SCR-Filter Model Order Reduction (2): Proper Orthogonal Decomposition and Artificial Neural Network

Seun Olowojebutu¹ · Thomas Steffen¹ · Phillip Bush²

Received: 14 May 2019 / Revised: 18 July 2020 / Accepted: 1 August 2020 / Published online: 27 August 2020

© The Author(s) 2020

Abstract

Catalysed diesel particulate filters (DPF) have been described as multifunctional reactor systems. Integration of selective catalytic reduction (SCR) functionality in the DPF creates an SCR-in-DPF system that achieves nitrous oxides (NO_x) treatment along with particulate matter (PM) collection. The physical and chemical aspects of the integrated SCR-filter complicate system modelling. The goal of this work is to develop low-complexity model of the SCR-filter system which retains high fidelity. A high-fidelity model of the SCR-coated filter has been developed and validated. The performance of the model was described in a previous paper. Model complexity reduction is attempted in this paper. The objective is to achieve simulation times that can support the deployment of the model for online system control in an engine control unit. Two approaches were taken for the SCR-coated filter model order reduction (MOR): a “grey-box” approach via proper orthogonal decomposition (POD) and a “black box” approach via artificial neural network (ANN) function approximation. The POD method is shown to deliver a significant MOR while maintaining a high degree of fidelity but with less than 5% improvement in simulation time. The ANN method delivers a substantial MOR with reduction of three orders of magnitude in simulation time. The accuracy of the ANN model is satisfactory with good generalisation to new test data but noticeably inferior to the POD method.

Keywords NH₃ SCR-in-DPF · Reduced order modelling · Proper orthogonal decomposition · Machine learning · Neural networks function approximation · Diesel emission after-treatment

1 Introduction

Diesel engines offer superior performance in fuel economy compared with gasoline engines [1], but the simultaneous control of soot/particulates (PM) and nitrogen oxides (NO_x) is challenging. Integration of the selective catalytic reduction (SCR) catalyst within a diesel particulate filter (DPF) monolith is an emerging technology for simultaneous control of soot and NO_x emissions.

Modelling of the integrated SCR-filter¹ unit is complicated. The interaction of physical and chemical considerations

within different phases of the monolith material over different timescales make the effort of developing adequate representation of the SCR-filter system complicated. The application of the system model in real-time online model-based controls further motivates the need to develop simple but adequate representative models.

A model of the SCR-filter system has been developed and validated in [2]. The model was primarily validated against the published data in Schrade et al. [3]. The model describes the data well; details are provided in [2]. A new implementation of the model delivered faster than real-time solution times. However, the full-order model is not easily implemented for online control in an engine control unit (ECU). Therefore, model order reduction is attempted in this work.

Two model order reduction (MOR) techniques are implemented in this paper: (a) proper orthogonal decomposition (POD) and (b) artificial neural networks (ANN).

The POD method is a multi-variate statistical method that aims to obtain a compact representation of data. In the POD method, large-scale system data is decomposed into its characteristic modes (or eigenvectors). A Galerkin projection of

¹ SCR-coated filter also referred to as SCRf, SDPF, SCR-in-DPF

This is the second of two papers on the model order reduction of integrated SCR filter systems.

✉ Seun Olowojebutu
seun.olowojebutu@cantab.net

¹ Loughborough University, Loughborough, UK

² Eminox Ltd, Gainsborough, UK

the large-scale data on the subspace spanned by the largest (or dominant) eigenvectors can be used to derive a lower dimensional surrogate of the original large-scale system.

The motivation for using the POD approach is that the low-order model directly derives from the high-fidelity model (or experimental data), thereby retaining the physics of the system.

POD has been applied to engine research to study turbulence and cyclic variation of flow and combustion properties in internal combustion engines [4, 5]. The method is used extensively in model reduction in CFD applications [6]. The Galerkin technique has been applied in exhaust after-treatment system (EATS) modelling in the context of mean weighted residuals method of solution of differential equations [7, 8].

The POD approach was applied to the SCR-filter model order reduction in [9]. In that work, the SCR-filter model was not complete and had not been validated. This work improves on the presentation in [9] with the full-order model completed and validated. A change in the representation of diffusion transport in the model led to a different implementation of the SCR-coated filter model and the POD method resulting in different observations on the performance of the POD method in this work compared to [9].

The alternative approach to model order reduction applied in this work is ANN. In this approach, data was obtained from the full-order model developed in [2] and a black-box function approximation is derived to represent the model performance.

The motivation for this approach is twofold: (a) the ANN method is well suited for nonlinear model function approximation [10–12]; and (b) ANN can deliver model simplicity and adequate representation required for deploying the SCR-coated filter model in an online application.

Neural network has been applied in EATS modelling [13–19]. In [19], ANN was used to approximate the SCR catalytic converter to accelerate ammonia dosing optimisation. This is similar to our objective here, namely, to use a neural network to approximate the full-order SCR-filter model for deployment in a model-based control application.

The POD and ANN techniques were employed for MOR in this work. We found that the POD delivered a reduced order model with a higher degree of fidelity to the HFM but with limited improvement in simulation times. The ANN delivered a reduced order model which achieved a more significant reduction in simulation time—on the scale of three orders of magnitude—but with a poorer level of fidelity to the reference HFM.

A note on terminology: “MOR” in this work refers to the process of reducing model complexity for control-oriented application. The success of our MOR technique is assessed on two factors: fidelity and simulation efficiency. Model fidelity measures how close the reduced model is to the original full-order model. The root mean square error (RMSE) metric is used as an indicator of model fidelity. Simulation efficiency

measures the change (mostly improvement) in model simulation times.

The rest of the paper is organised as follows. The POD method is introduced in the next section with some background on the application of the POD method to SCR-filter modelling. This is followed by the results of the POD applied to a sample problem. The ANN method is then introduced. Some of the considerations taken in developing ANN for the SCR-filter model application are presented. The results of the ANN method applied to a sample problem are then presented, followed by some concluding remarks.

2 Proper Orthogonal Decomposition

POD, also known as principal component analysis (PCA), or Karhunen-Loève decomposition (KLD), is a statistical method which achieves model order reduction (MOR) by extracting the dominant modes of the system and using those modes to devise a lower dimensional approximation. The main idea of the POD method is to decompose system data obtained from experiment or numerical simulation into linear combination of basis functions (POD modes) and associated temporal coefficients. Model order reduction is then achieved by Galerkin projection of the full-order model on the subspace spanned by the lower-dimensional POD modes.

The motivation for using the POD approach is that the low-order model directly derives from the high-fidelity model (or experimental data), thereby retaining the physics of the system. The derivation of the POD modes from experimental or numerical data makes this model order reduction approach “grey-box”.

The mathematical background of the POD method is available in several excellent works [20–23]; a focused summary of the mathematical background as it applies to this system is presented in [9]. This section extends the approach from [9] for the 1D + 1D SCR-in-DPF model with more than two physical state variables.

2.1 The Intuitive Basis of the POD Method

Consider the solution to a simple equation to describe the transport of heat in a slab that has one end exposed to a heat source. The distribution of temperature within the slab is a series of exponential temperature profiles which progressively spreads throughout the entire slab provided the heat source is retained. The system dynamics is composed of a repeating pattern of temperature transport profile which spreads over time within the slab. The repeating patterns can be called “basis functions” [23].

The temporal variation of the physical phenomena (either due to its initial value or influence of source terms) can be captured in temporal coefficients which, along with the fixed

repeating patterns, can describe the underlying phenomena. Thus, it is possible to expand a composite variable $g(x, t)$ into a set of spatial basis functions and temporal coefficients (Eq. 1)

$$g(x, t) = \sum_{i=1}^k a_i(t) \phi_i(x) \quad (1)$$

The idea of decomposition of physical quantities in expansions of the form Eq. 1 was first proposed in the work of Joseph Fourier in his memoir *On the Propagation of Heat in Solid Bodies* written in 1807 [24]. There he proposed to expand an arbitrary function in a series of trigonometric basis functions. The well-known Fourier series are examples of Eq. 1, where the repeating patterns consist of trigonometric functions. The main difference, however, is that the Fourier patterns or Fourier “basis functions” are not derived from the data. The trigonometric functions are aimed to approximate any arbitrary functions. In POD, the basis functions are derived from the system data [23].

