Available online at www.sciencedirect.com

ScienceDirect

Proceedings of the Combustion Institute 38 (2021) 2701–2709

Proceedings of the Combustion Institute

www.elsevier.com/locate/proci

An *a priori* evaluation of a principal component and artificial neural network based combustion model in diesel engine conditions

Deepak K. Dalakoti [a,*], Armin Wehrfritz [a], Bruno Savard [b], Marc S. Day [c], John B. Bell [c], Evatt R. Hawkes [a,d]

[a] *School of Mechanical and Manufacturing Engineering, The University of New South Wales, Sydney, NSW 2052, Australia*
[b] *Department of Mechanical Engineering, University of Ottawa, Ottawa K1N 6N5, Canada*
[c] *Center for Computational Sciences and Engineering, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[d] *School of Photovoltaic and Renewable Energy Engineering, The University of New South Wales, Sydney, NSW 2052, Australia*

Received 8 November 2019; accepted 9 June 2020
Available online 29 August 2020

Abstract

A principal component analysis (PCA) and artificial neural network (ANN) based chemistry tabulation approach is presented. ANNs are used to map the thermochemical state onto a low-dimensional manifold consisting of five control variables that have been identified using PCA. Three canonical configurations are considered to train the PCA-ANN model: a series of homogeneous reactors, a nonpremixed flamelet, and a two-dimensional lifted flame. The performance of the model in predicting the thermochemical manifold of a spatially-developing turbulent jet flame in diesel engine thermochemical conditions is *a priori* evaluated using direct numerical simulation (DNS) data. The PCA-ANN approach is compared with a conventional tabulation approach (tabulation using *ad hoc* defined control variables and linear interpolation). The PCA-ANN model provides higher accuracy and requires several orders of magnitude less memory. These observations indicate that the PCA-ANN model is superior for chemistry tabulation, especially for modelling complex chemistries that present multiple combustion modes as observed in diesel combustion. The performance of the PCA-ANN model is then compared to the optimal estimator, *i.e.* the conditional mean from the DNS. The results indicate that the PCA-ANN model gives high prediction accuracy, comparable to the optimal estimator, especially for major species and the thermophysical properties. Higher errors are observed for the minor species and reaction rate predictions when compared to the optimal estimator. It is shown that the prediction of minor species and reaction rates can be improved by using training data that exhibits a variation of parameters as observed in the turbulent

* Corresponding author.
E-mail address: d.dalakoti@unsw.edu.au (D.K. Dalakoti).

https://doi.org/10.1016/j.proci.2020.06.263

flame. The output of the ANN is analysed to assess mass conservation. It is observed that the ANN incurs a mean absolute error of 0.05% in mass conservation. Furthermore, it is demonstrated that this error can be reduced by modifying the cost function of the ANN to penalise for deviation from mass conservation.

Keywords: Principal component analysis; Artificial neural network; Low-dimensional manifold; Diesel engine; Spray A

1. Introduction

The simulation of reacting flows is computationally expensive. This is in part due to the large number of thermochemical variables that need to be transported and the associated computationally expensive evaluation of chemical source terms and transport coefficients. An accurate description of the chemical kinetics may require hundreds of species and reactions, which leads to a very high computational cost. A smaller thermochemical state space is thus desired. One way to achieve this is to limit the reaction mechanism to only a few species, *i.e.* a global mechanism. However, such global mechanisms may not accurately describe fundamental combustion characteristics such as two-stage ignition or pollutant formation. An alternative approach is based on mapping the high-dimensional state space spanned by the thermochemical state vector onto a low-dimensional manifold (LDM). Such methods can yield a huge reduction in computational cost as the evolution of the thermochemical system can be uniquely described by a small number of control variables and thus can be stored in look-up tables.

The use of tabulated chemistry was first introduced by Peters [1] in their steady laminar flamelet model for nonpremixed flames. A more general theoretical framework on using intrinsic low-dimensional manifolds (ILDM) has been proposed by Maas and Pope [2]. Several LDM based methods have subsequently been proposed that include the flame prolongation of ILDM (FPI) model by Gicquel et al. [3], the flamelet-generated manifold (FGM) method by van Oijen and de Goey [4] and the flamelet progress variable (FPV) method by Pierce and Moin [5]. In brief, these LDM combustion models rely on canonical configurations (*e.g.* one-dimensional premixed flames) to obtain a high-dimensional state space, which is then projected onto an LDM and tabulated in a look-up table as a function of a few control variables (hereby referred to as the conventional tabulation approach). Such models have been successfully applied in several studies to model turbulent premixed [4,6] and nonpremixed flames [5]. However, most of these studies have been limited to relatively simple fuels which do not exhibit the multiple combustion mode structure observed in diesel combustion.

