

## Accepted version on Author's Personal Website: C. R. Koch

Article Name with DOI link to Final Published Version complete citation:

Armin Norouzi, Charles Robert Koch, and Masoud Aliramezani. A correlation based model order reduction approach for a diesel engine nox and bmep dynamic model using machine learning. *International Journal of Engine Research*, 2020. doi: [10.1177/1468087420936949](https://doi.org/10.1177/1468087420936949)

### See also:

[https://sites.ualberta.ca/~ckoch/open\\_access/IJER\\_AN\\_2020.pdf](https://sites.ualberta.ca/~ckoch/open_access/IJER_AN_2020.pdf)

Post-print

As per publisher copyright is ©2020



This work is licensed under a  
[Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).



Article accepted version starts on the next page →

[Or link: to Author's Website](#)

# A correlation-based model order reduction approach for a diesel engine $\text{NO}_x$ and brake mean effective pressure dynamic model using machine learning

*International J of Engine Research*  
1–19

© IMechE 2020

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/1468087420936949

[journals.sagepub.com/home/ijer](https://journals.sagepub.com/home/ijer)**Armin Norouzi<sup>ID</sup>, Masoud Aliramezani and Charles Robert Koch<sup>ID</sup>**

## Abstract

A correlation-based model order reduction algorithm is developed using support vector machine to model  $\text{NO}_x$  emission and break mean effective pressure of a medium-duty diesel engine. The support vector machine-based model order reduction algorithm is used to reduce the number of features of a 34-feature full-order model by evaluating the regression performance of the support vector machine-based model. Then, the support vector machine-based model order reduction algorithm is used to reduce the number of features of the full-order model. Two models for  $\text{NO}_x$  emission and break mean effective pressure are developed via model order reduction, one complex model with high accuracy, called high-order model, and the other with an acceptable accuracy and a simple structure, called low-order model. The high-order model has 29 features for  $\text{NO}_x$  and 20 features for break mean effective pressure, while the low-order model has nine features for  $\text{NO}_x$  and six features for break mean effective pressure. Then, the steady-state low-order model and high-order model are implemented in a nonlinear control-oriented model. To verify the accuracy of nonlinear control-oriented model, a fast response electrochemical  $\text{NO}_x$  sensor is used to experimentally study the engine transient  $\text{NO}_x$  emissions. The high-order model and low-order model support vector machine models of  $\text{NO}_x$  and break mean effective pressure are compared to a conventional artificial neural network with one hidden layer. The results illustrate that the developed support vector machine model has shorter training times (5–14 times faster) and higher accuracy especially for test data compared to the artificial neural network model. A control-oriented model is then developed to predict the dynamic behavior of the system. Finally, the performance of the low-order model and high-order model is evaluated for different rising and falling input transients at four different engine speeds. The transient test results validate the high accuracy of the high-order model and the acceptable accuracy of low-order model for both  $\text{NO}_x$  and break mean effective pressure. The high-order model is proposed as an accurate virtual plant while the low-order model is suitable for model-based controller design.

## Keywords

Machine learning, support vector machine, model order reduction, diesel engine, emissions, control-oriented model

Date received: 26 February 2020; accepted: 29 May 2020

## Introduction

### *Diesel engine $\text{NO}_x$ 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.<sup>1,2</sup> The high combustion temperature and the lean air–fuel mixture of diesel engines lead to a relatively high  $\text{NO}_x$  emission. Reducing diesel engine  $\text{NO}_x$  and particulate matter (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.<sup>3,4</sup> This is mainly due to the inverse effect of in-cylinder peak

---

Department of Mechanical Engineering, University of Alberta, Edmonton, AB, Canada

### Corresponding author:

Armin Norouzi, Department of Mechanical Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada.

Email: [norouziy@ualberta.ca](mailto:norouziy@ualberta.ca)

temperature and air–fuel ratio on  $\text{NO}_x$  and PM.<sup>5–7</sup> Therefore, to develop an effective and practical  $\text{NO}_x$  control strategy, all other gaseous and PM emissions should also be considered.

The  $\text{NO}_x$  components of diesel engine exhaust gas typically contain approximately 1/7–1/3  $\text{NO}_2/\text{NO}$  fraction<sup>8</sup> while  $\text{NO}_2/\text{NO}$  ratio increases after the diesel oxidation catalyst (DOC) to approximately one.<sup>9</sup> Different methods have been used by the automotive industry to reduce diesel engine  $\text{NO}_x$  emission, including exhaust gas recirculation (EGR),<sup>10</sup> low-temperature combustion (LTC)<sup>11,12</sup> and most effectively, urea-based selective catalytic reduction (SCR).<sup>13–15</sup> Although all these methods can help reducing the engine tailpipe  $\text{NO}_x$  emissions, it has become more difficult to keep pace with increasingly stringent emission regulations by only using the conventional  $\text{NO}_x$  reduction approaches.<sup>16,17</sup> To address this issue, more complex engine control strategies and after treatment systems are needed.<sup>18–20</sup> 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 modeling internal combustion engines*

Data-driven models have become of especial interest to many of the researchers in the last two decades as an efficient way to predict, optimize or diagnose the performance of internal combustion engines (ICEs) mostly using artificial neural network (ANN).<sup>21–23</sup> Machine learning (ML) is increasingly used for ICE optimization and calibration. Machine learning methods, along with the state-of-the-art optimization algorithms such as genetic algorithm (ML-GA), were also used for ICE optimization. For instance, an ML-GA 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.<sup>24,25</sup> 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 has shown a better performance compared with the GA, micro GA, and particle swarm optimization (PSO).<sup>26</sup>

Support vector machine (SVM) is another popular data-driven method that is now being increasingly used for modeling ICEs mostly for steady-state prediction of engine performance or emissions. SVM is an ML technique which is capable of modeling complex and nonlinear input–output relations based on a sufficiently large training data set.<sup>27–29</sup> 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.<sup>30–34</sup> The simple structure of SVM, especially

using linear kernel SVM when compared to ANN, allows for improved learning performance which, when combined with the rapid in-cycle field programmable gate array (FPGA)–based calculations, could allow for online learning of in-cycle combustion metrics.<sup>12,35</sup>

An  $\text{NO}_x$  prediction model was developed for a hydrogen-enriched compressed natural gas engine using an optimal SVM method where PSO was used to find the regulatory parameters of SVM.<sup>36</sup> Also, the effect of SVM model parameters such as penalty factor kernel, insensitive band loss function, and the training sample size was studied in Duan et al.<sup>36</sup> An optimal SVM for diesel engine  $\text{NO}_x$  prediction was developed by Liu et al.,<sup>37</sup> 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.<sup>37</sup>

An ANN and an SVM model are developed by Niu et al.<sup>38</sup> to predict the performance and emission of a marine diesel engine. The developed ANN and SVM were compared and it 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.<sup>38</sup> 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 Zuo et al.<sup>39</sup> Predicting  $\text{NO}_x$  emission using SVM method is not limited to ICEs. SVM has also been used to predict  $\text{NO}_x$  emissions of gas turbines<sup>40</sup> and coal-fired utility boilers.<sup>41</sup> The least-square type support vector machine (LS-SVM) has also been used for  $\text{NO}_x$  emission prediction.<sup>42–44</sup> 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.<sup>45</sup>

### *Model order reduction and feature selection*

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.<sup>46</sup> 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 the so-called merit function are selected. Similarly, in MOR, the less significant features are removed based on pre-defined criteria. Typical FS steps are as follows: (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, and (4) repeat (2) and (3) until certain number of features is selected.<sup>47</sup>

Different methods have been developed for FS problem such as minimum-redundancy-maximum-relevance (mRMR),<sup>48</sup> conditional mutual information,<sup>49</sup> joint mutual information,<sup>50</sup> and correlation FS.<sup>51</sup> In a correlation-based FS approach, typically, a merit function is defined based on the feature-output correlation, and the criteria for FS are 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.<sup>51</sup> An FS for the SVM algorithm in time series forecasting was proposed in He et al.,<sup>52</sup> where hyperparameters were optimized based on the 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.<sup>52</sup> An SVM-based FS method was proposed for electric load forecasting in Tao et al.<sup>53</sup> where the mean absolute percentage error (MAPE) was used for the algorithm stoppage criteria.<sup>53</sup>

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 the so-called feature extraction methods.<sup>54</sup> The PCA is one of the 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).<sup>54,55</sup> 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.<sup>56</sup> The other popular MOR class includes time-domain approaches, such as Chebyshev and Wavelet.<sup>57,58</sup> Also, different ML techniques were used for MOR, including neural network, GA, fuzzy logic, PSO, simulated annealing, and ant colony.<sup>56,59,60</sup>

A wide range of methods have been used in the literature for FS; however, SVM-based MOR or FS 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.<sup>61</sup> 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 an MOR algorithm.

### Motivation and objectives

Several phenomenological and physical control-oriented models (COMs) of diesel engine performance and emission have been developed for feedback control of emissions.<sup>62–64</sup> Physical and chemical submodels 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 affect gaseous and particulate emissions. However, despite the ongoing advancements in the physics-based combustion and emission models, they cannot 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.<sup>65</sup>

While the physics-based combustion and emission modeling approach provides physical insight, an accurate detailed three-dimensional (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.<sup>65,66</sup> ML-based model that is trained with appropriate data and sufficiently large data size can provide an accurate, flexible, and fast data-driven prediction.<sup>67</sup> These data-driven models are typically faster than physics-based approaches and can be easily trained using experimental data.<sup>65</sup> Developing ML-based engine models also requires a deep understanding of the system and the modeling approach. Most of the state-of-the-art ML 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 article provides comprehensive examination of the design parameters of an SVM model to be used for ICE 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 data sets compared to other data-driven approaches used for ICEs. First, most of the data-driven models developed for ICEs have used nonlinear kernels without exclusively investigating feature interactions. Second, none of these models propose a COM to be used for engine control or observer design. In addition, use of a correlation-based FS 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  $\text{NO}_x$  sensors can now provide accurate input for engine feedback control. Therefore, developing accurate 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.<sup>68</sup>

A full-order model (FOM) SVM is first developed to predict engine  $\text{NO}_x$  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 FS approach is proposed and used to reduce the order of the FOM. Next, two high-order models (HOMs) and two low-order models (LOMs) were derived for  $\text{NO}_x$  and BMEP using the proposed MOR approach, and the results are compared with a two-layer ANN model (one hidden layer and one output layer). Finally, a COM was developed by adding a first-order lag that represents the system dynamics for BMEP and  $\text{NO}_x$ . The results show that the high-order and low-order COMs were both capable of accurately tracking the transient  $\text{NO}_x$  emission and engine BMEP at different engine speeds and loads for experimental data that were not used to develop the model.

This article is organized as follows. The “Experimental setup” section provides information about experimental setup, data collection, and collected data map. Section “SVM” presents background information about SVM and problem formulating. In section “FOM,” the FOM for  $\text{NO}_x$  and BMEP is developed. The “MOR” section introduces a novel correlation-based MOR algorithm, and the corresponding results for the MOR are provided for  $\text{NO}_x$  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  $\text{NO}_x$  nonlinear control-oriented model” presents a nonlinear COM along with the dynamic model validation with experimental data. The main conclusions of this article are detailed in the “Conclusion” section.