The intuition of the POD method is that for even the most complex physical phenomena, fundamental patterns/characteristic modes can be identified from the system data. The singular value decomposition (SVD) routine performs an operation similar to extracting the fundamental recurring patterns in a dataset.

The evolution of the specie concentration on a SCR filter system has repeating spatial components which are identified as the basis function from the SVD routine; the associated temporal coefficients track any transient component of the system. The combination of the spatial basis function and the temporal coefficient estimated from the POD-Galerkin method (Eq. 9) approximates the system data without much loss of fidelity. The aim of the POD method is to utilise only a few of the spatial basis of the SVD output to reconstruct the original dataset without much loss of fidelity.

2.2 POD Application to SCR-Filter Model

The first step in the application of the POD method is to obtain an ensemble data describing the system. The data can be obtained from experiments or from numerical simulation of the full-order model. The POD method in this work was applied using numerical data obtained from a validated high-fidelity model of the SCR-coated filter system [2]. For each of the cases considered, the numerical model is solved at the specified inlet conditions, to generate the system ensemble data for the POD basis. In this presentation, the numerical model solution is referred to as “high-fidelity model” or HFM, and the low-order POD-derived model is referred to as the “POD-model”.

The wall temperature, the ammonia surface coverage (ASC) and mass of soot retained in the system are selected as state variables on which to apply the POD method because

they are the explicit time-dependent variables in our system. (For ease of description, only POD basis associated with the ammonia surface coverage on catalyst site S_1 is described in the published results.) The remaining (quasi-steady state) equations are solved at each time step with the values of the state variables obtained from the POD model evolution.

Let $\theta, T_s, m_p \in \mathbb{R}^{N \times T}$ be the solution obtained from the solution of the HFM for ASC, wall temperature and soot mass retained, respectively, at N discrete units of the wall layer along the length of the catalyst, over the time period $[0, T]$. The mass of soot variable m_p is combined for deep-bed soot in the wall and soot on the cake layer. The POD modes associated with the ASC, wall temperature and soot mass data ensemble are extracted by SVD of the data. Let $\Phi_\theta, \Phi_{T_s}, \Phi_{m_p} \in \mathbb{R}^{N \times p}$ be the orthonormal matrices of the first p POD modes, where p is the number of modes need to capture a given level of variation energy from the data ensemble. The Galerkin projection of the subspace spanned by the POD modes on the HFM is then as follows:

Let $\hat{\theta} \in \mathbb{R}^{N \times T}$ be the approximate coverage θ obtained from the low-order POD matrix as:

$$\hat{\theta} = \Phi_\theta \mathbf{a}_\theta \quad (2)$$

and $\hat{T}_s$ be the approximate temperature T_s obtained from the low-order POD matrix as

$$\hat{T}_s = \Phi_{T_s} \mathbf{a}_{T_s} \quad (3)$$

and $\hat{m}_p$ be the approximate soot mass m_p obtained from the low-order POD matrix as

$$\hat{m}_p = \Phi_{m_p} \mathbf{a}_{m_p} \quad (4)$$

where $\mathbf{a}_\theta, \mathbf{a}_{T_s}, \mathbf{a}_{m_p} \in \mathbb{R}^{p \times T}$ are the a-coefficient matrices obtained from the inner product of the variable and its basis function. That is:

$$a_i(t) = \langle g(x, t), \phi_i(x) \rangle \quad (5)$$

$a_i(t) = \langle g(x, t), \phi_i(x) \rangle$ where $g(x, t)$ is a function over some domain of interest $\mathcal{Q}$, $\phi_i(x)$ is basis function, and $\langle g, \phi \rangle = \int_{\mathcal{Q}} g(x) \phi(x) d\mathcal{Q}$ denotes the inner product in domain $\mathcal{Q}$.

For compactness, let $\mathbf{X}$ represent θ, T_s , and m_p as follows:

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} \theta \\ T_s \\ m_p \end{bmatrix} \quad (6)$$

and

$$\hat{\mathbf{X}} = \begin{bmatrix} \hat{X}_1 \\ \hat{X}_2 \\ \hat{X}_3 \end{bmatrix} = \begin{bmatrix} \hat{\theta} \\ \hat{T}_s \\ \hat{m}_p \end{bmatrix} = \begin{bmatrix} \Phi_\theta \mathbf{a}_\theta \\ \Phi_{T_s} \mathbf{a}_{T_s} \\ \Phi_{m_p} \mathbf{a}_{m_p} \end{bmatrix} = \Phi_{\mathbf{X}} \mathbf{a}_{\mathbf{X}} \quad (7)$$

The time-explicit component of the SCR-coated filter model equations [2] can be written compactly as

$$\frac{\partial}{\partial t} [\mathbf{X}] = \begin{bmatrix} f_1(\mathbf{X}, \boldsymbol{\mu}) \\ f_2(\mathbf{X}, \boldsymbol{\mu}) \\ f_3(\mathbf{X}, \boldsymbol{\mu}) \end{bmatrix} = [\mathbf{f}(\mathbf{X}, \boldsymbol{\mu})] \quad (8)$$

where $\boldsymbol{\mu} = [y_i, T_i, \dots]$ is the parametric vector containing other system variables.

Using the POD modes as basis functions Φ_X and an approximation of $\mathbf{X}$ defined in equation 7, the POD-Galerkin projection is equivalent to

$$\frac{d}{dt} \mathbf{a}(t) = \Phi_X^T [\mathbf{f}(\Phi_X \mathbf{a}_X, \boldsymbol{\mu})] \quad (9)$$

with the initial condition of $\mathbf{a}_X(0) = \Phi_X^T \mathbf{X}(0)$.

$\mathbf{a}_X(t)$ is solved over time by Eq. 9, and $\hat{\mathbf{X}}(t)$ is estimated as $\hat{\mathbf{X}} = \Phi_X \mathbf{a}_X$ from Eq. 7. The other system variables are then obtained based on the approximated $\hat{\mathbf{X}}$ over time.

3 POD Results and Discussion

3.1 Steady State Application

The steady state NO_x conversion case reported in Schrade et al. [3] is used to evaluate the POD methodology. Our high fidelity model presented in [2] was validated with data from [3]; therefore it is considered a suitable reference against which to benchmark the performance of our POD model. Interested readers are referred to [3] for the details of the experimental set up and overall context for the case.

Table 1 presents the specification of the SCR-coated filter sample and the parameters of the steady-state NO_x conversion experiment [3]. The feed gas NH_3 and NO_x concentration trend is presented in Fig. 1.

The POD method was applied to capture two variation energy levels: 85% and 95%. This was done to indicate how the POD results approach the HFM as the variation energy level approaches 100%.

The POD results are presented in the following sections.

3.1.1 POD Modes

The spectrum of the singular values obtained from the decomposition of a state variable (ammonia surface coverage for sites S_1 and S_2 [2]) is presented in Fig. 2.

The result show that the singular values are dominated by a few points as most of the variation energy can be captured by those point. For example, in Fig. 2b, one POD mode captures 85% of the variation energy and two POD modes capture 99% of the variation energy.

The POD modes selected to capture the given variation energy level are then deployed to reduce the model order.

3.1.2 POD Basis Functions

Figure 3 provides some insight into the nature of the first two POD basis functions used to describe this steady-state application. The plots are presented for the ammonia surface coverage (ASC) variable in site S_1 .

Notice first in Fig. 3a, the plots show the filling of the wall with ammonia from over time, with ASC settling at around 0.7 at 5000 s. This is the HFM output data that is used to generate the POD basis functions. Figure 3b shows this variable averaged over the wall domain over time. Notice that the wall starts at empty of ammonia and gradually fills up to circa 0.7 at time 5000 s. (Note that ammonia is supplied in excess to this system with ANR at circa 1.25; therefore, there is no apparent reduction in surface coverage over time as NO_2/NO_x ratio changes.)