Several challenges arise when extending these methods to more complex fuels such as *n*-dodecane, a diesel fuel surrogate. First, fuels like *n*-dodecane exhibit complex two-stage ignition characteristics. Consequently, flames in diesel engines exhibit a multiple combustion mode structure consisting of low-temperature chemistry (LTC) and high-temperature chemistry (HTC). This makes selecting an appropriate LDM a challenging task. Second, it is not clear if an LDM consisting of one or two parametrising variables, as typically used in literature, gives an accurate representation of the full thermochemical system for complex fuels like *n*-dodecane. It is also not clear how LDMs with more than two control variables can be generated to tabulate the data obtained by parameter variations in canonical configurations. Third, tabulation using a structured uniformly-spaced approach requires large memory, which becomes computationally prohibitive when tabulation on several control variables is required and/or multiple variables are needed to be tabulated. Fourth, a structured uniformly-spaced approach may not give a smooth representation of the underlying nonlinear functions due to the prohibitively large number of data points required, and subsequently may give rise to numerical instabilities [7].

The limitations of the conventional tabulation approach can be addressed by using data driven techniques as described below. The first challenge can be addressed by using principal component analysis (PCA), a linear dimensionality reduction technique, to identify an LDM. The principal components obtained by PCA, unlike nonlinear dimensionality reduction techniques, are computationally cheap to compute, have well defined transport equations for *a posteriori* simulations [8,9] and form an orthogonal basis, which is desirable for presumed probability density function (PDF) closure methods [6,10]. This makes PCA an ideal technique to identify LDMs. PCA also addresses challenge number two as an arbitrary number of PCs (up to the dimension of the full chemical system) can be selected for parametrising the high dimensional state space. Sutherland and Parente [9] demonstrated *a priori* the applicability of PCA for obtaining an LDM. The variables identified by PCA were shown to be better at parametrising the high-dimensional state space

than the *ad hoc* defined control variables typically used in literature. *A posteriori* simulations using the PCA approach have also been conducted [8,11].

The third and fourth challenges can be addressed by using machine learning regression techniques to map, i.e. fit, the high-dimensional state space on to the LDM. Artificial neural networks (ANNs), in particular, are well suited for fitting highly nonlinear data. Using ANNs, tabulation of the thermochemical state space translates to storing the network architecture and the associated node weights and thus requires only a modest amount of data to be stored. Since ANNs learn the underlying functional form they are expected to yield smooth outputs, which is a shortcoming of the conventional tabulation approach [7].

A combustion modelling approach based on PCA to identify LDMs and ANN for tabulating (PCA-ANN approach) thus has the potential to address all of the challenges discussed above. A few studies employing ANNs for combustion modelling have appeared in literature. Sen et al. [12] used ANNs to tabulate syngas-air chemistry for large-eddy simulations (LES). The results with ANN chemistry tabulation showed good agreement with direct numerical simulation (DNS) statistics while providing significant speed-up compared to direct-chemistry integration. Similar observations were reported by Franke et al. [13] for LES simulations of methane-air combustion. Echekki et al. [11] used ANNs to tabulate principal component reaction rates for one-dimensional turbulence simulations of piloted methane-air flames. The PCA-ANN approach showed results close to the full transport approach. Similar results were reported by Malik et al. [14] and Isaac et al. [15]. Owoyele et al. [16] used ANNs to tabulate chemistry for simulation of flames in diesel engine conditions using Reynolds-averaged Navier–Stokes (RANS) and LES approaches. The results showed that ANNs reduce the memory requirement and the chemistry look-up time while well predicting the flame characteristics. PCA for selection of LDM was however not considered.

The studies discussed above have shown that the PCA-ANN approach improves the manifold reconstruction accuracy and addresses the challenges of chemistry tabulation discussed above. However, these studies have focussed on relatively simple fuels like syngas/methane/hydrogen and on simplified configurations like one-dimensional turbulence and partially-stirred reactors. No previous study has considered a PCA-ANN approach for combustion modelling of a 3D turbulent flame and in diesel engine conditions where a complex multiple combustion mode flame structure is observed. The objective of this work is thus to *a priori* evaluate the PCA-ANN combustion modelling approach in diesel engine thermochemical conditions using data from a recent DNS of a spatially developing turbulent jet flame [17]. The results from the PCA-ANN model are compared with optimal estimators and a conventional tabulation approach and strategies to improve the PCA-ANN predictions are further discussed.

2. Methodology

We use three canonical configurations to train the PCA-ANN model. The PCA-ANN model is then *a priori* evaluated using data from a DNS of a spatially-developing round jet turbulent flame in diesel engine conditions. The canonical configurations, the turbulent flame and the PCA-ANN methodology are described next.