## Experimental setup

### Diesel engine

A four-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.<sup>69</sup>

| Parameter            | Value                              |
|----------------------|------------------------------------|
| Engine type          | In-line, four-cylinder             |
| Displacement         | 4.5 L                              |
| Bore $\times$ stroke | 102 mm $\times$ 120 mm             |
| Peak torque          | 624 N m at 1500 r/min              |
| Peak power           | 123 kW at 2000 r/min               |
| Aspiration           | Turbocharged and charge air cooled |
| Certification level  | Tier 3/stage IIIA                  |

### Electrochemical $\text{NO}_x$ sensor

A production amperometric  $\text{NO}_x$  sensor (ECM-06-05) was used in the experiments. All the sensor working parameters were set using the sensor control module (ECM- $\text{NO}_x\text{CAN}$  P/N: 02-07). The sensor control module was connected to a computer via a controller area network (CAN) interface (*Kvaser Light HS*) to monitor and log the measurements.

### Fourier-transform infrared spectroscopy

A Fourier-transform infrared spectroscopy (FTIR) analyzer (*MultiGas 2030*) was used to validate the ECM  $\text{NO}_x$  sensor measurement and to measure the concentration of other species in the exhaust gas. The FTIR analyzer 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 °C to avoid water vapor condensation in the sample gas as schematically shown in Figure 1. The engine operating conditions and corresponding  $\text{NO}_x$  concentrations are shown in Figure 2.

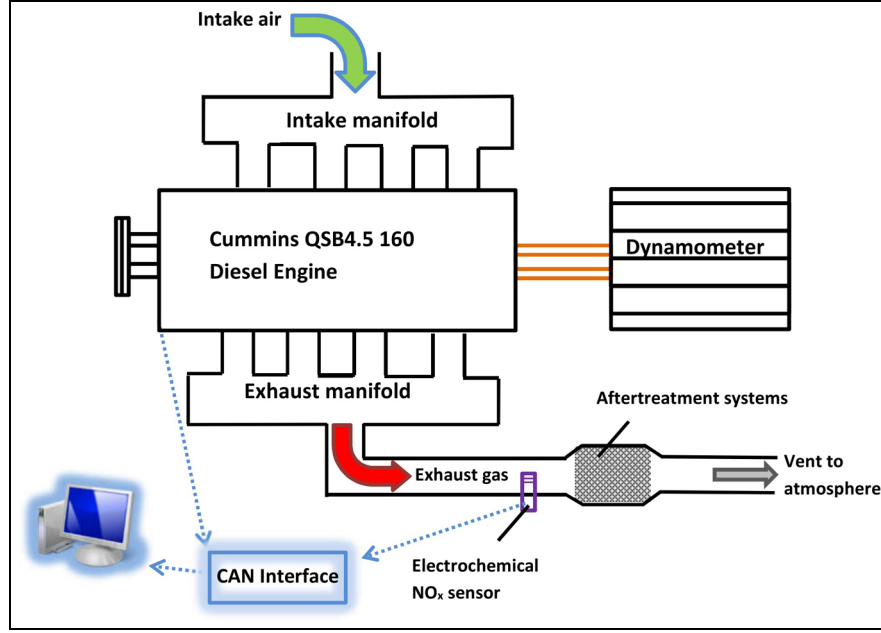
## SVM

### Convex optimization problem

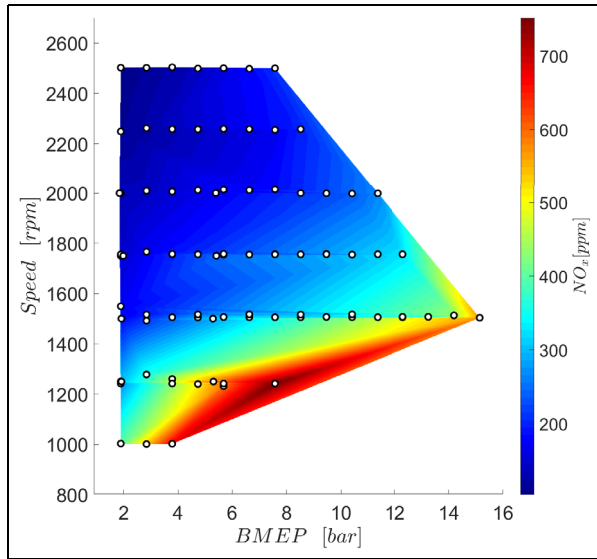
SVM, introduced by Vapnik,<sup>70,71</sup> is a supervised ML approach. SVM is typically used for classification of labeled data by creating a set of hyperplanes in an infinite-dimensional space.<sup>72</sup> SVM is also used for regression and function approximation, also called the support vector regression (SVR), which was introduced by Cortes and Vapnik.<sup>73</sup> 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 as follows:

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

In other words, the optimization problem is to find the flattest function for which the acceptable deviation



**Figure 1.** Experiment setup—internal combustion engine and fast response  $\text{NO}_x$  sensor.



**Figure 2.** Experimental engine operating points used for model training.

from training data is at most  $\epsilon$ . The optimal hyperplane to which describes the training data,  $\{\mathbf{u}_i, \mathbf{z}_i\}$ , can be defined as

$$\mathbf{y}(\mathbf{u}_i) = \mathbf{w}^T \mathbf{u}_i + \mathbf{b} \quad (1)$$

where  $\mathbf{w}$  and  $\mathbf{b}$  are found by solving the SVM algorithm for regression problems. Flatness of  $\mathbf{y}(\mathbf{u}_i)$  in equation (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  $\epsilon$ . Then, the

optimization problem to find the optimum  $\mathbf{y}(\mathbf{u}_i)$  is defined as

$$\begin{aligned} &\text{Minimize : } \frac{1}{2} \|\mathbf{w}\|_2^2 \\ &\text{Subject to : } \begin{cases} \mathbf{z}_i - \mathbf{w}^T \mathbf{u}_i - \mathbf{b} \leq \epsilon \\ \mathbf{w}^T \mathbf{u}_i + \mathbf{b} - \mathbf{z}_i \leq -\epsilon \end{cases} \quad i = 1, \dots, n \end{aligned} \quad (2)$$

The convex optimization problem, equation (2), is feasible when such  $\mathbf{y}(\mathbf{u}_i)$  exists which is as flat as possible and approximates all training data with at most  $\epsilon$  deviation. In other words, the convex optimization problem is feasible when

$$-\epsilon \leq \mathbf{z}_i - \mathbf{y}_i \leq \epsilon \quad (3)$$

So, the  $\epsilon$ -insensitive linear loss function is defined as<sup>73</sup>

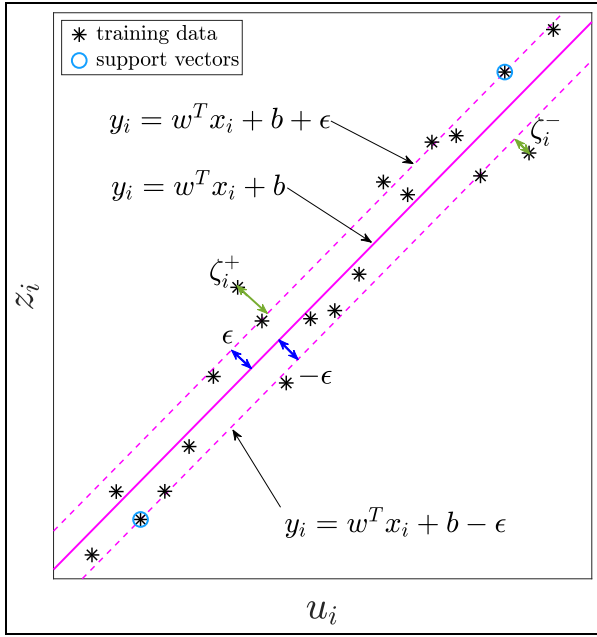
$$L_\epsilon(\mathbf{z}_i, \mathbf{y}_i) = \begin{cases} 0 & |\mathbf{z}_i - \mathbf{y}_i| \leq \epsilon \\ |\mathbf{z}_i - \mathbf{y}_i| - \epsilon & \text{otherwise} \end{cases} \quad (4)$$

where the loss function would be zero if training error is less than  $\epsilon$ . Also, the empirical risk function,  $R_{emp}$ , is defined based on the loss function as<sup>74</sup>

$$R_{emp}(\mathbf{w}, \mathbf{b}) = \frac{1}{N} \sum_{i=1}^N L_\epsilon(\mathbf{z}_i, \mathbf{y}_i) \quad (5)$$

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 equation (3) to overcome the above optimization problem infeasibility as

$$-\epsilon - \zeta_i^- \leq \mathbf{z}_i - \mathbf{y}_i \leq \epsilon + \zeta_i^+ \quad (6)$$



**Figure 3.** SVM regression and support vectors example.

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 equation (6) as

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

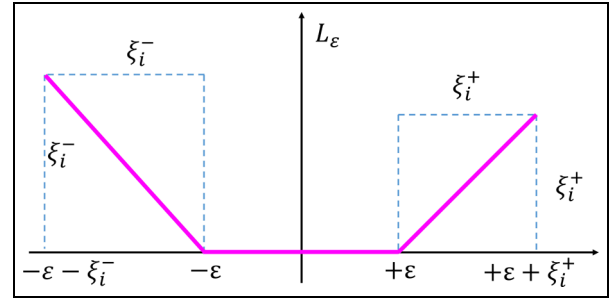
Then, the convex optimization problem is modified by adding the minimizing empirical risk function term to equation (2)

$$\begin{aligned} \text{Minimize : } & \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^n (\zeta_i^+ + \zeta_i^-) \\ \text{Subject to : } & \begin{cases} \mathbf{z}_i - \mathbf{w}^T \mathbf{u}_i - \mathbf{b} \leq \epsilon + \zeta_i^+ \\ \mathbf{w}^T \mathbf{u}_i + \mathbf{b} - \mathbf{z}_i \leq \epsilon + \zeta_i^- \\ \zeta_i^-, \zeta_i^+ \geq 0 \end{cases} \end{aligned} \quad (8)$$

where  $C$  is the 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 Figure 3. The  $\epsilon$ -insensitive linear loss function is schematically shown in Figure 4.