The first two POD basis functions are shown in Fig. 3c. The first POD basis function $\Phi_{1\theta S1}$ appears like an average of the profiles $X_{\theta S1}$ over time. It is a vector value of circa 0.1 flat across the wall domain. The ASC is relatively flat for most of the time in this example, lending further support to the interpretation of the first POD mode as a type of average of the system data.

The second POD basis function $\Phi_{2\theta S1}$ is a downward sloping plot from the inlet end of the filter wall towards the outlet end. Note that some of the values of the second POD basis function is negative, which underscores the need to apply caution in any direct physical interpretation of the POD basis function [22].

Although it can be risky to provide a physical interpretation of POD basis functions, these two functions align with the function the SCR system. The conversion depends on the total or average coverage, which is represented by the constant value in the first basis function $\Phi_{1\theta S1}$. However, the ammonia slip depends on the spatial distribution of the coverage, which is typically skewed to the front. The second POD basis function $\Phi_{2\theta S1}$ is a bias free linear function, which represents this skewing of the spatial distribution. It is worth noting that these two basis functions are reminiscent of the Legendre polynomials, which are known to be good generic basis functions for systems with strong spatial correlation like this one.

According to Eq. 1, the system data is decomposed into spatial basis functions and temporal coefficients. The associated temporal coefficients estimated from Eq. 9 are shown in Fig. 3d for the first two POD functions. It is seen that the first temporal coefficient $a_{1X_{\theta S1}}$ retains the form of the averaged $X_{\theta S1}$. The second temporal coefficient $a_{2X_{\theta S1}}$ quickly approaches zero to reduce the contribution of the negative values of the second POD basis function on the resulting system

Table 1 POD model steady-state application, filter specification and parameters of SGB tests on SCR filter sample

Parameter	Value	Parameter	Steady-state NO _x conversion
Filter material	Silicon carbide	Space velocity	20 k/h
Length [m]	0.152	Temperature	200 °C
Diameter [mm]	19.18 ^a	Soot load	0 g/l
Cell density [cps]	300	Inlet soot concentration	25 mg/Sm ³
Wall thickness [mil]	12	Feed gas composition	250 ppm NH ₃ , 200 ppm NO _x , 8% O ₂ , 5% H ₂ O, with NO ₂ /NO _x = [0,1]
Material porosity [%]	58	Simulation runtime	10,000 s in 20 s timesteps = 500 s
Mean pore size [μm]	22		
Catalytic coating [–]	Cu-Zeolite		
Degreening	2 h @ 600 °C		

^a Width and height are 17 mm in [3], converted here to diameter with the same area.

approximation. It is this property of the POD modes whereby the combination of the first few POD basis functions and associated temporal coefficient produce results that appear similar to the source system data that makes the POD method attractive as a MOR technique. Additional improvement in data approximation is achieved with more POD basis functions, by definition.

3.1.3 Output Concentration

The model output NO_x and NH₃ concentration data is shown in Fig. 4 along with the results from a POD-reduced model at 85% and 99% variation energy level. Results from POD modes at 85% variation energy level are labelled “POD₁”, while the results from POD modes at 99% variation energy level are labelled “POD₂”.

Outlet concentration plots show that the performance of POD models agree closely with the HFM. The POD results at 99% energy are closer to the HFM results compared with

the 85% POD model for the NO_x and NH₃ concentrations. This is expected from the definition of the POD approach.

The performance of the POD model is evaluated by the root mean square error (RMSE) defined as the square root of the MSE in Eq. 12. The RMSE summary data for the POD model in steady-state application is presented in Table 2.

The performance of the POD model is good in the steady-state application. The RMSE of outlet NO_x is at most 2 ppm which is circa 1% of the input NO_x concentration. The RMSE of outlet NH₃ is at most 14 ppm which is circa 5% of the inlet NH₃ concentration. This demonstrates the ability of the POD model to approximate the HFM without loss of fidelity in the steady-state application.

The performance of the POD model for other relevant parameters is summarised in Table 3.

The results show that the performance of the POD model is satisfactory with low RMSE across the relevant parameters. The RMSEs are lower for the case POD₂ as expected because more system variation energy is captured to develop that POD

Fig. 1 POD model steady-state application, feed gas NO_x and NH₃ concentration

Fig. 2 POD model steady-state application, singular values of the state variable ammonia surface coverage. Few modes capture the most of the variation energy

Fig. 3 Further insight into the nature of the POD basis function for the steady-state application. Plot (a) is the HFM output for ASC on active catalyst site S1 over time. Plot (b) is averaged ASC over the wall domain over time. Plot (c) is the first two POD basis functions. The first POD basis function appear like an average of the ASC data over time, a static spatial representation of the data. Plot (d) shows the associated temporal coefficients. System data is approximated by the POD basis functions and associated functions; better approximation is achieved by using more POD basis functions by definition

model. The magnitude of the RMSEs for both POD models are within 1–6% of the average HFM outputs which further underscore that the RMSEs are low across the relevant parameters.

3.1.4 Simulation Time

The evidence of model order reduction is the reduction in simulation time. The simulations are carried out on a PC with the following specification: Intel i5 4460 processor, 3.2 GHz CPU, 32GB RAM running MATLAB R2018b software. The same time step is specified for the HFM and POD models.

Fig. 4 POD model steady-state application, outlet NO_x and NH_3 concentration compared between HFM and two POD models implemented

The simulation time obtained in the steady-state application is presented in Table 4. The real-time factor (RTF) is the ratio of simulation time to the real time. A model faster than real time has $RTF < 1$, and $RTF > 1$ indicates model is slower than real time. The system real runtime is 500 s (model run for 10,000 s in 20s-timesteps).

The results in Table 4 show that our model is faster than real time by one order of magnitude. This is true for the full-order high-fidelity model and, of course, for the POD models.

The simulation time results also indicate that there is a measure of model order reduction achieved with the POD model. This is seen in the 4% reduction in simulation time for the POD model with 85% variation energy. Simulation time savings of 2% is also seen in the POD model with 99% variation energy.

The savings in simulation time is below expectation. The modest time savings indicate that most of the simulation time is not in the domain of the problem that has had its dimension reduced by the POD method. The majority of the simulation time $\sim 80\%$ is spent on solving the inlet channel-wall layer-outlet channel coupled nonlinear boundary value differential equations [2].

The POD method is applied to the state variables with explicit-time terms in their model equations. The POD excels

Table 2 POD model steady-state application RMSE for outlet NO_x and NH_3 concentration

Parameter	NO_x	NH_3
HFM		
POD ₁	2 ppm	14 ppm
POD ₂	< 1 ppm	5 ppm

Table 3 POD model steady-state application RMSE for other system parameters

Parameter	POD ₁ RMSE	% of average HFM output	POD ₂ RMSE	% of average HFM output
Soot mass ^a [mg/l]	0.022	< 1%	0.021	< 1%
Pressure drop [Pa]	0.05	< 1%	0.02	< 1%
Outlet CO ₂ [%]	0.04	4%	0.03	3%
Outlet CO [ppm]	0.4	4%	0.2	2%
Average ASC on S ₁ [–]	0.017	3%	0.003	1%
Average ASC on S ₂ [–]	0.039	6%	0.008	1%

^a Inlet soot concentration is specified at 25 mg/Sm³ and continuous soot oxidation is modelled resulting in minimal soot accumulation. The focus of the work with soot is in the influence on system NO₂ concentration and overall NO_x conversion

at reducing the simulation effort for the evolution of the state variables in time by solving a reduced-order initial value differential equation. In this application, the POD method achieves almost 10% reduction in simulation time for this component of the system. The time savings are dwarfed however by the invariant simulation requirements of the principal component of the model solution: the inlet channel-wall layer-outlet channel coupled nonlinear boundary value differential equations [2].

In the steady-state application, the POD method can deliver a modest order reduction of the SCR-coated filter model which retains high degree of fidelity. However, because of the non-linear nature of the system equation, the simulation of the model must be calculated via the reconstruction of the full state; therefore, only a moderate reduction (2–4%) in computation time is achieved. The low RMSE demonstrate the high fidelity of the reduced order model.

3.2 Transient State Application

The performance of the POD methodology to a transient state application is presented in this section. The transient state application is based on reference [9].