2.1. Turbulent flame

A DNS of a spatially-developing round jet turbulent flame in diesel engine conditions is considered. The DNS are fully described elsewhere [17]. The thermochemical conditions are the same as that of the Engine Combustion Network (ECN) Spray A flame. A gaseous partially-premixed *n*-dodecane fuel jet was injected into a quiescent oxidiser at 900K and 15% O_2 by volume. The simulation was conducted until a quasi-steady lifted flame was established. A 53-species chemical mechanism was employed [18]. Figure 1 shows a 3D rendering of the flame coloured by combustion modes [17]. The flame presents a complex, multi-mode structure featuring LTC and rich premixed, lean premixed and nonpremixed HTC [17].

2.2. Canonical configuration

In the diesel engine modelling literature, typically zero-dimensional (0D) homogeneous reactors and one-dimensional (1D) nonpremixed igniting flamelets are used to tabulate chemistry. In the current study we thus evaluate these two configurations. In addition, two-dimensional (2D) laminar lifted flames stabilised in a convective mixing layer are considered to generate a database that accounts for laminar diffusion effects in both the streamwise and cross-stream directions. The thermochemical conditions are matched to the turbulent flame described above in all configurations. The 0D cases simply consider the ignition of a series of uncoupled constant pressure reactors with initial conditions determined from the mixing line. The 1D igniting flamelets are solved in mixture fraction space using the formulation of Pitsch and Peters [19] as implemented in FlameMaster [20]. A scalar dissipation rate at stoichiometric mixture fraction (χ_{st}) of $20s^{-1}$ is selected to represent the values observed in the turbulent flame near the flame base. A unity Lewis number assumption was

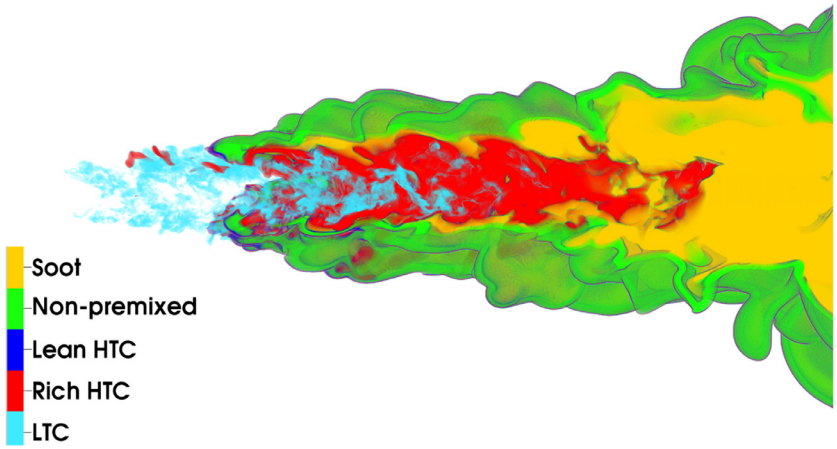

Fig. 1. 3D rendering of the flame coloured by combustion modes.

used for scalar transport. The 2D flames correspond to those presented in our previous study [21]. The fuel and oxidiser streams have the same inlet velocity, corresponding to the flame propagation velocity. Mixture-averaged transport properties were used for the 2D flames. Here, two cases with $\chi_{st} = 4\text{s}^{-1}$ and $\chi_{st} = 350\text{s}^{-1}$ at the inlet are used. Unless otherwise stated, results are presented with $\chi_{st} = 4\text{s}^{-1}$ case.

2.3. PCA

A $N \times m$ dimensional dataset $\mathbf{[\phi]}$ composed of N samples of the m-dimensional thermochemical state vector ($m = 54$, 53 species and temperature) is considered. The principal components $\mathbf{[\eta]}$ are defined as $\mathbf{[\eta] = [A^T][R]([\phi] - [M])}$ where $\mathbf{[A^T]}$ are the eigenvectors of the covariance matrix of $\mathbf{[\phi]}$, $\mathbf{[M]}$ is the component wise mean of $\mathbf{[\phi]}$ and $\mathbf{[R]}$ is the component wise inverse of the range of $\mathbf{[\phi]}$. Range scaling was chosen because it captures the highest variance of the dataset in the first few PCs while the performance for reconstructing species mass fractions is good for most species. For a detailed comparison of the various scaling methods see Ref. [22]. A perfect mapping from $\mathbf{[\eta]}$ to $\mathbf{[\phi]}$ can be obtained if all principal components are retained. LDMs can be identified by retaining a subset of $\mathbf{[\eta]}$. The choice of the number of retained η is guided by the desire to have the smallest LDM which captures a required level of variance in the dataset. In this work, we consider the first five components which are sufficient to capture $\approx 95\%$ of the variance.

The canonical configurations described in Section 2.2 are used to compute $\mathbf{[A^T]}$, $\mathbf{[R]}$ and $\mathbf{[M]}$. These are then used to compute the principal components for the turbulent flame dataset as, $\mathbf{[\eta_T] = [A^T][R]([\phi_T] - [M])}$, where $\mathbf{[\phi_T]}$ is the turbulent flame thermochemical database.