#### Dual optimization problem and computing matrix $\mathbf{w}$

To consider constraints of the convex optimization problem in equation (8), the Lagrangian function is calculated to change the convex optimization problem to a dual optimization problem (also called the primal problem) as<sup>61</sup>



**Figure 4.**  $\epsilon$ -insensitive loss function with slack variable based on Vapnik.<sup>74</sup>

$$\begin{aligned} L = & \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^N (\zeta_i^- + \zeta_i^+) \\ & - \sum_{i=1}^N \alpha_i^+ (-\mathbf{z}_i + \mathbf{y}_i + \epsilon + \zeta_i^+) - \sum_{i=1}^N \mu_i^+ \zeta_i^+ \\ & - \sum_{i=1}^N \alpha_i^- (\mathbf{z}_i - \mathbf{y}_i + \epsilon + \zeta_i^-) - \sum_{i=1}^N \mu_i^- \zeta_i^- \end{aligned} \quad (9)$$

where  $\alpha_i^+$ ,  $\alpha_i^-$ ,  $\mu_i^+$ , and  $\mu_i^-$  are the Lagrangian multipliers and  $\alpha_i^+, \alpha_i^-, \mu_i^+, \mu_i^- \geq 0$ . Based on the Saddle point 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<sup>61</sup>

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

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

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

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

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

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

The dual optimization problem is obtained by substituting equations (10a)–(10d) into equation (9) as

$$\begin{aligned} \text{Minimize : } L = & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i^+ - \alpha_i^-) (\alpha_j^+ - \alpha_j^-) \mathbf{u}_i^T \mathbf{u}_j \\ & - \sum_{i=1}^N (\alpha_i^+ - \alpha_i^-) \mathbf{z}_i + \epsilon \sum_{i=1}^N (\alpha_i^+ + \alpha_i^-) \\ \text{Subject to : } & \begin{cases} \sum_{i=1}^N (\alpha_i^+ - \alpha_i^-) = 0 \\ 0 \leq \alpha_i^+ \leq C \\ 0 \leq \alpha_i^- \leq C \end{cases} \end{aligned} \quad (12)$$

Equation (12) can be rewritten in a standard quadratic programming (QP) form<sup>75</sup>

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

where

$$\begin{aligned} \alpha &= \begin{bmatrix} \alpha^+ \\ \alpha^- \end{bmatrix}, H = \begin{bmatrix} H & -H \\ -H & H \end{bmatrix}, f = \begin{bmatrix} -\mathbf{z}_i + \epsilon \\ \mathbf{z}_i + \epsilon \end{bmatrix} \\ H &= [\mathbf{u}_i^T \mathbf{u}_j], A_{eq} = [1 \dots 1 - 1 \dots - 1], B_{eq} = [0] \end{aligned} \quad (14)$$

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

### Karush–Kuhn–Tucker conditions and computing vector $\mathbf{b}$

Based on Karush–Kuhn–Tucker (KKT) approach, the following equations must be fulfilled at the optimum point<sup>76,77</sup>

$$\alpha_i^+ (-\mathbf{z}_i + \mathbf{y}_i + \epsilon + \zeta_i^+) = 0 \quad (15a)$$

$$\alpha_i^- (\mathbf{z}_i - \mathbf{y}_i + \epsilon + \zeta_i^-) = 0 \quad (15b)$$

$$\mu_i^+ \zeta_i^+ = (C - \alpha_i^+) \zeta_i^+ \quad (15c)$$

$$\mu_i^- \zeta_i^- = (C - \alpha_i^-) \zeta_i^- \quad (15d)$$

Considering equations (15a)–(15d), only the following five cases are possible

$$\alpha_i^+ = \alpha_i^- = 0 \quad (16a)$$

$$0 < \alpha_i^+ < C, \quad \alpha_i^- = 0 \quad (16b)$$

$$0 < \alpha_i^- < C, \quad \alpha_i^+ = 0 \quad (16c)$$

$$\alpha_i^+ = C, \quad \alpha_i^- = 0 \quad (16d)$$

$$\alpha_i^- = C, \quad \alpha_i^+ = 0 \quad (16e)$$

For  $|\mathbf{z}_i - \mathbf{y}_i|$  to be exactly equal to  $\epsilon$ , the only equations (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 Figure 3). Hence, the support vector domain,  $S$ , is calculated as

$$S = \{i | 0 < \alpha_i^- + \alpha_i^+ < C\} \quad (17)$$

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  $\mathbf{y}_i + \text{sign}(\alpha_i^+ - \alpha_i^-) \epsilon$  ( $i \in S$ ). As a result,  $\mathbf{b}$  is calculated as

$$\mathbf{b} = \frac{1}{|S|} \sum_{i \in S} (\mathbf{z}_i - \mathbf{w}^T \mathbf{u}_i - \text{sign}(\alpha_i^+ - \alpha_i^-) \epsilon) \quad (18)$$

In summary, the convex problem (equation (8)) is changed to the dual problem (equation (12)). Then, by solving the QP, equation (13), and substituting it into

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

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

the support vector expansion, equation (10a),  $\mathbf{w}$  is calculated. Then, vector  $\mathbf{b}$  is calculated using equation (18) (KKT conditions). Finally, by substituting  $\mathbf{w}$  and  $\mathbf{b}$  into equation (1), the prediction model of a given data set ( $\{\mathbf{u}_i, \mathbf{z}_i\}$ ) is found as

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

In this study,  $\mathbf{y}(\mathbf{u})$  is used to predict steady-state diesel engine  $\text{NO}_x$  emission and BMEP. This function is used to predict steady-state behavior of engine and will be denoted as  $\mathbf{y}_{ss}(\mathbf{u})$  in the subsequent sections.

### FOM

The diesel engine model consists of 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  $\text{NO}_x$  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 repetition formula  $((n+r-1)!/r!(n-1)!)$ , where  $n$  is the number of original features (in our case,  $n=3$ ), and  $r$  is the order of interactions.<sup>78</sup> 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 is listed in Table 2.

The total number of experimental points used for training is 62. To simultaneously minimize the model bias and to avoid overfitting, orders 1–4 of the original inputs and their interactions are considered as the base FOM model (34 features) to predict the steady-state values of  $\text{NO}_x$  and BMEP. The FOM features are listed in Table 3.



**Table 3.** Features of the full-order model (FOM) of NO<sub>x</sub> and BMEP.

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

The feature vector,  $U_j$ , is defined using Table 3 as

$$U_j = \{u_i\}_j \quad i = 1, \dots, n, \quad j = 1, \dots, 34 \quad (20)$$

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.<sup>79</sup> Particularly, for SVMs, the training time can be significantly reduced by normalizing the features.<sup>80</sup> Here the rescaling or also called min-max normalization method is used to normalize feature for the SVM

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

The system output vector is defined as

$$Z = \{z_i\} = [NO_{x,i} \quad BMEP_i]^T \quad i = 1, \dots, n \quad (22)$$

Then, the predicted steady-state NO<sub>x</sub> and BMEP are

$$y_{ss} = [NO_{x,ss} \quad BMEP_{ss}]^T \quad (23)$$

By solving the SVM algorithm for a given training data set,  $\{\bar{U}_j, Z\}$ , where  $\bar{U}_j$  and  $Z$  are calculated from equations (20) and (22), respectively, the approximate function,  $y_{ss}$ , is obtained to predict the steady-state values of NO<sub>x</sub> 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<sub>x</sub> 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 is analyzed. The proposed cost function to find  $C$  is defined as

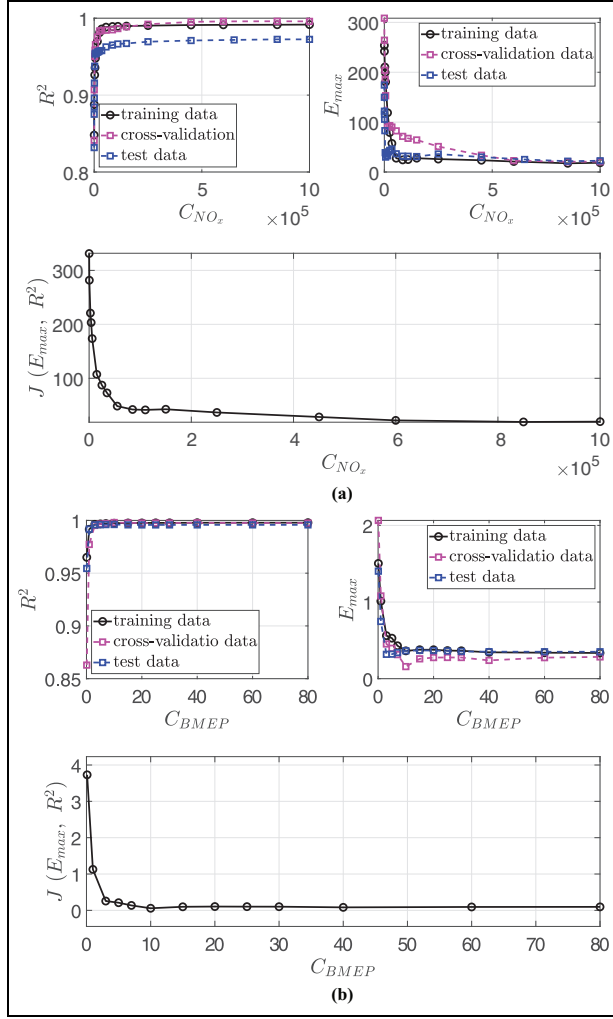
$$J(E_{max}, R^2) = \sqrt{\frac{E_{max,tr} E_{max,ts}}{R_{tr}^2 R_{ts}^2}} \quad (24)$$

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 the squared correlation coefficients 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 equation (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<sub>x</sub> and BMEP model are shown in Figure 5. The regulatory parameter,  $C$ , is a trade-off between the model flatness and the tolerated error. Based on the results shown in Figure 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 decreases 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, that is, the overfitting probability has increased. As a result, the model is less robust for new test data due to possible overfitting. Therefore, by setting  $C = C_o$ , the model performance is maximized while overfitting constraints are fulfilled. To ensure that 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 Figure 5,  $C_o$  for NO<sub>x</sub> and BMEP is selected to be  $C_{o,NO_x} = 85,000$  and  $C_{o,BMEP} = 60$ , respectively. The prediction versus the actual value for FOM NO<sub>x</sub> and BMEP is shown in Figure 6. Here, the cross-validation portion of training data is shown for both NO<sub>x</sub> and BMEP; however, to reduce the complexity of figures, for the rest of the article, combined cross-validation and training data are illustrated as training data. It should be noted that the regulatory parameter remains constant throughout the MOR process.

## MOR

Next, using the proposed FOM and by solving the SVM algorithm for regression, matrices  $w$  and  $b$  are obtained. The best  $C$  value for the SVM problem,  $C_o$ ,



**Figure 5.** Maximum error ( $E_{max}$ ), correlation coefficient ( $R^2$ ), and cost function ( $J(E_{max}, R^2)$ ) versus regulatory parameter  $C$ : (a)  $NO_x$  and (b) BMEP.

is found using different criteria. In this section, the 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 non-linear reduced control-oriented model (NRCOM) is found through the flowchart in Figure 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 algorithm for  $NO_x$  and BMEP, two types of models are proposed as follows:

1. 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 an  $NO_x$  sensor fault detection algorithm or VPs to evaluate a controller in simulation.