Transient input conditions are obtained from a sample world harmonized test cycle (WHTC) conditions reported in ref. [25]. The engine-out NO_x emission concentrations and the exhaust gas temperatures are obtained from the reference.² Details of the test engine conditions can be found in ref. [25]. The exhaust gas flow rate was constant for the simulation. In the experiment, liquid urea solution was injected; thus, the input NH₃ gas concentration was not measured. For our analysis, NH₃ gas is injected at ANR = 1.0. The inlet NO_x and NH₃ concentrations and exhaust gas temperatures are presented in Fig. 5. The filter substrate of Table 1 is retained for this application.

² The NO₂ concentration was increased by a factor of 10 in this analysis. This is done to increase NO₂/NO_x ratio so that the interaction of SCR and soot oxidation reactions can be realised.

The performance of the POD model in a transient state application is discussed in the following sections.

3.2.1 POD Modes

The fundamental idea exploited in the POD method is that in a high-dimensional data ensemble of independent variables, the dominant components of the data can be captured with a surprisingly few number of orthogonal modes [26]. In the transient mode, the singular value decomposition of the ammonia surface coverage shows that only a few singular modes describe most of the variation energy in the system. This is consistent with the observation from applying the POD to the steady-state case.

In Fig. 6b, it is seen that the relative energy of a few POD modes dominates over other POD modes. In this example, two modes are needed to capture 85% of the variation energy in the ammonia surface coverage data and four POD modes can capture 99% of the variation energy. Therefore, the POD method is suited for model order reduction in the transient state application.

3.2.2 POD Basis Functions

In a similar analysis as the steady-state application, Fig. 7 presents an insight into the nature of the POD basis functions for the transient state application.

Figure 7a shows the profile of ammonia surface coverage on active site S₁ over the wall at different times in the simulation. Due to the transient state scenario, the ASC is not

Table 4 POD model steady-state application, simulation times showing the extent of model order reduction achieved

Parameter	Time [secs]	RTF [–]	% change
HFM	70.47	0.141	
POD ₁	67.64	0.135	–4%
POD ₂	69.02	0.138	–2%

Fig. 5 POD model transient state application, feed gas temperature and NO_x and NH_3 concentration

monotonically increasing like the steady-state scenario. The variation in ASC can be seen in Fig. 7b. The variation in ASC is due to variation in inlet specie concentration and consumption/storage of species within the wall.

While it can be difficult to offer any clear physical interpretation of the POD basis functions as shown in Fig. 7c, the first POD basis function Φ_{10S1} can be considered essentially constant, representing the average of the ammonia surface coverage data ensemble. The other POD basis functions are somewhat similar to the Legendre polynomials, but they are showing more variation towards the front of the catalyst, where higher gradients are to be expected. With increasing

order, the basis functions show increasing variance over distance, helping to resolve finer variations in coverage.

The a-temporal coefficients associated with the POD basis functions are shown in Fig. 7d with the first coefficient $a_{1X_{0S1}}$ showing a profile similar in form to the average ammonia surface coverage over the wall domain (Fig. 7b). This supports the interpretation of the first POD basis function as a form of system data average. The other a-temporal coefficients are to be interpreted as the statistical complements of the POD basis functions that enable an approximation of the system data to be achieved by the POD method.

3.2.3 Output Conversion

The performance of the POD model is here investigated with the cumulative NO_x conversion and cumulative NH_3 slip over the simulation period.

Figure 8 presents the cumulative NO_x conversion and NH_3 slip for the high-fidelity model (HFM) and two POD models. POD model developed with POD modes which capture 85% of the variation energy, namely, POD_1 and POD model developed with POD modes which capture 99% of the variation energy, POD_2 .

The POD model results show close alignment with the HFM result for NO_x conversion and NH_3 slip. The performance of POD model which captures 99% of the variation energy exceeds that of POD_1 which is designed to capture 85% variation energy. This is expected by definition of the POD method.

Table 5 presents the RMSE of the outlet NO_x and NH_3 concentrations.

Fig. 6 POD model transient state application, singular values of the state variable ammonia surface coverage. Few modes capture the most of the variation energy

Fig. 7 Further insight into the nature of the POD basis function for the transient state application. Plot (a) is the HFM output for ASC on active catalyst site S_1 over time. Plot (b) is averaged ASC over the wall domain over time. Plot (c) is the first four POD basis functions. The first POD basis function appear like an average of the ASC data over time, a static spatial representation of the data. Plot (d) shows the associated temporal coefficients. System data is approximated by the POD basis functions; better approximation can be achieved by using more POD basis functions by definition

The performance of the POD model is good in the transient state application. The RMSE errors in NO_x outlet concentration is circa 3% of the mean inlet NO_x concentration for POD_1 and less than 0.3% for POD_2 . The NH_3 RMSE is circa 12% of mean NH_3 inlet concentration at POD_1 and about 0.4% at POD_2 . The low RMSE, particularly for POD model #2 where 99% of the variation energy is captured, demonstrates that the reduced model order retains a high degree of fidelity for this transient case application.

Table 6 show RMSE for other relevant system parameters.

The results show that the performance of the POD model is satisfactory with low RMSE across the relevant parameters. The RMSEs are lower for the case POD_2 as expected because

more system variation energy is captured to develop that POD model. The magnitude of the RMSEs for both POD models are within 1–11% of the average HFM outputs which further underscore that the RMSEs are low across the relevant parameters.

3.2.4 Simulation Time

The simulation times recorded for the HFM and the POD models for the transient state application are presented in Table 7. The simulation real runtime is 391 s (model run for 1955s in 5 s time steps). Consistent with the steady-state application, the models achieve faster than real-time simulation time based on their real-time factor $RTF < 1$.

The results show poor performance in model reduction as there is only a marginal reduction in simulation time less than 1% for POD_2 model. There is no apparent improvement in simulation time with the POD_1 model.

Consistent with the steady-state application, much of the simulation time is spent in solving the coupled non-linear boundary value (BVP) equation of the inlet channel, wall layer and outlet channel species transport balance. See Figure 9 for a split of the simulation time. The POD method is applied to the state variables with explicit-time terms in their

Fig. 8 POD model transient state application, cumulative NO_x conversion and cumulative NH_3 slip compared between HFM and two POD models implemented

Table 5 POD model transient state application RMSE for outlet NO_x and NH_3 concentration

Parameter	NO_x	NH_3
HFM		
POD_1	17 ppm	67 ppm
POD_2	<1 ppm	2 ppm

Table 6 POD model transient state application RMSE for other system parameters

Parameter	POD ₁	% of average HFM output	POD ₂	% of average HFM output
Soot mass [mg/l]	0.17	1%	0.17	1%
Pressure drop [Pa]	0.46	< 1%	0.1	< 1%
Outlet CO ₂ [ppm]	< 1	4%	< 1	3%
Outlet CO [ppm]	< 1	4%	< 1	2%
Average ASC on S ₁ [–]	0.02	5%	0.001	< 1%
Average ASC on S ₂ [–]	0.03	11%	0.002	1%

model equations, but this contribute less than 15% of the total simulation time and the improvements in solving those equations by the POD model do not transfer sufficiently to the overall simulation time savings.

The POD model delivers improvement in the “Other” component of the simulation time, but not enough to improve overall simulation time. In the case of POD₁, the increase in the time taken to solve the BVP component of the model (due to larger initial deviation and more iterations to reach desired error tolerances) altogether reverses any benefit in the POD application.

The POD method—although delivers a high-fidelity low order approximation of the full order model—has not achieved any significant reduction in simulation time. So, while theoretically valid, the practical use remains limited.

4 Artificial Neural Networks

The second approach to model order reduction applied in this work is artificial neural networks (ANN). In this approach, data was obtained from the high-fidelity model presented in reference [2] and a black-box approximation of the HFM is derived.

The motivation for this approach is twofold: (a) the ANN method is well suited for nonlinear model function approximation [10–12]; and (b) ANN can deliver model simplicity and adequate representation required for deploying the SCR-coated filter model in an online application.