In *a posteriori* simulations transport equations for η need to be solved in addition to the usual transport equations for mass, momentum and energy. This subsequently requires the chemical source terms (ω_η), as well as the mixture thermal conductivity (λ), heat capacity (C_p), viscosity (μ), diffusion coefficients (D_η) and mean molecular mass ($\overline{W}$) to be obtained as functions of η. It should be further noted that mass fractions are not required in the *a posteriori* simulations, they are however of interest for post-processing. To ensure a consistent evolution of the thermodynamic state variables in the flow solver and the thermochemical database, the approach developed for tabulated chemistry by Vicquelin et al. [23] can be used.

The chemical source terms for η are defined as, $\mathbf{[\omega_\eta] = [A^T][R][\omega_\phi]}$, where $\mathbf{[\omega_\phi]}$ is the reaction rate matrix of the high-dimensional thermochemical basis. Following Biglari et al. [8], the diffusion coefficients for η are defined as, $\mathbf{[D_\eta] = [A^T][D_\phi]}$, where $\mathbf{[D_\phi]}$ are the mixture-averaged diffusion coefficients of the high-dimensional thermochemical basis. ANNs are used to store these quantities parametrised by η. Note that we only considered a linear PCA approach to avoid the high cost of computing the kernel function associated with non-linear PCA [24] for the large training dataset ($\mathcal{O}(10^6)$ samples) used in this work.

2.4. ANN

Fully connected dense ANNs are used to fit the thermochemical quantities on η. A single, multi-input, multi-output ANN is used for all species and temperature. Separate ANNs are used for each ω_η and the thermophysical properties. The architecture of the neural networks is described in the supplementary material. The ANNs are trained using data from the canonical configurations described in Section 2.2. The training data are divided into

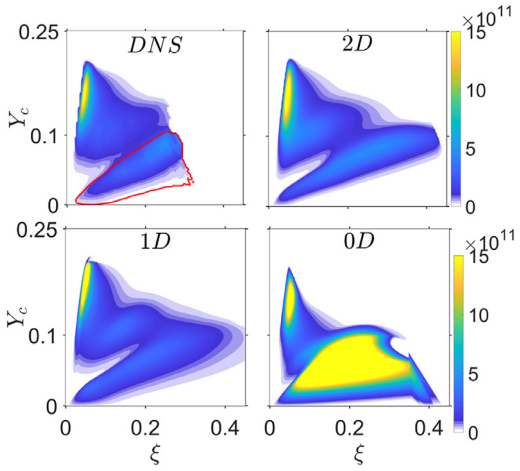

Fig. 2. Heat release rate conditioned on ξ and Y_c for the DNS and the various canonical configurations. The red contour marks the LTC region (where $Y_{C_12H_25O_2}$, an LTC marker, is greater than 5% of the maximum [17,21]).

training and validation sets in a 4:1 ratio. The training set is used to train the ANNs while the validation set is used to assess the model performance on new data and avoid overfitting. The trained ANNs are then used to predict the quantities for the turbulent flame based on η_T. The adequacy of this mapping is evaluated using the turbulent flame DNS data. The error in prediction of a variable y_i is defined as, $\epsilon_i = \sqrt{\overline{(y_i - y_{p,i})^2}/\overline{y_i^2}}$, where $y_{p,i}$ represents the value predicted by the model, y_i represents the value from the DNS and $\overline{y_i}$ represents the mean value of y_i. In addition, the coefficient of determination scores are presented in the supplementary material. The mean and median error in the prediction of all species is then defined as mean and median of $\{\epsilon_i \mid i = 1, N_s\}$, where N_s is the number of species. To put the results of PCA-ANN approach in context, we also present results from the optimal estimator ($< \phi | \eta_1, \ldots, \eta_5 >$), the conditional mean from the DNS which minimises the error [25]. Note that sufficient snapshots from the DNS were used to get converged statistics for the optimal estimator.

3. Results

3.1. Choice of canonical configuration

In this section, we evaluate the canonical configurations and compare their predictions with the DNS. Figure 2 shows the heat release rate (HRR) conditioned on mixture fraction (ξ) and progress variable ($Y_c = Y_{CO} + Y_{CO_2} + Y_{H_2} + Y_{H_2O}$), the LDM control variables typically used in literature [26], for the DNS as well as the three canonical configurations. A few observations follow. First, the DNS presents a complex combustion structure consisting of regions of LTC (region inside the red contour) and HTC (region outside the red contour). Second, qualitatively both 1D flamelet and 2D flame resemble the DNS flame structure. Third, the 0D reactors, commonly used in literature for combustion modelling, show highly enhanced LTC and HTC regions that do not resemble the DNS. Fourth, the 2D flame appears closest to the DNS structure. These observations are consistent with our previous study of LDM combustion modelling [27] where 2D flames were shown to be better in predicting the flame structure, both qualitatively and quantitatively compared to 0D reactors and 1D flamelets. Thus, we will only consider data from the 2D flames for generating the thermochemical database in the following. Results obtained with the 0D reactors and 1D flamelets are presented in the supplementary material.