2. 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.<sup>81</sup>

As shown in Figure 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 MOR 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 MOR to the LOM. To avoid significant loss of the model accuracy, the least significant features are removed one by one until the relative difference between the cost functions  $J_m$  and  $J_{m-1}$  becomes more higher than the acceptable threshold. Then, the corresponding feature number is the number of features of the LOM,  $m_{LOM}$ . The relative difference between  $J_m$  and  $J_{m-1}$  is defined as

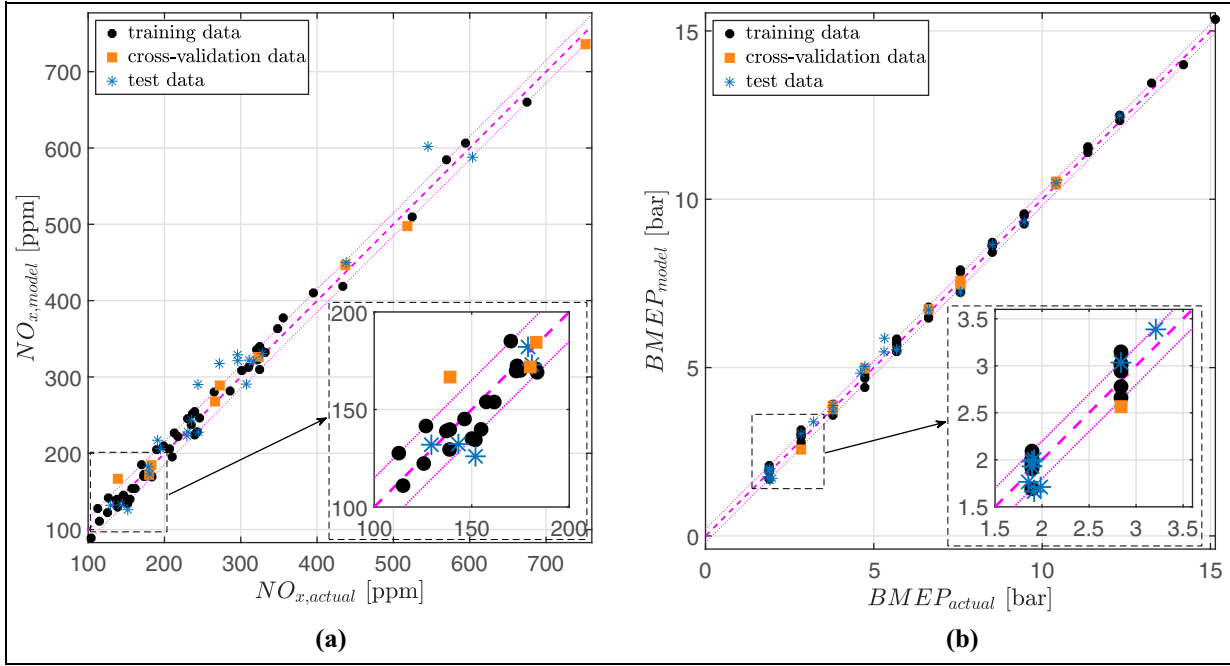
$$d_r(J_m, J_{m-1}) = \frac{|J_m - J_{m-1}|}{\max(J_m, J_{m-1})} \quad (25)$$

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

### $NO_x$ steady-state model

The squared correlation 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 Figure 8.

Based on Figure 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 are 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 Figure 8 by starting from  $m = 34$ , the first relative difference larger than 25% occurs for a nine-feature model ( $m_{LOM, NO_x} = 9$ ), the models with nine features are chosen as the LOM for  $NO_x$ , that is, the model with nine 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) is compared with an ANN using the same set of features. This



**Figure 6.** Prediction versus actual data for  $\text{NO}_x$  and BMEP full-order model (FOM): (a) prediction versus actual for  $\text{NO}_x$  FOM and (b) prediction versus actual for BMEP FOM.

provides a standard to compare these results to an 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.<sup>38</sup> The selection of hidden layer and neuron number was based on similar ANN-based studies in the literature. To make sure that the number of neurons is 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.<sup>22,38</sup> 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. In addition, the risk of overfitting is higher for ANN for the same size of training data.<sup>82</sup> 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  $\mathbf{b}$ .

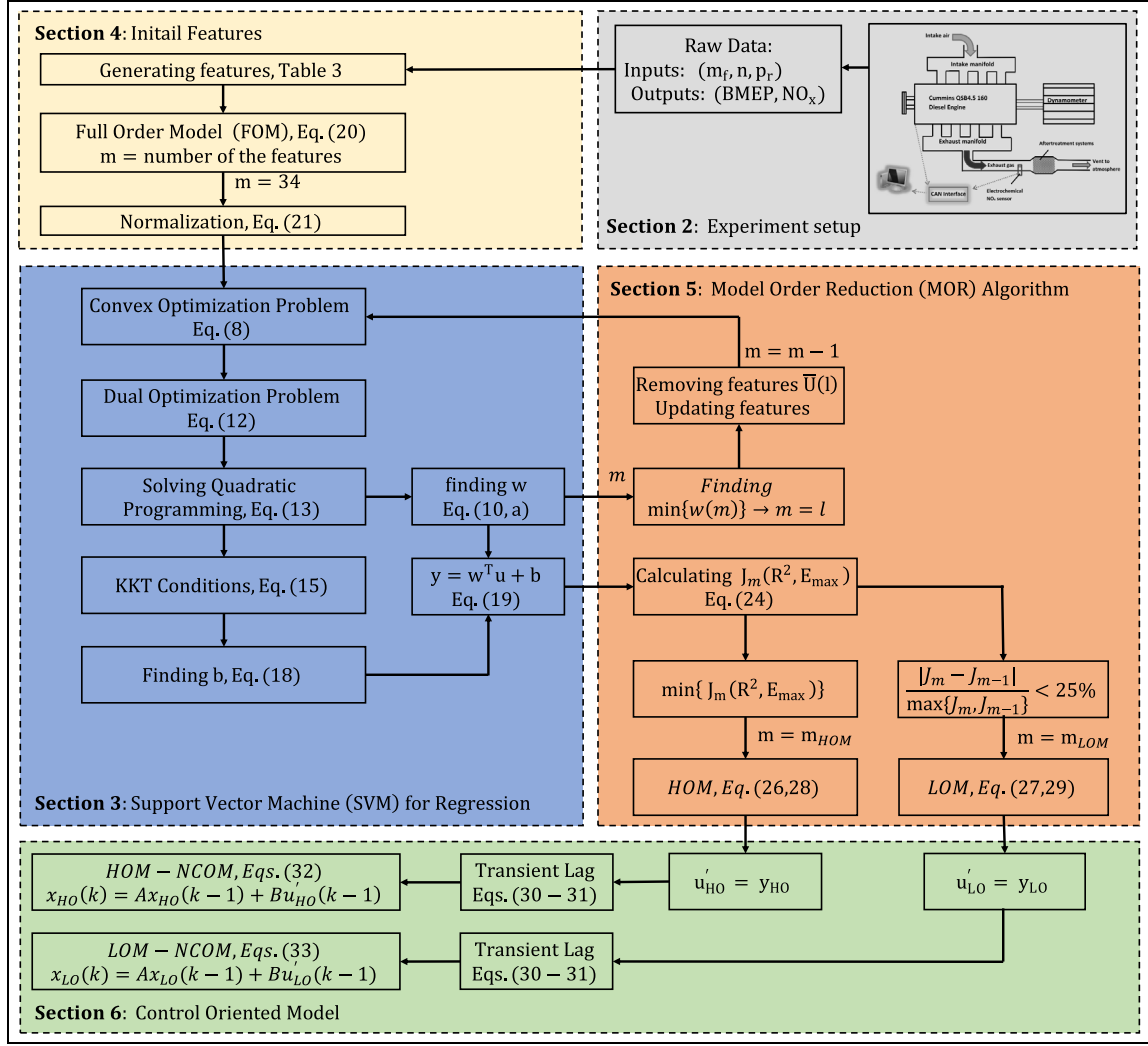
It should be noted that the performance of HOM is even better than FOM as a result of removing unnecessary features that affect the flatness of the SVM algorithm. Based on Table 4, the accuracy of LOM is acceptable where the error is above the defined threshold.

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

$$\bar{\mathbf{U}}_{\text{NO}_x, \text{HO}} = \bar{\mathbf{U}}_j \quad j = 1 - 9, 11 - 15, 17, 19 - 27, 29, 31 - 34 \quad (26)$$

$$\bar{\mathbf{U}}_{\text{NO}_x, \text{LO}} = \bar{\mathbf{U}}_j \quad j = 2, 5, 8, 15, 21, 22, 27, 32, 33 \quad (27)$$

where  $i$  is the data index, and  $j$  is the feature index. By solving the SVM algorithm for  $\text{NO}_x$ , the features of HOM and LOM are obtained. The equation of  $\text{NO}_{x, \text{HO}, \text{ss}}$  and  $\text{NO}_{x, \text{LO}, \text{ss}}$  is listed in Appendix 1, equations (40) and (41). The predicted steady-state  $\text{NO}_x$  versus the actual value for both the high-order and the low-order steady-state  $\text{NO}_x$  models is shown in Figure 9. Based on Figure 9(a), most of the test and the training data are within the defined tolerance  $\epsilon$  for HOM of  $\text{NO}_x$ . However, as shown in Figure 9(b), the accuracy of LOM is not consistent



**Figure 7.** Control-oriented model (COM) development and SVM-based MOR algorithm.

throughout all data points for both training and test points and item number of outliers is greater than the HOM.

### BMEP steady-state model

Similar to the  $\text{NO}_x$  steady-state model, the BMEP reduced steady-state model is obtained. The squared correlation coefficient ( $R^2$ ), the maximum error between prediction and actual data ( $E_{\max}$ ) for both the training and the test data, and the defined cost function ( $J(R^2, E_{\max})$ ) with respect to the number of features are shown in Figure 10. Based on Figure 10, a 20-feature model ( $m_{\text{HOM}, \text{BMEP}} = 20$ ) and a six-feature model ( $m_{\text{LOM}, \text{BMEP}} = 6$ ) 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  $\text{NO}_x$  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 six-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