Neural network has been applied in exhaust after-treatment system (EATS) modelling [13–19]. Chi in [13] used ANN to simulate the monolith bed temperature for more accurate

estimation of the effect of radial mal-distribution of energy on reaction and by extension EATS performance. ANNs have been used to facilitate the application of NO_x virtual sensor for emissions monitoring and online on-board diagnostics [14, 15]. Brahma et al. [16] coupled neural networks with phenomenological soot modelling to improve the performance of a DPF unit within a broader EATS control design and optimisation framework. ANNs have also been used to evaluate the relative performance of different SCR catalyst formulation on NO_x conversion [17, 18]. In this work, we apply ANN to the problem of approximating the SCR-filter model for control-oriented application. The authors are not aware of any previous application of the ANN method in order reduction of the SCR filter model. The performance of the ANN method is compared against an alternative MOR technique.

The process of developing our ANN is as follows. Firstly, data was collected over the desired range of operating conditions. Thereafter, a multilayer neural network was created and trained with 85% of the data collected. Cross-validation was carried out with 10% of the training data, and model testing was with the remaining 15% of the data. Model performance

Fig. 9 POD model transient state application, percentage contribution of BVP solver to overall simulation time. Although POD model improves simulation time of the “other” component of simulation time, but being a small element of the total, results in, at best, a marginal improvement in overall simulation time

Table 7 POD model transient state application, simulation times showing the extent of model order reduction achieved

Parameter	Time [secs]	RTF [–]	% change
HFM	59.76	0.15	
POD ₁	66.78	0.17	+ 10%
POD ₂	59.39	0.15	< – 1%

was further tested with a new dataset not used for model training.

This section provides further remarks on the development process of the ANNs.

4.1 System Data

System data was collected from the validated model of the SCR-coated filter developed in reference [2]. Table 8 presents the input and output data. The system state variables are also presented. The state variables are those with time-explicit terms in the SCR-coated filter model. In this model, the state variables drive the evolution of system properties over time. In Table 8, N is the number of discrete unit along the filter channel, w is the number of discrete units along wall, and T is the number of data samples (Fig. 17).

Note that the lumped parameter approach of Kladopoulou et al. [27] is applied to the wall temperature dynamics in the HFM [2]. This approach was taken to simplify the energy model in line with the objective of model development for system control. In our experimental set-up for model validation, thermocouples were inserted at various positions in the monolith to cover a wide range of the radial and axial surface. A volume-weighted average of the temperature data is correlated against the wall temperature predictions of the lumped model with satisfactory performance in a similar way to the work of Kladopoulou et al [27].

Figure 10 shows a schematic of the artificial neural network system showing input, output and state variables.

The process of data generation is stepwise random excitation of the input data within a defined upper and lower limit. The upper and lower limits of the variables are defined to cover the range expected from the model operation. The HFM was validated against Schrade et al. data [3] in this work. Therefore the range of our input parameters were dictated by the range of the reported experimental data in [3]. It is possible to extend our model beyond this application if good-quality data is available for model calibration. The data is discrete, generated at 1 s time steps. The dimension of each variable is included in Table 8 with the number of datapoints equivalent to T . In order to excite the system state variables sufficiently within the operational range, the input signal is held for 100 s between excitation levels. The maximum residence time for the Schrade et al. system [3] is circa 6 s.

Figure 11 shows the input signals in a matrix plot. The diagonal of the matrix plot show histograms of the input signals over the range specified. The histogram of each input variable shows that the variable is drawn from a uniform distribution within the specified range. The only exception is the mass of urea (plot in diagonal #4) which is skewed left to match the NH_3/NO_x ratio (ANR) specified for the analysis. The plot highlights that the exposure of the input sufficiently covered the range of interest.

Table 8 System data for ANN, includes input, output and state variables

Parameter	Symbol, units	Dimension
Input		
Inlet temperature	T_{in} , °C	[1xT]
Inlet NO _x conc	$\text{NO}_x - \text{in}$, ppm	[1xT]
NO ₂ /NO _x ratio	NNR, [-]	[1xT]
Inlet soot conc	$[\text{PM}]_{\text{in}}$, mg/Sm ³	[1xT]
Space velocity	SV , k/h	[1xT]
Urea injection (ANR)	m_{Urea} , kg/h	[1xT]
Inlet CO conc	CO_{in} , ppm	[1xT]
Inlet CO ₂ conc	$\text{CO}_2 - \text{in}$, %	[1xT]
Output		
Outlet NO _x conc	$\text{NO}_x - \text{out}$, ppm	[1xT]
Outlet NH ₃ conc	$\text{NH}_3 - \text{out}$, ppm	[1xT]
Outlet CO conc	CO_{out} , ppm	[1xT]
Outlet CO ₂ conc	$\text{CO}_2 - \text{out}$, %	[1xT]
Pressure drop	ΔP , kPa	[1xT]
Outlet temperature	T_{out} , °C	[1xT]
State		
Ammonia surface coverage S1, S2	$[\theta_{S1}, \theta_{S2}]^T$, [-]	[2NwxT]
Wall temperature	T_{wall} , °C	[1xT]
Soot mass in wall	m_{Wall} , kg	[NwxT]
Soot mass in cake	m_{Cake} , kg	[NxT]

The size of the data generated is 100,000 samples. Eighty-five percent of this data was used for training the ANN. Ten percent of the training data was used for cross-validation. The remaining dataset was used to test the trained ANN model. A new dataset with 100,000 samples was used to further test model performance.

4.2 ANN Architecture

The ANN architecture is a classic multi-layer perceptron comprising input node, a single hidden layer, an output layer and output nodes. See Fig. 12 for a schematic of the network.

The size of the input node is shown as $8 + X$, the size of the output node is shown as $6 + X$, where 8 is the number of input parameters and 6 is the number of output parameters as per Table 8, and X is the number of the state variables after dimension reduction from the original dimension of $N + 1 + 3Nw$. The approach to dimension reduction is presented in Section 4.4.

The number of hidden layers and the number of neurons in each hidden layer are selected to provide adequate function representation without over-fitting. One hidden layer of 40 neurons was found to be acceptable for this exercise.

The transfer function for the hidden layer is the hyperbolic tangent sigmoid (tansig), because it is a good choice for many

Fig. 10 Schematic of ANN showing input, output and state variables

Fig. 11 Matrix plot of input signals to ANN showing the limits of signal excitation. Histogram in diagonal show uniform distribution of samples within the parameter space

Fig. 12 Schematic of the ANN architecture. Typical network contain input, hidden and output layers

nonlinear functions [17]. The transfer function for the output layer is linear (purelin). The transfer function equations are given in Eq. 10 and 11 respectively.

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (10)$$

$$f(x) = x \quad (11)$$

During network training, the weights and biases associated with each neuron are determined to minimise the mean sum of square errors (MSE) between the model output (y_o) and target data (y_t). The expression for MSE is given in Eq. 12

$$MSE = \frac{1}{T} \sum_j (\hat{y}_t - \hat{y}_o)^2 \quad (12)$$

where $\hat{y}_t$ and $\hat{y}_o$ are the normalised³ dataset designed to exclude any influence of the relative magnitude of the system variables on the overall MSE.

Network training is achieved via the back-propagation method using the Levenberg-Marquardt algorithm [28]. The ANN training is implemented in the Deep Learning Toolbox of MATLAB R2018b.

4.3 Static vs. Dynamic Networks

Many ANN structures used for engineering applications are either static (feedforward) neural networks or dynamic (recurrent) neural networks with feedback. In the static ANN (Fig. 13) neurons respond instantaneously to input signals and are well suited for pattern recognition applications where both input and output variables represent spatial patterns independent of time [29].

Dynamic networks (Fig. 14) employ feedback between the neurons of a layer and/or between the layers of the network. The dynamic network more closely represents the working of the biological neurons with the feedback denoting local memory attributes [29].

Comparatively, static ANNs are conditionally stable and better at function approximation in real-time applications [30]. Static ANNs do not take into account time delays and are weaker in modelling nonlinear systems [29].

Dynamic ANN on the other hand exploit feedback and time delays to achieve better performance in nonlinear systems in general and are particularly appropriate for system identification, control and filtering applications [29]. Dynamic ANNs are more complex, require larger training datasets and take longer to train [30].