3.2. Comparison with conventional tabulation

Before presenting the results of the proposed PCA-ANN model, we present a comparison with the conventional tabulation approach. In the conventional tabulation approach, typically ξ and Y_c are used as control variables for tabulation. In the PCA-ANN approach we use η as control variables. To test the adequacy of PCA to identify an LDM, we first compare the predictive performance of tables parametrised by ξ and Y_c (Tab_{ξ,Y_c}) with tables parametrised by η_1 and η_2 (Tab_{PCA}). The error in the prediction of HRR is used for the comparison due to its importance in combustion modelling and its complex functional form. Several uniformly discretised tables with different grid spacing were obtained from the 2D flame data (at $\chi_{st} = 4\text{s}^{-1}$) for the two sets of control variables (the sensitivity of table resolution is presented in the supplementary material). The tables with the smallest error were then chosen. The results show that Tab_{ξ,Y_c} incurs an error of 74% while Tab_{PCA} incurs an error of 67% in the prediction of HRR. The lower error for Tab_{PCA} shows that η are better in parametrising the high-dimensional manifold. Next, we evaluate the use of ANNs instead of a simple linear interpolation for retrieving the data from the LDM. Using an ANN reduces the prediction error to 54% and 56% for the LDMs with ξ and Y_c, and η_1 and η_2, respectively.

The advantage of the PCA-ANN approach is further highlighted when using more than two control variables for tabulation. Our test revealed that using first five η to train the ANN leads to a prediction error of 42% for HRR. This is significantly smaller than that obtained with two components. Using a five-dimensional table with the conventional linear interpolation approach becomes infeasible. For instance, tabulating a single variable parametrised by five control variables (as used here) with 100 points in each direction translates to storing 10 billion points or $\approx$ 80

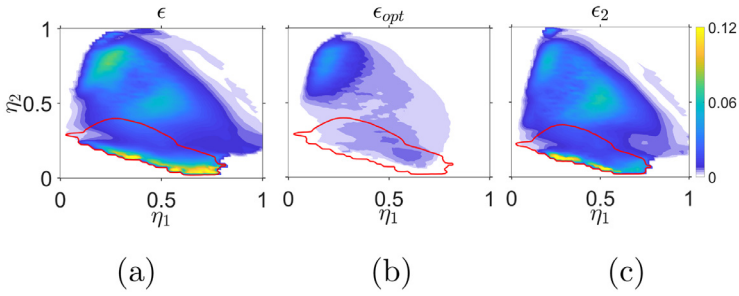

Fig. 3. Error incurred in the prediction of reaction rates conditioned on η_1 and η_2. Note that η_1 and η_2 have been normalised.

Table 1
Error in the prediction of ω_η from the ANNs and optimal estimators.

ω_η	ϵ	ϵ_{opt}	ϵ_χ
ω_{η_1}	42%	25%	38%
ω_{η_2}	50%	30%	34%
ω_{η_3}	46%	26%	31%
ω_{η_4}	56%	28%	36%
ω_{η_5}	85%	10%	56%

gigabytes in memory. Conversely, the ANN used in this paper (see Section 3.3) to fit all species and temperature (54 variables) only requires $\approx$ 32,000 coefficients (0.25 megabyte of memory). The retrieval of the thermochemical manifold using the PCA-ANN approach was also found to be about 2 times faster than the conventional tabulation approach. It should however be noted that the computational cost for retrieving data from ANNs will strongly depend on the ANN architecture (number of nodes, activation functions etc.). These results clearly illustrate the advantages of using ANNs over conventional tabulation approach.