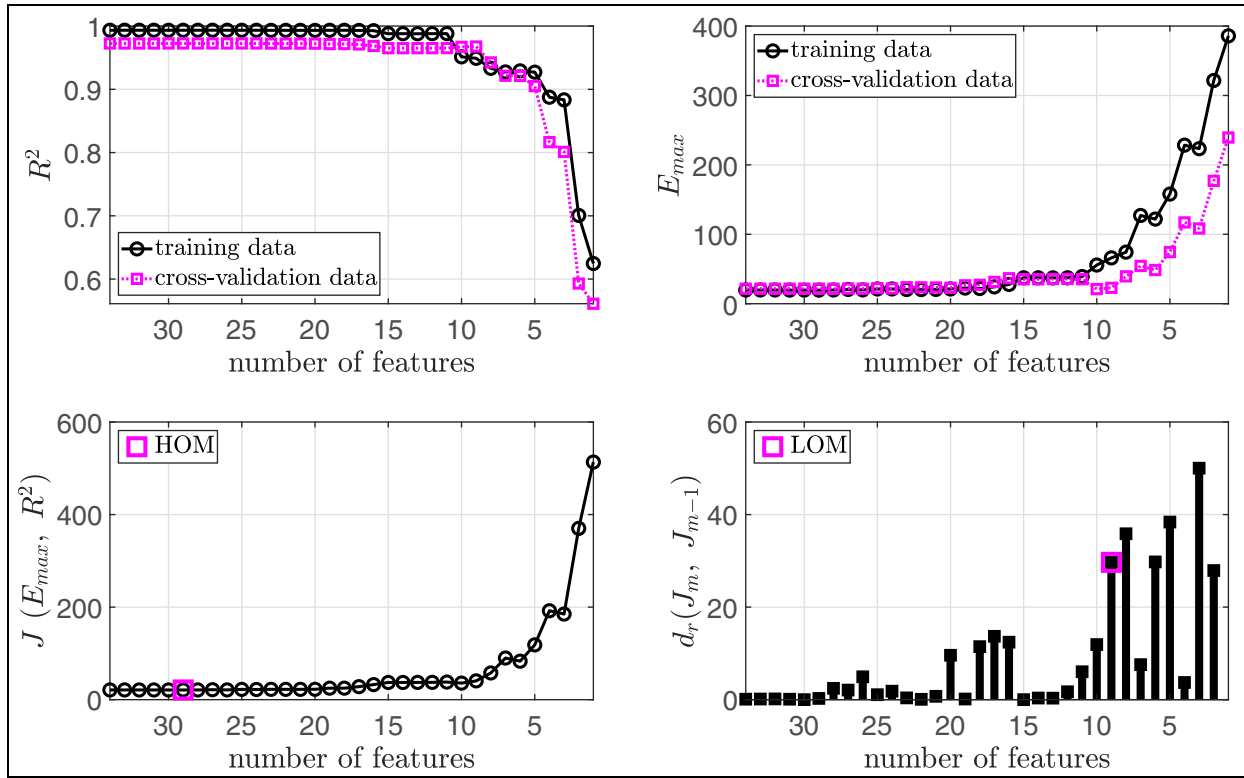
$$\bar{U}_{\text{BMEP}, \text{HO}} = \bar{U}_j$$

$$j = 1, 4, 7 - 10, 17 - 21, 24 - 27, 29 - 32, 34$$
(28)

$$\bar{U}_{\text{BMEP}, \text{LO}} = \bar{U}_j$$

$$j = 1, 18, 21, 25, 27, 30$$
(29)

where  $i$  is the 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  $\text{BMEP}_{\text{HO}, \text{ss}}$  and  $\text{BMEP}_{\text{LO}, \text{ss}}$  is listed in Appendix 1, equations (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 is shown in Figure 11. As shown in Table 5,



**Figure 8.** Maximum error ( $E_{max}$ ), squared correlation coefficient ( $R^2$ ), and cost function ( $J(E_{max}, R^2)$ ) versus the number of features of prediction function for steady-state  $\text{NO}_x$  prediction.

**Table 4.** Performance of the  $\text{NO}_x$  full-order model (FOM), high-order model (HOM), and low-order model (LOM).

| Model type              | FOM     |         | HOM     |         | LOM    |         |
|-------------------------|---------|---------|---------|---------|--------|---------|
| Number 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  | 0.9664  | 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   |

SVM: support vector machine; ANN: artificial neural network.

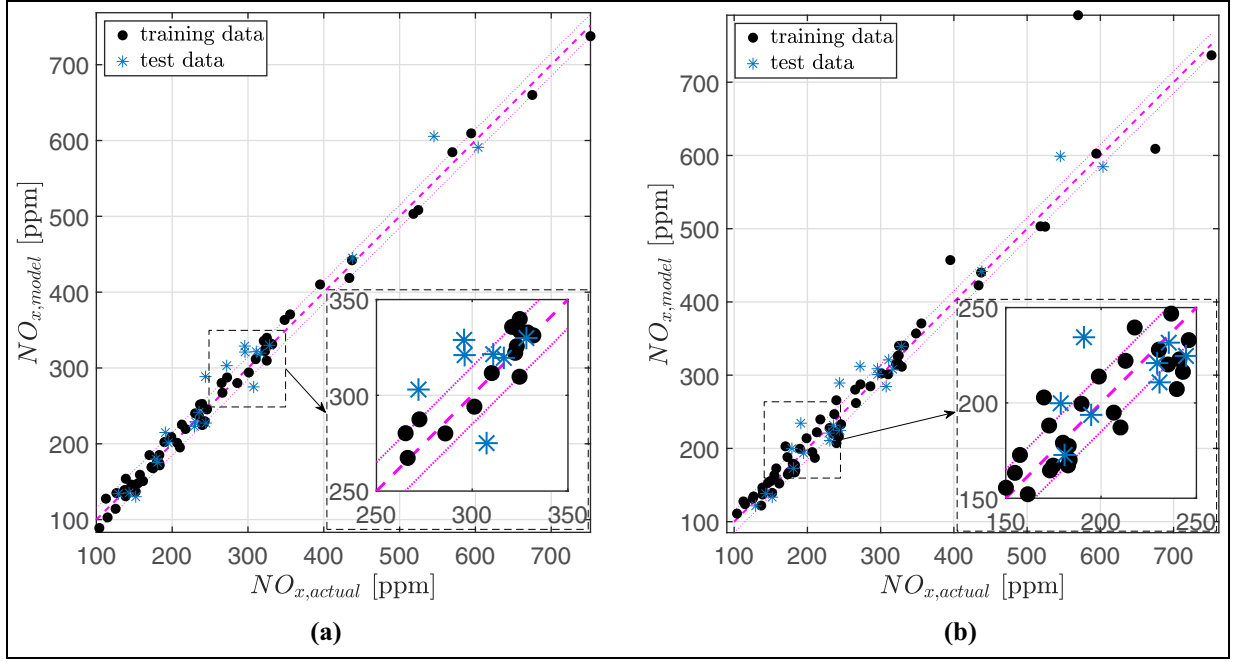
the HOM and LOM have an acceptable accuracy while the HOM has a higher accuracy than LOM. However, the LOM of BMEP has only six features, which makes it a simple model that requires a low computational effort. Most of the test and the training data for both HOM and LOM of BMEP are within the defined tolerance  $\epsilon$ , as shown in Figure 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 (Tables 4 and 5) is that by increasing the 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.<sup>83</sup> This trend appears in the HOM and LOM models.

### BMEP and $\text{NO}_x$ nonlinear control-oriented model

The model described above is used to determine steady-state  $\text{NO}_x$  and BMEP. Now a simple first-order



**Figure 9.** Prediction versus actual data for the low-order model (LOM) and the high-order model (HOM) of  $\text{NO}_x$ : (a) HOM of  $\text{NO}_x$  (with 29 features) and (b) LOM of  $\text{NO}_x$  (with nine features).

dynamic model for transient operation will be defined. To derive the discrete-time dynamic COM,  $\text{NO}_x$  concentration at step  $k$  for a sampling interval of  $T$  is calculated as follows

$$\text{NO}_x(k) = \left(1 - \frac{T}{\tau_{\text{NO}_x} + T}\right) \text{NO}_x(k-1) + \frac{T}{\tau_{\text{NO}_x} + T} \text{NO}_{x,ss}(k-1) \quad (30)$$

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

$$\text{BMEP}(k) = \left(1 - \frac{T}{\tau_{\text{BMEP}} + T}\right) \text{BMEP}(k-1) + \frac{T}{\tau_{\text{BMEP}} + T} \text{BMEP}_{ss}(k-1) \quad (31)$$

where  $\text{NO}_{x,ss}(k-1)$  and  $\text{BMEP}_{ss}(k-1)$  are the steady-state  $\text{NO}_x$  and BMEP calculated using equations (40) and (43). Hence,  $\tau$  is the sample interval and  $k$  is the sample time while  $\tau_{\text{NO}_x}$  and  $\tau_{\text{BMEP}}$  are the time constant parameters for  $\text{NO}_x$  and BMEP, respectively, which are estimated based on the experimental data and are found to be 1 and 0.2s for  $\text{NO}_x$  and BMEP, respectively.<sup>69</sup> The state space of the COM for both HOM and LOM can be defined as

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

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

where vector  $\mathbf{x}(\mathbf{k})$  contains two model states

$$\mathbf{x}_{\text{HO}}(\mathbf{k}) = [\text{NO}_{x,\text{HO}}(k) \quad \text{BMEP}_{\text{HO}}(k)]^T \quad (34)$$

$$\mathbf{x}_{\text{LO}}(\mathbf{k}) = [\text{NO}_{x,\text{LO}}(k) \quad \text{BMEP}_{\text{LO}}(k)]^T \quad (35)$$

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

$$\hat{\mathbf{u}}_{\text{HO}}(\mathbf{k}) = \begin{bmatrix} \text{NO}_{x,\text{HO},ss} \\ \text{BMEP}_{\text{HO},ss} \end{bmatrix} \quad (36)$$

$$\hat{\mathbf{u}}_{\text{LO}}(\mathbf{k}) = \begin{bmatrix} \text{NO}_{x,\text{LO},ss} \\ \text{BMEP}_{\text{LO},ss} \end{bmatrix} \quad (37)$$

where  $\text{NO}_{x,\text{HO},ss}$ ,  $\text{BMEP}_{\text{HO},ss}$ ,  $\text{NO}_{x,\text{LO},ss}$ , and  $\text{BMEP}_{\text{LO},ss}$  are listed in Appendix 1. The vector  $\mathbf{y}$  contains two model outputs