In this work, we implemented both the static and dynamic networks to compare their performance for this application. Further details on the specific architecture of the static and dynamic networks are provided as follows; however, only the results of the dynamics ANN are reported thereafter due to space constraints.

4.3.1 Static Network

The static network architecture is presented as a schematic in Fig. 15. In the static network, the state variables set as input and the derivatives are outputs. These are connected via the Simulink integrator block to simulate feedback.

4.3.2 Dynamic Network

The dynamic recurrent network is presented in a schematic in Fig. 16. The state variables are set as input to the dynamic network at time, t ; while the outputs are the state variable at time, $t + 1$. The feedback loop is connected by a unit-delay block “D”.

Unlike the static network above, which approximates the derivative of the state, this one approximates the state itself for the next time step. In a linear interpretation, there is no difference between the two approaches. But since the network is non-linear, this approach tends to be inherently more stable.

4.4 State Variable Dimension Reduction

The solution of the HFM was over an inlet channel-wall layer-outlet channel representation of the SCR-coated filter. The solution domain was discretised into N units in the axial dimension and w units in the wall dimension [2]. Figure 17 shows the solution domain.

The size of the ANN input node is $8 + N + 1 + 3Nw$ which for $N = 10$ and $w = 5$ results in 169 input nodes. The size of the

³ minmax normalisation shrink data to [-1,1]

Fig. 13 Schematic of a static network

output node is $6 + N + 1 + 3Nw$ which is 167 for the same N and w discrete units.

The dimension of the state variables within the input and output nodes were reduced to make the training of the ANN memory efficient. Table 9 shows the lumping technique used for each state variable. The reduced dimension of the input node is 13 and the output node is 11. This approach achieves a significant reduction of the number of state variables and therefore of the model order.⁴ Unlike the POD described in Section 2, the chosen basis is not ideal, but it has a clear physical interpretation.

It is noted that this state variable dimension reduction step results in what can be considered a zero-dimensional (0D) model; this is coincidental. The aim in this section was to lump the state variable in a manner that supports the ANN framework, so that model training is memory efficient. The physical interpretation of the lumping process mirrors the assumptions made in deriving a 0D model, namely, that the system can be considered as a CSTR which is homogenous in the spatial domain. Consideration of further overlaps in the ANN model and the conventional 0D model is outside the scope of this paper.

5 ANN Results and Discussion

This section presents the performance of the static and dynamic neural networks developed.

5.1 Overall

The MSE performance of the network for the training, validation and test data is presented in Figure 18. Training was

⁴ We found lumping by averaging and summing as described in Table 9 to be effective for our objective. Readers are invited to experiment with other dimension reduction methods.

relatively fast for this network with convergence achieved in 4.5 h.

The performance metric in Fig. 18 is for the normalised dataset. The best normalised MSE is 0.0148. The actual network performance when the data is de-normalised is presented in Table 10 for the training and test datasets.

The R^2 parameter is about 98% for both the training and test datasets. This means that the ANN model can explain 98% of the variation within the datasets. This is good performance as an R^2 of 1.0 indicates perfect correlation.

The R^2 parameter is sensitive to outliers; therefore, a modified index of agreement d_1 parameter is also reported. The index of agreement parameter is based on absolute errors and reflects the proportional contribution of each error estimate to the overall model performance without inflation by their square values [31]. The index of agreement at 0.97 indicates good agreement between the model prediction and the true output observations.

The R^2 and d_1 results show that the ANN approximation of the full-order SCR filter model performs adequately.

5.2 Specific Output Variables

Figure 19 presents a further look at the performance of the outlet NO_x concentration. Figure 19a shows the ANN model output of outlet NO_x concentrations against the target HFM data. The R^2 of the ANN for outlet NO_x in the test dataset is 0.95.

The results in Fig. 19a show that for the training and test datasets, the majority of model output outlet NO_x concentration congregates around the diagonal line of fit.

Figure 19b shows the distribution of the deviation in outlet NO_x concentration between the ANN model output and the target HFM data for the test dataset. The results show that most of the points are around the 0–25 ppm error and 80% of the observations are within ± 80 ppm deviation.

Fig. 14 Schematic of a dynamic recurrent network with feedback from output to input layer

Fig. 15 Schematic of Simulink implementation of static ANN. State variables and derivatives are neural network input and output, respectively. Simulink integrator block to simulate feedback

Figure 19c shows the distribution of the NO_x conversion predicted in the ANN and HFM. The results show that the ANN model tracks the HFM with similar relative frequencies of NO_x conversion over the test data.

A closer look at the ANN performance on outlet NH_3 concentration is presented in Fig. 20.

Figure 20a shows the model outlet NH_3 concentration against the target data for the train and test datasets. The results show that most of the observations congregate around the line of fit. The test data R^2 value shown on the plot indicate that the ANN model can explain 90% of the variation in the target HFM data. The relatively low R^2 value for the outlet

NH_3 concentration is due to few outlier values with large deviation.

The distribution of the deviation between the ANN model output and the target data for the test dataset is presented in Fig. 20b. The histogram corroborates the observation in Fig. 20a with most of the observation around 0–25 ppm. Ninety percent of the model output NH_3 concentrations are within 50 ppm of the target HFM NH_3 concentrations.

Figure 20c provides further insight to the performance of the ANN model in the distribution of the prediction of NH_3 slip between the ANN and HFM reported for the test data. The results show that the ANN model tracks the performance of

Fig. 16 Schematic of Simulink implementation of dynamic ANN. State variables as network input and output at one timestep ahead. Simulink unit delay block to simulate feedback

Fig. 17 The full-order SCR-filter model solution domain

the HFM satisfactorily with similar relative frequencies of NH_3 slip reported.

In general, the results of the model on outlet NO_x and NH_3 concentrations show that the ANN can approximate the SCR-filter model.

The performance of the other output variables against the target data in our model is presented in Fig. 21. The plot shows the performance for the training dataset (blue) and the test dataset (red) for outlet CO and CO_2 concentrations, filter pressure drop and the outlet temperature. The model R^2 for the test data is overlaid on each plot.

The summary of the model performance for the output variables is presented for the training and test dataset in Table 11.

The low RMSE, MAE and the high R^2 values for the training dataset show that the trained neural network can approximate the full order SCR-filter model to a reasonable level of accuracy. Similar level of performance on the test data show that the trained neural network can be generalised as a function approximation.

Note that R^2 performance for outlet CO and CO_2 reported in Table 11 shows apparent perfect correlation between the ANN and HFM ($R^2 = 1$). This is due primarily to the almost inert nature of those components in this system. Diesel oxidation catalyst is not modelled in the HFM and inlet soot concentration is relatively small. Therefore, CO and CO_2 concentration change within the system is insignificant. These results are reported for completeness. Figure 22 shows how the ANN

model tracks the increase in CO and CO_2 due to soot burn-off in our model.

5.3 Simulation Time

The training time was relatively short at 4.5 h. The system real runtime was 100,000 s in 1 s time steps. The model simulation time was 156 s. This is equivalent to a real time factor of circa 1.6×10^{-3} for this case. This means that the neural network model can achieve three orders of magnitude reduction in simulation time compared to real time.

5.4 Trained Network on New Dataset

To minimise the random effects of the training data and assess whether the trained model can be generalised, the trained model was tested against new, previously unseen dataset. (For ease of reference, the new dataset is called “Test Data #2”.) The new dataset was different from the

Table 9 State variables dimension reduction by lumped parameter

Parameter	Original dimension & dimension reduction technique	Reduced dimension
θ_{S1}, s_2	$[2Nw \times T]$, mean	$[2 \times T]$
m_{Wall}	$[Nw \times T]$, sum	$[1 \times T]$
m_{Cake}	$[N \times T]$, sum	$[1 \times T]$

Fig. 18 Dynamic ANN lumped parameter. Overall network MSE performance

Table 10 Dynamic ANN lumped parameter. Network performance metrics on training and test data

Parameter	Train	Test
R^2	0.98	0.97
d_1^a	0.97	0.96

^a Modified index of agreement d_1 is defined as follows: $d_1 = 1 - \frac{\sum |y_i - y_o|}{\sum (|y_o - \bar{y}_i| + |y_i - \bar{y}_i|)}$ where y_i is target observation, y_o is model prediction/output, and $\bar{y}_i$ is the mean of target observation [31].

original training data because of the different initial values of the input variables and random samples in the bulk of the dataset. Good model performance on the new dataset indicates that the model can be generalised beyond the specific training dataset.