3.3. Prediction of thermochemical manifold

In this section we evaluate *a priori* the performance of the PCA-ANN model in predicting the thermochemical manifold of the turbulent flame. The chemical source terms that are the most important quantities controlling the evolution of the thermochemical manifold in *a posteriori* simulations are considered first. Table 1 lists the error in the prediction of ω_η from the PCA-ANN model along with the error from the optimal estimator (ϵ_{opt}). Significant errors are incurred in the prediction of all ω_η by both the PCA-ANN model and the optimal estimator. Compared to the optimal estimator, the PCA-ANN incurs significantly higher errors for all ω_η. To identify the regions contributing to this error we present the total error incurred for all reaction rates conditioned on the first two η ($[\sqrt{\Sigma_n(f_i - \omega_{\eta_i})^2} \mid \eta_1, \eta_2]/\Sigma_n max_i(\omega_{\eta_i})$ where f_i is the predicted value for ω_{η_i}) in Fig. 3[1]. Also presented is the result for the optimal estimator. The red contour in the figure marks the LTC region. A few observations follow. First, consistent with the observations listed in Table 1, the ANN incurs higher errors than the optimal estimator in all regions. Second, the highest errors for the PCA-ANN model are concentrated in regions of LTC (within the red contour) and edge-flame (at high η_2 values). Quantitatively, the error in the LTC region is 77% while the error in the edge-flame region is 43% for the PCA-ANN model indicating that modelling LTC is much more challenging than HTC. The reasons for high errors will be discussed later in this section. As outlined in Section 2.3, other than the chemical source terms the thermophysical properties (C_p, μ, λ, $\overline{W}$, D_η) need to be obtained as a function of η in *a posteriori* simulation. The PCA-ANN model was found to yield excellent predictions for these properties with values within a 1.5% margin from those of the optimal estimator. For brevity, these results are not presented here but included in the supplementary material.

The predictions for the species mass fractions are next compared to the optimal estimator. The mean error in the prediction of all species is 20% and the median error is 17%. The corresponding mean and median errors from the optimal estimator are 16% and 10%, respectively. This result shows that the PCA-ANN model can predict the thermochemical state with good accuracy, *i.e.* comparable to the optimal estimator. This result is significant considering that the thermochemical state is parametrised by only five variables and only a single neural network, fitted on a canonical configuration is able to predict all species observed in the complex turbulent flame with good accuracy.

The results for CO_2, OH, CH_2O and $C_12H_25O_2$, the key species in the two-stage diesel engine combustion are now discussed. OH is a

[1] Note that although the results are obtained with five η, the conditioning is performed on the first two η for the purpose of representation in 2D space.

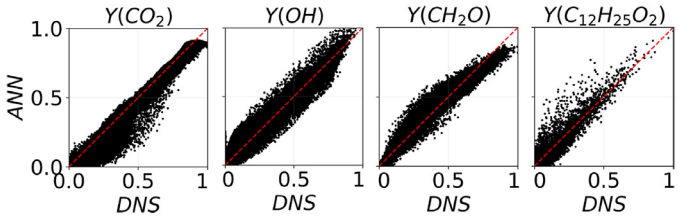

Fig. 4. Scatter plot of mass fraction of major species with DNS data on the abscissa and the ANN prediction on the ordinate. Note that normalised values are presented.

Table 2
Error in the prediction of minor species from the ANN and the optimal estimator.

Species	ϵ	ϵ_{opt}
CH_2^*	53%	40%
CH_2	52%	41%
C_5H_9	50%	32%
pC_4H_9	37%	19%
$OC_{12}H_{23}OOH$	36%	26 %

high-temperature marker, $C_{12}H_{25}O_2$ is an LTC marker, CH_2O is a product of LTC and is often used in experimental studies to identify LTC regions, and CO_2 is a product of the overall combustion process. Figure 4 presents the scatter plot of mass fractions of these species against the values predicted by the ANN. Although some scatter is observed, the ANN predictions are largely concentrated near the DNS values as indicated by the mean error which is 11%, 18%, 13% and 30% for CO_2, OH, CH_2O and $C_{12}H_{25}O_2$, respectively. The PCA-ANN model thus predicts the major species with good accuracy. Table 2 lists the five species that incur the highest error in the PCA-ANN model. Also listed is the error from the optimal estimator (ϵ_{opt}). The results indicate that the species incurring highest errors are radicals. The highest error is 53% observed for CH_2^*, a short lived radical. Similar to the observations for reaction rates, compared with the optimal estimator, a significantly higher error is incurred by the PCA-ANN model for the prediction of minor species.

It should be noted that although a perfect prediction of the turbulent flame manifold is not expected, the results obtained from the PCA-ANN model are worse than the benchmark of the optimal estimator (best possible prediction for an LDM). This could be due to the following reasons. First, the canonical configuration used to train the ANN is not representative of the thermochemical manifold observed in the turbulent flame. Turbulence-chemistry interactions can cause the thermochemical manifold in the turbulent flame to deviate from that of laminar flames. Turbulence can lead to a large variation in local parameters like scalar dissipation rate, thus affecting the evolution of the thermochemical manifold. Using training data that accounts for the variation of such parameters can potentially improve predictions, especially for minor species and ω_η, which are very sensitive to parameter changes. This will be revisited in Section 3.4. Second, the ANN is not properly optimised for all species. Owoyele et al. [16] showed that the accuracy of the thermochemical basis predicted by the ANN can be increased if a different ANN is used for prediction of each species in contrast to using a single ANN for all species. This was the case in that study because the authors used the same ANN architecture for both cases. Fitting an individual ANN for each species is cumbersome for training, requires *ad hoc* assumptions for grouping species in clusters and can be computationally more expensive. Another advantage of using a single ANN for all species is that the constraint for mass conservation can be explicitly specified. This will be further discussed in Section 3.5. For the present case, tests with fitting an individual ANN for each species revealed that no appreciable improvement in the prediction error is observed. Moreover, the errors in the species mass fractions predicted by the ANN are close to the optimal estimator values (mean error of 20% vs 16% for the optimal estimator) which further indicates that the ANN used in this study is well optimised for all species.

