning theory. Springer science & business media; 2013.
76. Bellman R, et al. The theory of dynamic programming. Bulletin of the American Mathematical Society. 1954;60(6):503–515.
77. Karush W. Minima of functions of several variables with inequalities as side constraints. M Sc Dissertation Dept of Mathematics, Univ of Chicago. 1939.
78. Kuhn HW, Tucker AW. Nonlinear programming, in (J. Neyman, ed.) Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley; 1951.
79. Brualdi RA. Introductory combinatorics. Pearson Education India; 1977.
80. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:150203167. 2015.
81. Juszczak P, Tax D, Duin RP. Feature scaling in support vector data description. In: Proc. ASCI. Citeseer; 2002. p. 95–102.

82. Norouzi A, Aliramezani M, Koch CR. Diesel Engine NOx Reduction Using a PD-type Fuzzy Iterative Learning Control with a Fast Response NOx Sensor. In: Proceedings of Combustion Institute - Canadian Section (CICS); 2019. p. 7–12.
83. Ren J. ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging. Knowledge-Based Systems. 2012;26:144–153.
84. Shalev-Shwartz S, Srebro N. SVM optimization: inverse dependence on training set size. In: Proceedings of the 25th international conference on Machine learning. ACM; 2008. p. 928–935.
85. Norouzi A, Ebrahimi K, Koch CR. Integral Discrete-time Sliding Mode Control of Homogeneous Charge Compression Ignition (HCCI) Engine Load and Combustion Timing. IFAC-PapersOnLine. 2019;52(5):153–158.

Steady-state $\mathrm{NO_x}$ Model

$$\begin{aligned}\mathbf{NO_{x,HO,ss}} &= \mathbf{w^T_{NO_x}}\ \bar{\mathbf{U}}_{\mathbf{NO_x,HO}} + b_{NO_x,HO} \\ &= 528.93\ \bar{\mathbf{U}}_{\mathbf{1}} - 2411.93\ \bar{\mathbf{U}}_{\mathbf{2}} + 1544.89\ \bar{\mathbf{U}}_{\mathbf{3}} - 1177.86\ \bar{\mathbf{U}}_{\mathbf{4}} + 3654.47\ \bar{\mathbf{U}}_{\mathbf{5}} - 1495.65\ \bar{\mathbf{U}}_{\mathbf{6}} \\ &- 836.09\ \bar{\mathbf{U}}_{\mathbf{7}} + 1123.75\ \bar{\mathbf{U}}_{\mathbf{8}} - 512.52\ \bar{\mathbf{U}}_{\mathbf{9}} - 1230.75\ \bar{\mathbf{U}}_{\mathbf{11}} - 1319.38\ \bar{\mathbf{U}}_{\mathbf{12}} + 2846.76\ \bar{\mathbf{U}}_{\mathbf{13}} \\ &- 612.65\ \bar{\mathbf{U}}_{\mathbf{14}} + 1574.67\ \bar{\mathbf{U}}_{\mathbf{15}} + 529.47\ \bar{\mathbf{U}}_{\mathbf{17}} - 2329.74\ \bar{\mathbf{U}}_{\mathbf{19}} + 1214.66\ \bar{\mathbf{U}}_{\mathbf{20}} - 573.96\ \bar{\mathbf{U}}_{\mathbf{21}} \\ &+ 1076.88\ \bar{\mathbf{U}}_{\mathbf{22}} - 1129.99\ \bar{\mathbf{U}}_{\mathbf{23}} - 2012.74\ \bar{\mathbf{U}}_{\mathbf{24}} - 1001.16\ \bar{\mathbf{U}}_{\mathbf{25}} + 815.56\ \bar{\mathbf{U}}_{\mathbf{26}} + 1440.14\ \bar{\mathbf{U}}_{\mathbf{27}} \\ &- 1571.92\ \bar{\mathbf{U}}_{\mathbf{29}} + 789.89\ \bar{\mathbf{U}}_{\mathbf{31}} - 2795.74\ \bar{\mathbf{U}}_{\mathbf{32}} + 1602.41\ \bar{\mathbf{U}}_{\mathbf{33}} + 2640.48\ \bar{\mathbf{U}}_{\mathbf{34}} + 782.81\end{aligned} \tag{40}$$