The network performance on the new dataset is presented in Table 12 alongside the training and test performance on the original dataset.

The R^2 value is 0.96 for the new dataset which compares very well with the performance on the original test set. The d_1 value is 0.95 which compares well with the performance of the original test set. The outlet NO_x and NH_3 concentration RMSE and MAE errors are small and of similar magnitude to the original dataset error metrics. This demonstrates the ability of the trained dynamic model to approximate the performance of the full-order SCR-coated filter model for previously unseen data.

The network performance for outlet NO_x and NH_3 concentration is presented in Fig. 23 with similar performance as already discussed for the original dataset. The apparent better performance in the error distribution of Test Data #2 outlet NH_3 (Fig. 23d) compared with Test Data outlet NH_3 (Fig. 20b) even though the NH_3 RMSE and MAE is worse (Table 12) is due to the larger sample size of Test Data #2. Test Data #2 is 100,000 sample size while Test #1 is 15,000 samples, and the error deviation histogram output is in relative frequency.

5.5 Practical Considerations for ANN Modelling in EATS

This section presents a few thoughts on further practical aspects to consider for applying ANN to EATS modelling.

5.5.1 Quality of Training Data

The performance of an ANN depends on the quality of data available to train the model. System-specific data can be generated from high-fidelity models as we did in this study. This enables a degree of customisation of the range, domain and frequency of the data. The pitfall of this approach is that specific shortcomings of the HFM carry forward into the trained ANN. Alternatively, data

Fig. 19 Dynamic ANN lumped parameter. (a) Network output of outlet NO_x concentration against HFM for train and test data [in ppm]. (b) Distribution of deviation in network output outlet NO_x concentration (on test data) from HFM [in ppm]. (c) Distribution of NO_x conversion predicted from HFM and ANN (on test data)

Fig. 20 Dynamic ANN lumped parameter. (a) Network output of outlet NH_3 concentration against HFM for train and test data. (b) Distribution of deviation in network output outlet NH_3 concentration (on test data) from HFM

can be obtained from telemetry feedback from real-life deployment of the specific units of interest. More effort is required to process this data for use in the ANN framework. The practical restrictions in the measurement of some system variables (e.g. ammonia surface coverage,

NO_x speciation) further limit the ability of the output models to generalise acceptably. It is not practical to generate data for the ANN from SGB-type experiments because of the efforts required to generate the volume of data required to train a neural network. Once good-

Fig. 21 Dynamic ANN lumped parameter. Network performance on other output variables on train and test data

Table 11 Dynamic ANN lumped parameter. Network performance on other output variables on train and test data

	Parameter	Train	Test
NO _x	R ²	0.96	0.95
	RMSE (ppm)	89	105
	MAE ^a (ppm)	55	68
NH ₃	R ²	0.95	0.90
	RMSE (ppm)	35	57
	MAE (ppm)	20	27
CO	R ²	1.0	1.0
	RMSE (ppm)	0.08	0.1
	MAE (ppm)	0.06	0.07
CO ₂	R ²	1.0	1.0
	RMSE (%)	0.006	0.006
	MAE (%)	0.005	0.005
DP	R ²	0.99	0.99
	RMSE (kPa)	0.02	0.03
	MAE (kPa)	0.02	0.02
Temp	R ²	0.98	0.96
	RMSE (°C)	4.5	5.0
	MAE (°C)	2.0	2.1

^a MAE is mean absolute error. It is defined as: $\frac{1}{T} \sum_J |y_i - y_o|_J$.

quality data is available, the process of developing ANN models is relatively straightforward.

Table 12 Dynamic ANN lumped parameter. Network performance on new test data

Parameter	Train	Test	Test #2
Overall R ²	0.98	0.97	0.96
Overall d ₁	0.97	0.96	0.95
NO _x RMSE (ppm)	89	105	116
NO _x MAE (ppm)	55	68	70
NH ₃ RMSE (ppm)	35	57	93
NH ₃ MAE (ppm)	20	27	31

5.5.2 Online Training

By online training, we mean setting up a neural network which continues to accept data input and continually updates the model parameters considering the new data. This is an important aspect of ANN for current and future applications in EATS reduced order modelling. It is envisaged that most of the data for ANN development will be via telemetry feedback from real-life deployment of the specific unit of the EATS. The initial ANN will be trained perhaps on the first tranche of HFM data but with functionality to update “on the fly”. The fundamental steps to setting up the base ANN model is described in the previous sections. The online training algorithm will incorporate some dynamic backpropagation. The interest reader is referred to [12] for additional information on customising the ANN model for online training.

Fig. 22 Dynamic ANN lumped parameter. Network performance on output CO and CO₂ concentration. Percentage increase in CO and CO₂ concentration due to soot oxidation in the system. Soot content is small therefore apparent perfect correlation is reported between ANN and HFM. **a** Distribution of increase in outlet CO concentration. **b** Distribution of increase in outlet CO₂ concentration

6 Conclusions

An SCR-coated filter model was developed in [2]. A new implementation approach which delivers faster than real-time simulation times was demonstrated. The full-order SCR-filter model (HFM) is however complicated and does not lend itself to easy implementation for control application on a vehicle ECU.

To achieve the objective of model-based control development, it is necessary to reduce the complexity of the SCR-filter model. The SCR-filter model order reduction (MOR) has been attempted in this work. Two MOR techniques were implemented: (a) proper orthogonal decomposition, a grey-box approach, and (b) artificial neural networks for function approximation, a black-box approach.

The performance of both MOR approaches is compared in Table 13.

The POD delivers a reduced order model with a high degree of fidelity to the HFM. This is evident in the low RMSE values.

Fig. 23 Dynamic ANN lumped parameter. Network performance on new test data

The reduction in simulation time delivered by the POD is marginal when compared with the simulation times of the HFM.

The POD method delivers a high-fidelity approximation to the HFM because the POD model is developed from the HFM—the POD bases are derived from system data—thus retaining some physical aspects of the full-order model. The reduction in simulation time is modest because most of the simulation time is spent in the solution of the coupled channel wall layer species boundary value differential equation.

The ANNs deliver a reduced order model which achieves a significant reduction in simulation time. The speed-up achieved is in the order of a thousand times faster than the HFM. The

fidelity of the ANN model is satisfactory, although significantly poorer than obtained in the POD approach.

Acknowledgements The authors would like to acknowledge the support of the Engineering and Physical Sciences Research Council (EPSRC) and Eminox Ltd. for the funding for this research.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Nomenclature θ , Ammonia surface coverage; T_s , Wall temperature; m_p , Soot mass retained; N , Number of discrete units in channel; T , Time period; Φ_p , Orthonormal matrix of first p POD modes; θ , Approximate ammonia coverage; a_X , a-temporal coefficient; X , Surrogate variable for compact representation of θ , T_s , $m_p \in \mathbb{R}^{N \times T}$; $\tilde{X}$, Approximate surrogate variables; $g(x, t)$, (variable); $f(X, \mu)$, SCR-F model equation; μ , SCR-F model equation parameter other than X ; t , (time); w , Number of discrete units in wall; T_{in} , Inlet temperature; NO_{x-in} , Inlet NOx conc; NNR , NO_2/NO_x ratio; $[PM]_{in}$, Inlet soot conc; SV , Space velocity; $mUrea$, Urea injection (ANR); CO_{in} , Inlet CO conc; CO_{2-in} , Inlet CO2 conc; NO_{x-out} , Outlet NOx conc; NH_{3-out} , Outlet NH3 conc; CO_{out} , Outlet CO conc; CO_{2-out} , Outlet CO2 conc; ΔP , Pressure drop; T_{out} , Outlet temperature; $[\theta_{S1}, \theta_{S2}]^T$, Ammonia surface coverage S1, S2; $T_{wall,lt}$, Wall temperature; m_{Wall} , Soot mass in wall; m_{Cake} , Soot mass in cake