3.4. Influence of more data

As discussed above, training the ANN with data representative of parametric changes in the DNS is expected to improve predictions. To test this, we repeat the analysis presented above but with ANNs trained using data from 2D flames at two different scalar dissipation rates ($4\ s^{-1}$ and $350\ s^{-1}$). A minor improvement is observed in this case for prediction of species mass fraction. The mean error is reduced from 20% to 18%. The maximum error also reduced from 53% to 47%. However, a major improvement is observed in the prediction of reaction rates. This is confirmed from the quantitative error values (ϵ_χ) obtained with new ANNs that are reported in Table 1. Since a different dataset (*i.e.* a combination of two 2D flames) is used here for computing η, the η obtained with the new dataset may not be the same as that obtained previously.

Thus, a one-to-one comparison of the error in ω_η may not be meaningful. However, as reported in Table 1, the error is reduced for all ω_η, which is the desired outcome. Figure 3(c) shows the error with the new ANN conditioned on the first two principal components. In comparison to Fig. 3(a), the error in the edge-flame region is considerably reduced, while a minor reduction is observed in the LTC region. Quantitatively, the error in the LTC region is reduced from 77% to 65% while in the edge-flame region a decrease in error from 43% to 22% is observed. This result is consistent with previous studies that show that HTC is more sensitive to χ than LTC [28]. Overall, this clearly shows that using additional data can significantly reduce the prediction errors. The advantage of the PCA-ANN approach is further apparent when using multiple datasets for tabulation. It provides a straightforward way of identifying LDMs and storing/retrieving data, which can be challenging in the conventional tabulation approach.

3.5. Conservation properties

A limitation of using ANN for tabulation is that it is not clear whether the predicted output satisfies mass conservation and gives physically consistent outputs (*e.g.* non-negative mass fractions). To ensure physically consistent results, the absolute values of the mass fraction predicted by the ANN were considered in this paper[2]. The output of the ANN is now checked for conservation of mass. The percentage error in mass conservation is defined as $\epsilon_m = 100 \times |1 - \Sigma_i Y_i|$. The mean ϵ_m across all samples predicted by the ANN for the turbulent flame is less than 0.05% while the maximum ϵ_m is 4%. Elemental mass conservation is also satisfied within a maximum 5% error.

To more stringently satisfy mass conservation, one can modify the ANN cost function to penalise deviation from mass conservation in addition to the mean squared deviation for mass fraction of each species. Our test with the modified cost function revealed that the maximum error in mass conservation of the ANN output can be reduced to $\approx 1\%$ by simply adjusting the contribution of the two components of the cost function without any appreciable decrease in prediction accuracy. The ANN can thus be tuned to give physically consistent outputs.

3.6. Discussion

The results show that the PCA-ANN approach is able to predict species mass fractions and thermochemical properties with good accuracy in a 3D turbulent jet flame at diesel engine conditions. However, high errors are observed in the prediction of the source terms. The errors reported in this work should be considered in the context of diesel engine combustion which presents, in comparison to previous PCA-ANN studies [8,11,14,15], much more significant modelling challenges due to its complex multiple combustion mode structure. It should also be considered that in the present work a 3D turbulent flame has been modelled using laminar flame data, a much more challenging problem than previous PCA studies which consider the same configuration for training and testing.

Nevertheless, the PCA-ANN approach showed significant improvements over the conventional tabulation approach which demonstrates its potential for combustion modelling in diesel engine conditions. As discussed earlier, consideration of training data with a variation of parameters (χ in the present paper) yields better results and future work will consider using multiple datasets to reduce the error within acceptable limits. It should also be noted that the errors have been presented for the prediction of instantaneous quantities. In *a posteriori* simulations, mean reaction rates are required, which could compensate for some of the errors observed for instantaneous quantities. Building upon the present work, future studies will consider these aspects.

4. Conclusions

A principal component (PCA) and artificial neural network (ANN) based chemistry tabulation approach was *a priori* evaluated using data from a recent direct numerical simulation (DNS) of a spatially developing round jet turbulent flame in diesel engine conditions. Three canonical configurations, namely, zero-dimensional (0D) ignition reactors, one-dimensional (1D) nonpremixed flamelets and two-dimensional (2D) laminar flames were used to train the PCA-ANN model. The performance of the PCA-ANN model was evaluated against the optimal estimator and the conventional tabulation approach. The main conclusions are listed below.