$$\mathbf{y}(\mathbf{k}) = [x_1(k) \quad x_2(k)] \quad (38)$$

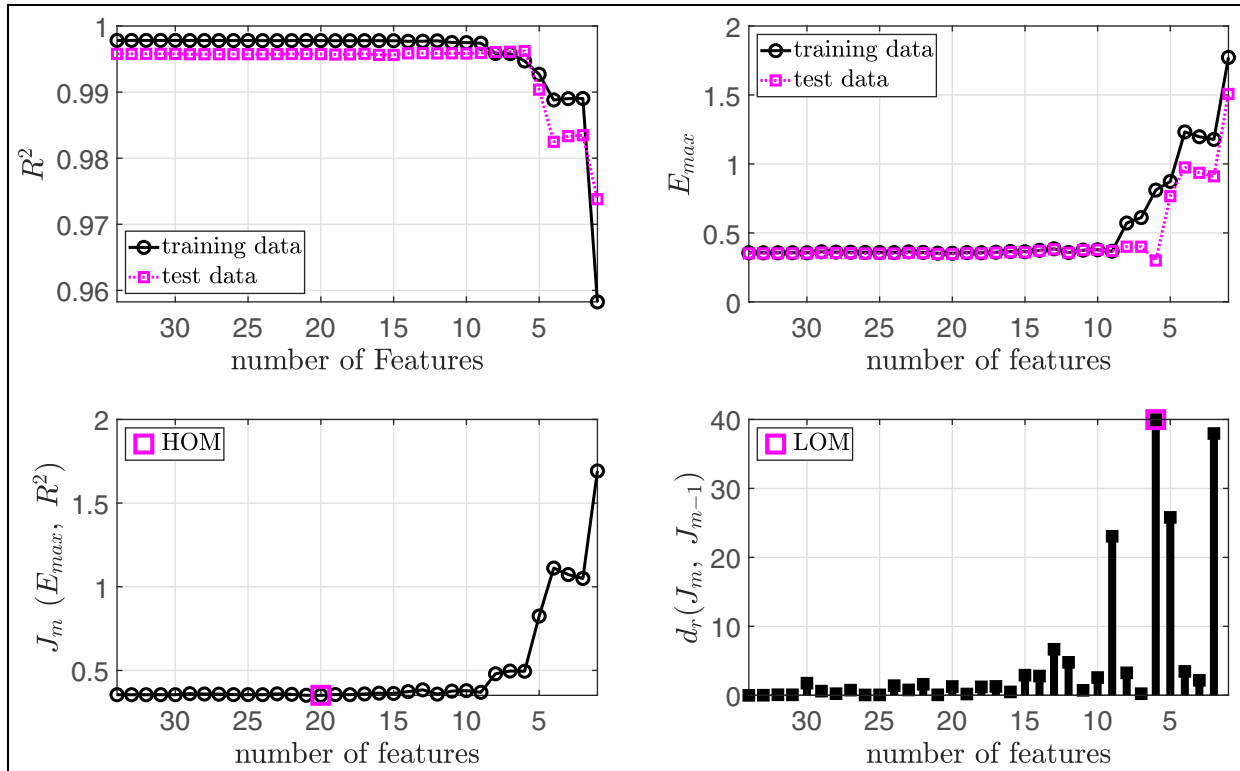
Matrices  $\mathbf{A}$  and  $\mathbf{B}$  are

$$\mathbf{A} = \begin{bmatrix} 1 - \frac{T}{\tau_{\text{NO}_x} + T} & 0 \\ 0 & 1 - \frac{T}{\tau_{\text{BMEP}} + T} \end{bmatrix} \quad (39)$$

$$\mathbf{B} = \begin{bmatrix} \frac{T}{\tau_{\text{NO}_x} + T} & 0 \\ 0 & \frac{T}{\tau_{\text{BMEP}} + T} \end{bmatrix}$$

Therefore, the HOM and LOM nonlinear control-oriented models (NCOMs) for  $\text{NO}_x$  and BMEP are obtained as equations (31) and (30). The open-loop response for HOM-NCOM and LOM-NCOM for  $\text{NO}_x$  and BMEP at four different engine speeds of 1250, 1500, 1750, and 2000 r/min is shown in Figures 12–15. In all cases,  $P_r$  and  $m_f$  are the system inputs that are applied to both HOM-NCOM and LOM-NCOM. In each transient test, the engine speeds remain constant (with a 10-r/min tolerance). In all of these plots, the open-loop response of both HOM-NCOM and LOM-NCOM based on the model versus actual measurements is shown. For BMEP, both the HOM-NCOM and LOM-NCOM follow the experimental closely as





**Figure 10.** Maximum error ( $E_{max}$ ), squared correlation coefficient ( $R^2$ ), and cost function ( $J(E_{max}, R^2)$ ) versus the 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    |        |
|-------------------------------|--------|--------|--------|--------|--------|--------|
| Number of features            | 34     |        | 20     |        | 6      |        |
| Training method               | SVM    | ANN    | SVM    | ANN    | SVM    | ANN    |
| $E_{max, tr}(\text{ppm})$     | 0.3560 | 0.4006 | 0.3526 | 0.3859 | 0.810  | 0.5435 |
| $E_{max, ts}(\text{ppm})$     | 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  |
| $J(E_{max}, R^2)(\text{ppm})$ | 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  |

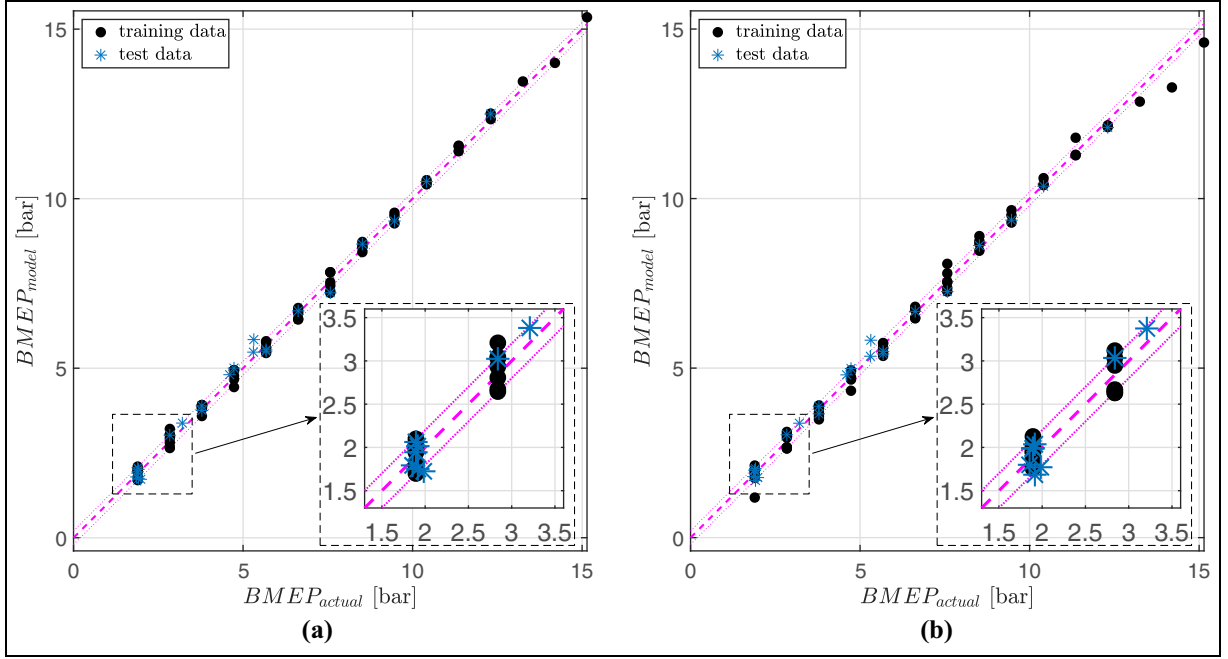
SVM: support vector machine; ANN: artificial neural network.

expected when looking at Figure 11. However,  $\text{NO}_x$  response for LOM-NCOM has different accuracies at different engine speeds. For instance, in Figures 13 and 15, the LOM-NCOM  $\text{NO}_x$  response is less accurate than the HOM model. Nonetheless,  $\text{NO}_x$  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, it is an accurate model for possible use as a VP to simulate the designed controller before implementation in a real-time system. In addition, it can be used as an accurate model for an  $\text{NO}_x$  sensor fault detecting algorithm. As the LOM-NCOM has a simple structure, it is quite suitable for designing a model-based robust controller such

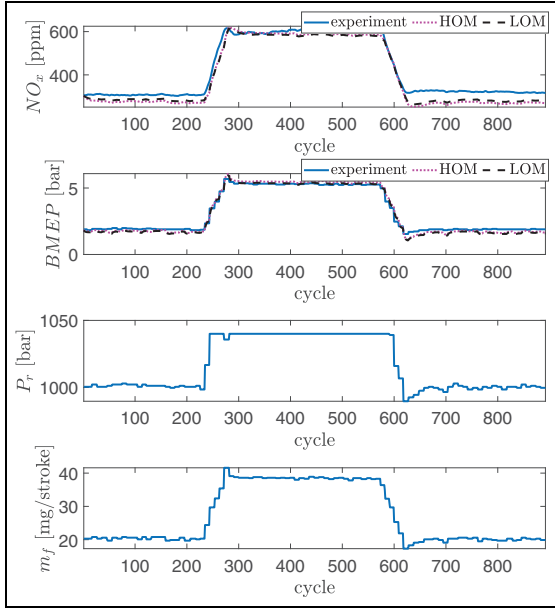
as sliding mode controller (SMC)<sup>84</sup> and is also capable of predicting samples ahead based on the current states and the inputs of the system. A robust controller can be used to overcome the model mismatch between LOM-NCOM and HOM-NCOM.

## Conclusion

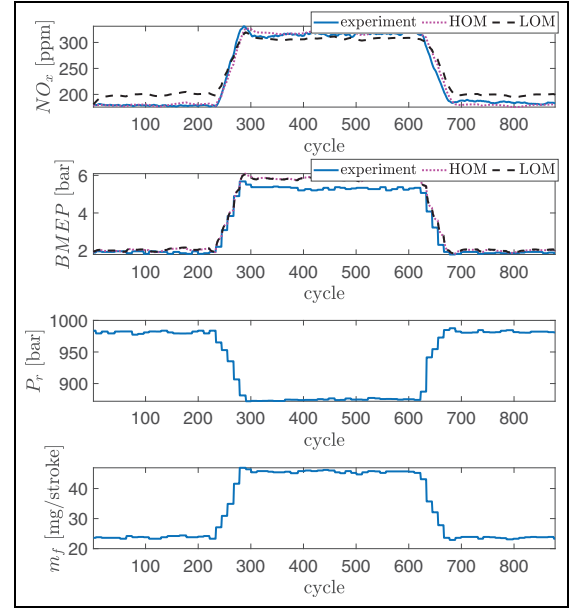
An MOR algorithm is developed using the SVM approach to predict the steady-state  $\text{NO}_x$  and BMEP of a medium-duty diesel engine. Based on the proposed SVM-based MOR algorithm and starting with a 34-feature FOM, an HOM and an LOM are developed to predict the steady-state  $\text{NO}_x$  emission and BMEP. The



**Figure 11.** Prediction versus actual data for high-order model (HOM) and low-order model (LOM) of BMEP: (a) HOM of BMEP (with 20 features) and (b) LOM of BMEP (with six features).