Abbreviations ANN, artificial neural network; ANR, ammonia NOx ratio; ASC, ammonia surface coverage; BVP, boundary value problem; CSTR, continuously stirred tank reactor; DPF, diesel particulate filter; EATS, exhaust after-treatment system; ECU, engine control unit; HFM, high fidelity model; MOR, model order reduction; PCA, principal component analysis; POD, proper orthogonal decomposition; RMSE, root mean square error; RTF, real time factor; SCR, selective catalytic reduction; SVD, singular value decomposition; WHTC, world harmonised transient cycle

Table 13 Comparative performance of the POD & ANN MOR techniques. POD delivered high model fidelity but modest simulation time improvements. ANN delivered excellent simulation time improvements with satisfactory model fidelity

	NO _x RMSE [ppm]	NH ₃ RMSE [ppm]	RTF
HFM			0.15
POD Steady State	<1	5	0.14
POD Transient State	<1	2	0.15
Dynamics ANN	105	57	0.0016
Static ANN ^a	148	51	0.0069

^aDetails of the static ANN not provided here. The main difference between static and dynamic ANN is that it takes longer to train the dynamic ANN because of time-ordered samples that cannot be parallelised on a high-performance computer. The model performance is similar for static and dynamic ANNs with the same order of magnitude reduction in simulation time achieved

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Fleming RD. "Fuel economy of light-duty diesel vehicles". 1976. doi:<https://doi.org/10.4271/760592>
- Olowojebutu, S., Steffen, T., Bush, P.: SCR-filter model order reduction (1): development and validation of the base "high-fidelity" model. *Emiss Control Sci Technol.* **6**(1), 1–17 (2019). <https://doi.org/10.1007/s40825-019-00150-1>
- Schrade, F., Brammer, M., Schaeffner, J., Langeheinecke, K., Kraemer, L.: Physico-chemical modeling of an integrated SCR on DPF (SCR/DPF) system. *SAE Int. J. Engines.* **5**(3), 958–974 (2012). <https://doi.org/10.4271/2012-01-1083>
- Chen, H., Reuss, D.L., Hung, D.L.S., Sick, V.: A practical guide for using proper orthogonal decomposition in engine research. *Int J Engine Res.* **14**(4), 307–319 (2012). <https://doi.org/10.1177/1468087412455748>
- Fogleman, M., Lumley, J., Rempfer, D., Haworth, D.: Application of the proper orthogonal decomposition to datasets of internal combustion engine flows. *J. Turbul.* **5**(23), 1–3 (2004)
- Alonso, D., Velazquez, A., Vega, J.M.: A method to generate computationally efficient reduced order models. *Comput. Methods Appl. Mech. Eng.* **198**(33), 2683–2691 (2009)
- Cavendish, J.C., Oh, S.H.: A computationally efficient Galerkin technique for approximating transient diffusion—reaction equations in composite media. *Chem. Eng. J.* **17**(1), 41–54 (1979)
- Oh, S.H., Cavendish, J.C.: Transients of monolithic catalytic converters. Response to step changes in feedstream temperature as related to controlling automobile emissions. *Ind. Eng. Chem. Prod. Res. Dev.* **21**(1), 29–37 (1982)
- Olowojebutu, S., Steffen, T.: A Low Order Model of SCR-in-DPF Systems with Proper Orthogonal Decomposition. *SAE Tech Pap.* (2018). <https://doi.org/10.4271/2018-01-0953>
- Becerra, V.M., Deng, J., Nasuto, S.: A novel dynamic neural network structure for nonlinear system identification. *IFAC Proc.* **38**(1), 348–353 (2005)
- Funahashi, K.-I.: On the approximate realization of continuous mappings by neural networks. *Neural Netw.* **2**(3), 183–192 (1989). [https://doi.org/10.1016/0893-6080\(89\)90003-8](https://doi.org/10.1016/0893-6080(89)90003-8)
- Gupta M, Jin L, Homma N. "Static and dynamic neural networks: from fundamentals to advanced theory". John Wiley & Sons; 2004
- Chi JN. "Application of artificial neural networks to aftertreatment thermal modeling". 2012. doi:<https://doi.org/10.4271/2012-01-1302>
- Arsie I, Cricchio A, De Cesare M, Pianese C, Sorrentino M. "A methodology to enhance design and on-board application of neural network models for virtual sensing of NOx emissions in automotive diesel engines". 2013. doi:<https://doi.org/10.4271/2013-24-0138>
- De Cesare M, Covassin F. "Neural network based models for virtual NOx sensing of compression ignition engines". 2011. doi:<https://doi.org/10.4271/2011-24-0157>
- Brahma I, Rutland CJ, Foster DE, He Y. "A new approach to system level soot modeling". 2005. doi:<https://doi.org/10.4271/2005-01-1122>
- Panahi, P.N., Niaei, A., Tseng, H.-H., Salari, D., Mousavi, S.M.: Modeling of catalyst composition–activity relationship of supported catalysts in NH₃–NO–SCR process using artificial neural network. *Neural Comput. Applic.* **26**(7), 1515–1523 (2015)
- de Oliveira, M.L.M., Silva, C.M., Moreno-Tost, R., Farias, T.L., Jiménez-López, A., Rodríguez-Castellón, E.: A study of copper-exchanged mordenite natural and ZSM-5 zeolites as SCR–NOx catalysts for diesel road vehicles: simulation by neural networks approach. *Appl. Catal. B Environ.* **88**(3–4), 420–429 (2009)
- Majd Faghihi E, Shamekhi AH. "Development of a neural network model of an SCR catalytic converter and ammonia dosing optimization using multi objective genetic algorithm". 2011. doi:<https://doi.org/10.4271/2011-01-1332>
- Chatterjee, A.: An introduction to the proper orthogonal decomposition. *Curr. Sci.* **78**(7), 808–817 (2000)
- Holmes P, Lumley JL, Berkooz G. "Turbulence, coherent structures, dynamical systems and symmetry". Cambridge university press; 1998
- Kerschen, G., Golinval, J., Vakakis, A.F., Bergman, L.A.: The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: an overview. *Nonlinear Dyn.* **41**(1), 147–169 (2005). <https://doi.org/10.1007/s11071-005-2803-2>
- Astrid P. "Reduction of process simulation models: a proper orthogonal decomposition approach". Technische Universiteit Eindhoven Ph. D.-thesis; 2004
- Fourier JBJ. "On the propagation of heat in solid bodies". Mem read before CI Instut Fr. 1807
- Jeon, J., Lee, J.T., Park, S.: Nitrogen compounds (NO, NO₂, N₂O, and NH₃) in NO_x emissions from commercial EURO VI type heavy-duty diesel engines with a urea-selective catalytic reduction system. *Energy Fuel.* **30**(8), 6828–6834 (2016). <https://doi.org/10.1021/acs.energyfuels.6b01331>
- Liang, Y.C., Lee, H.P., Lim, S.P., Lin, W.Z., Lee, K.H., Wu, C.G.: Proper orthogonal decomposition and its applications—part I: theory. *J. Sound Vib.* **252**(3), 527–544 (2002)
- Kladopoulou EA, Yang SL, Johnson JH, Parker GG, Konstandopoulos AG. "A study describing the performance of diesel particulate filters during loading and regeneration - a lumped parameter model for control applications". 2003. doi:<https://doi.org/10.4271/2003-01-0842>
- Hagan, M.T., Menhaj, M.B.: Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Netw.* **5**(6), 989–993 (1994). <https://doi.org/10.1109/72.329697>
- Sinha NK, Gupta MM, Rao DH. "Dynamic neural networks: An overview". In: *Industrial Technology 2000. Proceedings of IEEE International Conference on.* Vol 1 IEEE; 2000:491–496
- Al-Zohairy TA. "A comparison study between static and dynamic recurrent neural networks based adaptive control of nonlinear multivariable systems". In: *2008 3rd International Design and Test Workshop.* 2008;301–306. doi:<https://doi.org/10.1109/IDT.2008.4802518>
- Legates, D.R., McCabe Jr., G.J.: Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **35**(1), 233–241 (1999)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.