$$\begin{aligned}\mathbf{NO_{x,LO,ss}} &= \mathbf{w^T_{NO_x}}\ \bar{\mathbf{U}}_{\mathbf{NO_x,LO}} + b_{NO_x,LO} \\ &= -1654.78\ \bar{\mathbf{U}}_{\mathbf{2}} + 1082.55\ \bar{\mathbf{U}}_{\mathbf{5}} + 269.24\ \bar{\mathbf{U}}_{\mathbf{8}} + 1682.28\ \bar{\mathbf{U}}_{\mathbf{15}} - 946.71\ \bar{\mathbf{U}}_{\mathbf{21}} - 320.70\ \bar{\mathbf{U}}_{\mathbf{22}} \\ &+ 2549.50\ \bar{\mathbf{U}}_{\mathbf{27}} - 4713.23\ \bar{\mathbf{U}}_{\mathbf{32}} + 2608.52n^2\ \bar{\mathbf{U}}_{\mathbf{33}} + 1017.98\end{aligned} \tag{41}$$

Steady-state BMEP Model

$$\begin{aligned}\mathbf{BMEP_{HO,ss}} &= \mathbf{w^T_{BMEP}}\ \bar{\mathbf{U}}_{\mathbf{BMEP,HO}} + b_{BMEP,HO} \\ &= +3.03\ \bar{\mathbf{U}}_{\mathbf{1}} + 1.41\ \bar{\mathbf{U}}_{\mathbf{4}} + 1.57\ \bar{\mathbf{U}}_{\mathbf{7}} + 2.04\ \bar{\mathbf{U}}_{\mathbf{8}} - 2.05\ \bar{\mathbf{U}}_{\mathbf{9}} + 0.78\ \bar{\mathbf{U}}_{\mathbf{10}} + 1.47\ \bar{\mathbf{U}}_{\mathbf{117}} \\ &- 2.03\ \bar{\mathbf{U}}_{\mathbf{18}} + 1.05\ \bar{\mathbf{U}}_{\mathbf{19}} + 0.93\ \bar{\mathbf{U}}_{\mathbf{20}} - 1.16\ \bar{\mathbf{U}}_{\mathbf{21}} - 2.18\ \bar{\mathbf{U}}_{\mathbf{24}} + 1.66\ \bar{\mathbf{U}}_{\mathbf{25}} - 1.78\ \bar{\mathbf{U}}_{\mathbf{26}} \\ &+ 1.49\ \bar{\mathbf{U}}_{\mathbf{27}} - 0.75\ \bar{\mathbf{U}}_{\mathbf{29}} - 2.12\ \bar{\mathbf{U}}_{\mathbf{30}} + 1.35\ \bar{\mathbf{U}}_{\mathbf{31}} + 0.83\ \bar{\mathbf{U}}_{\mathbf{32}} - 0.81\ \bar{\mathbf{U}}_{\mathbf{34}} + 6.52\end{aligned} \tag{42}$$

$$\begin{aligned}\mathbf{BMEP_{LO,ss}} &= \mathbf{w^T_{BMEP}}\ \bar{\mathbf{U}}_{\mathbf{BMEP,LO}} + b_{BMEP,LO} \\ &= +6.36\ \bar{\mathbf{U}}_{\mathbf{1}} - 3.54\ \bar{\mathbf{U}}_{\mathbf{18}} - 5.00\ \bar{\mathbf{U}}_{\mathbf{21}} + 6.10\ \bar{\mathbf{U}}_{\mathbf{25}} + 3.89\ \bar{\mathbf{U}}_{\mathbf{27}} - 2.95\ \bar{\mathbf{U}}_{\mathbf{30}} + 6.62\end{aligned} \tag{43}$$