**Figure 12.** Transient response at engine speed of 1250 r/min.

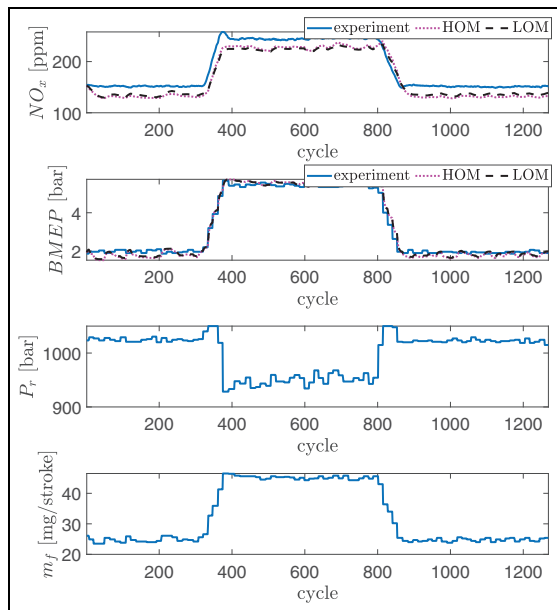


**Figure 13.** Transient response at engine speed of 1500 r/min.

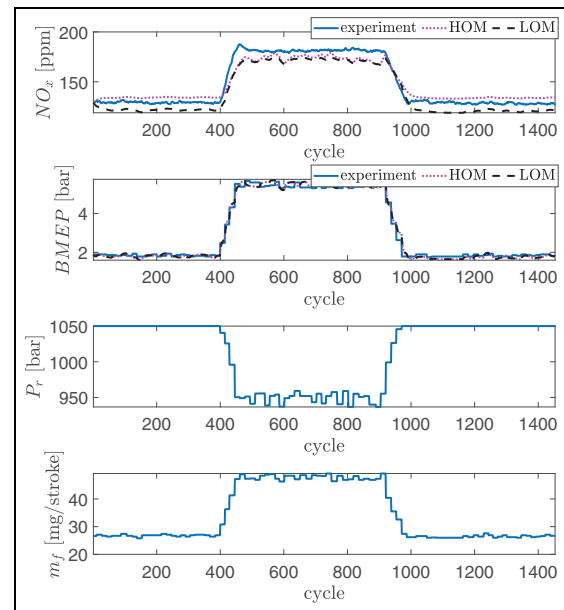
features of the models are calculated based on orders 1–4 of the main model inputs and the interactions of them. In this study, 84 engine operating points are considered, 74% of which is used to train the steady-state  $\text{NO}_x$  and BMEP, and 26% is used as test data. The model inputs are engine 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 for  $\text{NO}_x$  and 20 features for BMEP. For the steady-

state  $\text{NO}_x$  model, the squared correlation coefficient of test ( $R_{ts}^2$ ) is equal to 0.9724, 0.9725, and 0.9677 for FOM, HOM, and LOM, respectively. The  $R_{ts}^2$  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  $\text{NO}_x$  and BMEP is enhanced while the HOM complexity decreases 27.9% with respect to the FOM. The LOM model has an





**Figure 14.** Transient response at engine speed of 1750 r/min.



**Figure 15.** Transient response at engine speed of 2000 r/min.

acceptable accuracy with squared correlation coefficient of 0.9393 for  $\text{NO}_x$  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  $\text{NO}_x$  and BMEP are compared with an ANN, and the 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 is at least 5–14 times faster than the corresponding ANN models with the same set of features for  $\text{NO}_x$  and BMEP, respectively. In addition, the use of a linear kernel in the SVM makes it more suitable for real-time applications and for COMs.

Then, an NCOM is developed based on the developed SVM models to predict the transient behavior of the system. A fast response electrochemical  $\text{NO}_x$  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 different engine speeds for rising and falling step changes of the fuel rail pressure and the injected fuel amount for HOM. In addition, the LOM model has an accurate response at different speeds for BMEP; however, the  $\text{NO}_x$  prediction with LOM has varying accuracy at different engine speeds. It can be concluded that the HOM can predict  $\text{NO}_x$  and BMEP over a wide range of operating points, which makes it ideal to be used as a VP 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 an  $\text{NO}_x$  sensor fault detection algorithm based on this model will be investigated in future work.


### Declaration of conflicting interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was financially supported by the Natural Sciences Research Council of Canada (grant no. 2016-04646) and the Canada First Research Excellence Fund through Future Energy Systems.

### ORCID iDs

Armin Norouzi  <https://orcid.org/0000-0003-2690-0739>

Charles Robert Koch  <https://orcid.org/0000-0002-6094-5933>

### References

1. Vuorinen A. *Planning of optimal power systems*. Vammala: Vammalan Kirjapaino Oy, 2009.
2. Tutak W, Jamrozik A and Bereczky Lukacs ÁK. 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 and Chen SJ. Use of water containing acetone–butanol–ethanol for  $\text{NO}_x$ -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 and Roskilly AP. Particulate number and  $\text{NO}_x$  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; 8: 1152–1167.

6. Praveena V and Martin MLJ. A review on various after treatment techniques to reduce NO<sub>x</sub> emissions in a CI engine. *J Energy Inst* 2018; 91: 704–720.
7. Geng P, Tan Q, Zhang C, Wei L, He X, Cao E, et al. Experimental investigation on NO<sub>x</sub> and green house gas emissions from a marine auxiliary diesel engine using ultralow sulfur light fuel. *Sci Total Environ* 2016; 572(Suppl. C): 467–475.
8. Blanco-Rodriguez ID. *Modelling and observation of exhaust gas concentrations for diesel engine control*. Cham: Springer, 2014.
9. Koebel M, Elsener M and Kleemann M. Urea-SCR: a promising technique to reduce NO<sub>x</sub> emissions from automotive Diesel engines. *Catal Today* 2000; 59(34): 335–345.
10. Maiboom A, Tauzia X and 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 and 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. *Int J Engine Res* 2019; 20: 1101–1113.
13. Chavannavar P. Development and implementation of a mapless, model based SCR control system. *SAE Int J Engines* 2014; 7: 1113–1124.
14. Aliramezani M, Koch CR and Hayes RE. Estimating tail-pipe NO<sub>x</sub> concentration using a dynamic NO<sub>x</sub>/ammonia cross sensitivity model coupled to a three state control oriented SCR model. *IFAC-PapersOnLine* 2016; 49(11): 8–13.
15. Triantafyllopoulos G, Katsaounis D, Karamitros D, Ntziachristos L and Samaras Z. Experimental assessment of the potential to decrease diesel NO<sub>x</sub> emissions beyond minimum requirements for Euro 6 Real Drive Emissions (RDE) compliance. *Sci Total Environ* 2018; 618: 1400–1407.
16. Senda T and Harumi K. Prospects and challenges for the future of marine power systems. *Mar Eng* 2018; 53(3): 279–284.
17. Tang W, Siani A, Chen F and Chen B. On developing advanced catalysts systems to meet china new regulations. SAE technical paper 2019-01-0978, 2019.
18. Wong PK, Wong HC and Vong CM. Online time-sequence incremental and decremental least squares support vector machines for engine air-ratio prediction. *Int J Engine Res* 2012; 13(1): 28–40.
19. Aliramezani M, Koch CR and Patrick R. Phenomenological model of a solid electrolyte NO<sub>x</sub> and O<sub>2</sub> sensor using temperature perturbation for on-board diagnostics. *Solid State Ionics* 2018; 321: 62–68.
20. Ansari E, Menucci T, Shahbakhti M and 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. *Appl Energy* 2019; 239: 948–956.
21. Potenza R, Dunne J, Vulli S, Richardson D and King P. Multicylinder engine pressure reconstruction using NARX neural networks and crank kinematics. *Int J Engine Res* 2007; 8(6): 499–518.
22. He Y and Rutland C. Application of artificial neural networks in engine modelling. *Int J Engine Res* 2004; 5(4): 281–296.
23. Rezaei J, Shahbakhti M, Bahri B and Aziz AA. Performance prediction of HCCI engines with oxygenated fuels using artificial neural networks. *Appl 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 Int J Commer Veh* 2018; 11(5): 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*, Chicago, IL, 20–23 October 2019, vol. 59346, p.V001T06A007. New York: American Society of Mechanical Engineers.
26. Owoyele O and 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*, Chicago, IL, 20–23 October 2019. New York: American Society of Mechanical Engineers.
27. Xu Y, Guo R and Wang L. A twin multi-class classification support vector machine. *Cogn Comp* 2013; 5(4): 580–588.
28. Tanveer M. Robust and sparse linear programming twin support vector machines. *Cogn Comp* 2015; 7(1): 137–149.
29. Bertram AM and Kong SC. Computational optimization of a diesel engine calibration using a novel SVM-PSO method. SAE technical paper 2019-01-0542, 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. *Int J Engine Res*. Epub ahead of print 14 March 2019. DOI: 10.1177/1468087419833269.
31. Yusri IM, Majeed APPA, Mamat R, Ghazali MF, Awad OI and 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. *Renew Sustain Energ Rev* 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 and 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 and Koch CR. Support vector machine for a diesel engine performance and NO<sub>x</sub> emission control-oriented model. In: *21st IFAC World Congress, IFAC PapersOnLine*, Berlin, Germany, 11–17 July, 2020.

35. 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. *Int J Engine Res* 2020; 21(3): 421–436.
36. Duan H, Huang Y, Mehra RK, Song P and 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.
37. Liu B, Hu J, Yan F, Turkson RF and Lin F. A novel optimal support vector machine ensemble model for NOx emissions prediction of a diesel engine. *Measurement* 2016; 92: 183–192.
38. Niu X, Yang C, Wang H and Wang Y. Investigation of ANN and SVM based on limited samples for performance and emissions prediction of a CRDI-assisted marine diesel engine. *Appl Therm Eng* 2017; 111: 1353–1364.
39. Zuo Q, Zhu X, Liu Z, Zhang J, Wu G and Li Y. Prediction of the performance and emissions of a spark ignition engine fueled with butanol-gasoline blends based on support vector regression. *Environ Prog Sustain Energy* 2019; 38(3): e13042.
40. Azzam M, Awad M and Zeaiter J. Application of evolutionary neural networks and support vector machines to model NOx emissions from gas turbines. *J Environ Chem Eng* 2018; 6(1): 1044–1052.
41. Zhou H, Zhao JP, Zheng LG, Wang CL and Cen KF. Modeling NOx emissions from coal-fired utility boilers using support vector regression with ant colony optimization. *Eng Appl Artif Intel* 2012; 25(1): 147–158.
42. Lv Y, Liu J, Yang T and 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.
43. Lv Y, Yang T and Liu J. An adaptive least squares support vector machine model with a novel update for NOx emission prediction. *Chemometr Intell Lab Syst* 2015; 145: 103–113.
44. Wong KI, Wong PK, Cheung CS and Vong CM. Modeling and optimization of biodiesel engine performance using advanced machine learning methods. *Energy* 2013; 55: 519–528.
45. Ye J and Xiong T. SVM versus least squares SVM. In: *Artificial intelligence and statistics*, 2007, pp.644–651, <http://proceedings.mlr.press/v2/ye07a/ye07a.pdf>
46. Dietterich T. Overfitting and undercomputing in machine learning. *ACM Comp Surv* 1995; 27(3): 326–327.
47. Brown G, Pocock A, Zhao MJ and Luján M. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. *J Mach Learn Res* 2012; 13: 27–66.
48. Peng H, Long F and Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Patt Anal Mach Intel* 2005(8): 1226–1238.
49. Nguyen XV, Chan J, Romano S and 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*, 24–27 August 2014, pp.512–521. New York: ACM, <https://dl.acm.org/doi/pdf/10.1145/2623330.2623611>
50. Yang HH and Moody J. Data visualization and feature selection: new algorithms for nongaussian data. In: *Advances in neural information processing systems*, 2000, pp.687–693. New York: ACM, <https://papers.nips.cc/paper/1779-data-visualization-and-feature-selection-new-algorithms-for-nongaussian-data.pdf>
51. Hall MA. *Correlation-based feature selection for machine learning*. PhD Thesis, The University of Waikato, Hamilton, New Zealand, 1999.
52. He W, Wang Z and Jiang H. Model optimizing and feature selecting for support vector regression in time series forecasting. *Neurocomputing* 2008; 72(1–3): 600–611.
53. Tao X, Renmu H, Peng W and 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*, Hong Kong, China, 5–8 April 2004. pp.510–514, vol. 2. New York: IEEE.
54. Hira ZM and Gillies DF. A review of feature selection and feature extraction methods applied on microarray data. *Adv Bioinformatics* 2015; 2015: 198363.
55. Schölkopf B, Smola A and Müller KR. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation* 1998; 10(5): 1299–1319.
56. Salah Mohamed K. *Machine learning for model order reduction*. Cham: Springer, 2018.
57. Wang JM, Chu CC, Yu Q and Kuh ES. On projection-based algorithms for model-order reduction of interconnects. *IEEE Trans. Circuit Syst I* 2002; 49(11): 1563–1585.
58. Alam M, Nieuwoudt A and Massoud Y. Wavelet-based passivity preserving model order reduction for wideband interconnect characterization. In: *8th international symposium on quality electronic design (ISQED'07)*, San Jose, CA, 26–28 March 2007, pp.432–437. New York: IEEE.
59. Barrasso D, Tamrakar A and Ramachandran R. A reduced order PBM-ANN model of a multi-scale PBM-ANN description of a wet granulation process. *Chem Eng Sci* 2014; 119: 319–329.
60. Parmar G, Mukherjee S and Rasad R. Reduced order modelling of linear MIMO systems using genetic algorithm. *Int J Simulat Model* 2007; 6(3): 173–184.
61. Smola AJ and Schölkopf B. A tutorial on support vector regression. *Statist Comp* 2004; 14(3): 199–222.
62. Kihás D, Pachner D, Baramov L, Uchanski M, Naik P and Khaled N. Concept analysis and initial results of engine-out NOx estimator suitable for on ECM implementation. SAE technical paper 2016-01-0611, 2016.
63. Quérel C, Grondin O and Letellier C. Semi-physical mean-value NOx model for diesel engine control. *Cont Eng Prac* 2015; 40: 27–44.
64. Guardiola C, Martín J, Pla B and Bares P. Cycle by cycle NOx model for diesel engine control. *Appl Ther Eng* 2017; 110: 1011–1020.
65. Rezaei R, Hayduk C, Alkan E, Kemski T, Delebinski T and Bertram C. Hybrid phenomenological and mathematical-based modeling approach for diesel emission prediction. SAE technical paper 2020-01-0660, 2020.
66. Irdmoussa BK, Rizvi SZ, Veini JM, Nabert J and Shahbakhti M. Data-driven modeling and predictive control of combustion phasing for RCCI engines. In: *2019 American control conference (ACC)*, Philadelphia, PA, 10–12 July 2019, pp.1617–1622. New York: IEEE.

67. Oğuz H, Saritas I and Baydan HE. Prediction of diesel engine performance using biofuels with artificial neural network. *Expert Syst Appl* 2010; 37(9): 6579–6586.
68. Tschanz F, Amstutz A, Onder CH and Guzzella L. Feed-back control of particulate matter and nitrogen oxide emissions in diesel engines. *Cont Eng Prac* 2013; 21(12): 1809–1820.
69. Aliramezani M, Norouzi A, Koch CR and Hayes RE. A control oriented diesel engine NOx emission model for on board diagnostics and engine control with sensor feed-back. In: *Proceedings of combustion institute-Canadian section (CICS)*, Kelowna, BC, Canada, 13–16 May 2019, pp.1–6, <https://www.researchgate.net/publication/333223847>
70. Vapnik V and Lerner A. Generalized portrait method for pattern recognition. *Automat Rem Contr* 1963; 24(6): 774–780.
71. Vapnik V and Chervonenkis A. A note on class of perceptron. *Automat Rem Contr* 1964; 25: 103–109.
72. Drucker H, Burges CJ, Kaufman L, Smola AJ and Vapnik V. Support vector regression machines. In: *Advances in neural information processing systems*, 1997, pp.155–161, <http://papers.nips.cc/paper/1238-support-vector-regression-machines.pdf>
73. Cortes C and Vapnik V. Support-vector networks. *Mach Learn* 1995; 20(3): 273–297.
74. Vapnik V. *The nature of statistical learning theory*. Berlin: Springer Science & Business Media, 2013.
75. Bellman R. The theory of dynamic programming. *B Am Math Soc* 1954; 60(6): 503–515.
76. Karush W. *Minima of functions of several variables with inequalities as side constraints*. MSc Dissertation, Department of Mathematics, University of Chicago, Chicago, IL, 1939.
77. Kuhn HW and Tucker AW. Nonlinear programming. In: Neyman J (ed.) *Proceedings of the second Berkeley symposium on mathematical statistics and probability*. Berkeley, CA: University of California Press, 1951, pp.481–492.
78. Brualdi RA. *Introductory combinatorics*. Delhi, India: Pearson Education, 1977.
79. Ioffe S and Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* 2015, <https://arxiv.org/abs/1502.03167>
80. Juszczak P, Tax D and Duin RP. Feature scaling in support vector data description. In: *Proceedings of the 8th Annual Conference of the Advanced School for Computing and Imaging (ASCI)*, Delft, The Netherlands, June 2002, pp.170–177.
81. Norouzi A, Aliramezani M and 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)*, Kelowna, BC, Canada, 13–16 May 2019, pp.7–12, <https://www.researchgate.net/publication/333224129>
82. Ren J. and ANN vs. SVM: which one performs better in classification of MCCs in mammogram imaging. *Knowl Based Syst* 2012; 26: 144–153.
83. Shalev-Shwartz S and Srebro N. SVM optimization: inverse dependence on training set size. In: *Proceedings of the 25th international conference on machine learning*, 2008, pp.928–935. New York: ACM, <http://icml2008.cs.helsinki.fi/papers/266.pdf>
84. Norouzi A, Ebrahimi K and 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.

## Appendix I

### Steady-state NO<sub>x</sub> model

$$\begin{aligned}
 \text{NO}_{x, \text{HO}, \text{ss}} &= \mathbf{w}_{\text{NO}_x}^T \bar{\mathbf{U}}_{\text{NO}_x, \text{HO}} + b_{\text{NO}_x, \text{HO}} \\
 &= 528.93\bar{\mathbf{U}}_1 - 2411.93\bar{\mathbf{U}}_2 + 1544.89\bar{\mathbf{U}}_3 - 1177.86\bar{\mathbf{U}}_4 + 3654.47\bar{\mathbf{U}}_5 - 1495.65\bar{\mathbf{U}}_6 \\
 &\quad - 836.09\bar{\mathbf{U}}_7 + 1123.75\bar{\mathbf{U}}_8 - 512.52\bar{\mathbf{U}}_9 - 1230.75\bar{\mathbf{U}}_{11} - 1319.38\bar{\mathbf{U}}_{12} + 2846.76\bar{\mathbf{U}}_{13} \\
 &\quad - 12.65\bar{\mathbf{U}}_{14} + 1574.67\bar{\mathbf{U}}_{15} + 529.47\bar{\mathbf{U}}_{17} - 2329.74\bar{\mathbf{U}}_{19} + 1214.66\bar{\mathbf{U}}_{20} - 573.96\bar{\mathbf{U}}_{21} \\
 &\quad + 1076.88\bar{\mathbf{U}}_{22} - 1129.99\bar{\mathbf{U}}_{23} - 2012.74\bar{\mathbf{U}}_{24} - 1001.16\bar{\mathbf{U}}_{25} + 815.56\bar{\mathbf{U}}_{26} + 1440.14\bar{\mathbf{U}}_{27} \\
 &\quad - 1571.92\bar{\mathbf{U}}_{29} + 789.89\bar{\mathbf{U}}_{31} - 2795.74\bar{\mathbf{U}}_{32} + 1602.41\bar{\mathbf{U}}_{33} + 2640.48\bar{\mathbf{U}}_{34} + 782.81
 \end{aligned} \tag{40}$$

$$\begin{aligned}
 \text{NO}_{x, \text{LO}, \text{ss}} &= \mathbf{w}_{\text{NO}_x}^T \bar{\mathbf{U}}_{\text{NO}_x, \text{LO}} + b_{\text{NO}_x, \text{LO}} \\
 &= -1654.78\bar{\mathbf{U}}_2 + 1082.55\bar{\mathbf{U}}_5 + 269.24\bar{\mathbf{U}}_8 + 1682.28\bar{\mathbf{U}}_{15} - 946.71\bar{\mathbf{U}}_{21} \\
 &\quad - 320.70\bar{\mathbf{U}}_{22} \\
 &\quad + 2549.50\bar{\mathbf{U}}_{27} - 4713.23\bar{\mathbf{U}}_{32} + 2608.52n^2\bar{\mathbf{U}}_{33} + 1017.98
 \end{aligned} \tag{41}$$

### Steady-state BMEP model

$$\begin{aligned}
 \text{BMEP}_{\text{HO}, \text{ss}} &= \mathbf{w}_{\text{BMEP}}^T \bar{\mathbf{U}}_{\text{BMEP}, \text{HO}} + b_{\text{BMEP}, \text{HO}} \\
 &= +3.03\bar{\mathbf{U}}_1 + 1.41\bar{\mathbf{U}}_4 + 1.57\bar{\mathbf{U}}_7 + 2.04\bar{\mathbf{U}}_8 - 2.05\bar{\mathbf{U}}_9 + 0.78\bar{\mathbf{U}}_{10} + 1.47\bar{\mathbf{U}}_{117} \\
 &\quad - 2.03\bar{\mathbf{U}}_{18} + 1.05\bar{\mathbf{U}}_{19} + 0.93\bar{\mathbf{U}}_{20} - 1.16\bar{\mathbf{U}}_{21} - 2.18\bar{\mathbf{U}}_{24} + 1.66\bar{\mathbf{U}}_{25} - 1.78\bar{\mathbf{U}}_{26} \\
 &\quad + 1.49\bar{\mathbf{U}}_{27} - 0.75\bar{\mathbf{U}}_{29} - 2.12\bar{\mathbf{U}}_{30} + 1.35\bar{\mathbf{U}}_{31} + 0.83\bar{\mathbf{U}}_{32} - 0.81\bar{\mathbf{U}}_{34} + 6.52
 \end{aligned} \tag{42}$$

$$\begin{aligned}
 \text{BMEP}_{\text{LO}, \text{ss}} &= \mathbf{w}_{\text{BMEP}}^T \bar{\mathbf{U}}_{\text{BMEP}, \text{LO}} + b_{\text{BMEP}, \text{LO}} \\
 &= +6.36\bar{\mathbf{U}}_1 - 3.54\bar{\mathbf{U}}_{18} - 5.00\bar{\mathbf{U}}_{21} + 6.10\bar{\mathbf{U}}_{25} + 3.89\bar{\mathbf{U}}_{27} - 2.95\bar{\mathbf{U}}_{30} + 6.62
 \end{aligned} \tag{43}$$