1. A qualitative comparison of the canonical configurations with the DNS revealed that the DNS flame structure is best represented by the 2D flame followed by the 1D flamelet. The 0D reactors fail to qualitatively represent the DNS structure.
2. The PCA-ANN model was found to exhibit higher accuracy while requiring several orders of magnitude less memory than a conventional tabulation approach.
3. The PCA-ANN model gave excellent predictions, comparable to the optimal estimator for the thermophysical properties and major

[2] The problem of negative mass fractions can although be alleviated by pre-processing mass fractions using a log-scaling, it introduces other problems like potential mass fraction values of greater than one upon rescaling. As the net benefit was unclear, log-scaling was not used.

species. However, higher errors were observed in the prediction of minor species and reaction rates.

4. The error in reaction rates predicted by the PCA-ANN model was found to be concentrated in the low-temperature chemistry and edge-flame regions. Training the PCA-ANN network with data with a variation of scalar dissipation rate was shown to reduce errors especially in the edge-flame region.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work was supported by the Australian Research Council (ARC, DP180103923 and DP150104393) and the Australian Renewable Energy Agency. Computational resources were provided by the Australian Government through the Pawsey Supercomputing Centre and the National Computational Infrastructure under the National Computational Merit Allocation Scheme, by ARC projects LE170100032 and LE160100002 managed by Intersect Pty Ltd, and by the University of New South Wales. Access to computing systems at the National Energy Research Scientific Computing Center provided by the United States Department of Energy is acknowledged.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.proci.2020.06.263.

References

[1] N. Peters, *Prog. Energy Combust. Sci.* 10 (3) (1984) 319–339.
[2] U. Maas, S.B. Pope, *Combust. Flame* 88 (3–4) (1992) 239–264.
[3] O. Gicquel, N. Darabiha, D. Thévenin, *Proc. Combust. Inst.* 28 (2) (2000) 1901–1908.
[4] J. Van Oijen, L. De Goey, *Combust. Sci. Technol.* 161 (1) (2000) 113–137.
[5] C.D. Pierce, P. Moin, *J. Fluid Mech.* 504 (2004) 73–97.
[6] B. Fiorina, O. Gicquel, L. Vervisch, S. Carpentier, N. Darabiha, *Proc. Combust. Inst.* 30 (1) (2005) 867–874.
[7] M. Bode, N. Collier, F. Bisetti, H. Pitsch, *Combust. Theor. Model.* (2019) 1–26.
[8] A. Biglari, J.C. Sutherland, *Combust. Flame* 162 (10) (2015) 4025–4035.
[9] J.C. Sutherland, A. Parente, *Proc. Combust. Inst.* 32 (1) (2009) 1563–1570.
[10] L. Vervisch, R. Hauguel, P. Domingo, M. Rullaud, *J. Turbul.* 5 (4) (2004) 1–8.
[11] T. Echekki, H. Mirgolbabaei, *Combust. Flame* 162 (5) (2015) 1919–1933.
[12] B.A. Sen, E.R. Hawkes, S. Menon, *Combust. Flame* 157 (3) (2010) 566–578.
[13] L.L. Franke, A.K. Chatzopoulos, S. Rigopoulos, *Combust. Flame* 185 (2017) 245–260.
[14] M.R. Malik, B.J. Isaac, A. Coussement, P.J. Smith, A. Parente, *Combust. Flame* 187 (2018) 30–41.
[15] B.J. Isaac, J.N. Thornock, J. Sutherland, P.J. Smith, A. Parente, *Combust. Flame* 162 (6) (2015) 2592–2601.
[16] O. Owoyele, P. Kundu, M.M. Ameen, T. Echekki, S. Som, *Int. J. Eng. Res.* (2019). 1468087419837770
[17] D.K. Dalakoti, B. Savard, E.R. Hawkes, A. Wehrfritz, H. Wang, M.S. Day, J.B. Bell, *Combust. Flame* 217 (2020) 57–76.
[18] T. Yao, Y. Pei, B.-J. Zhong, S. Som, T. Lu, K.H. Luo, *Fuel* 191 (2017) 339–349.
[19] H. Pitsch, N. Peters, *Combust. Flame* 114 (1–2) (1998) 26–40.
[20] H. Pitsch, Flamemaster, a C++ computer program for 0D combustion and 1D laminar flame calculations, 1998.
[21] D.K. Dalakoti, A. Krisman, B. Savard, A. Wehrfritz, H. Wang, M.S. Day, J.B. Bell, E.R. Hawkes, *Proc. Combust. Inst.* 37 (2) (2019) 1961–1969.
[22] A. Parente, J.C. Sutherland, *Combust. Flame* 160 (2) (2013) 340–350.
[23] R. Vicquelin, B. Fiorina, S. Payet, N. Darabiha, O. Gicquel, *Proc. Combust. Inst.* 33 (1) (2011) 1481–1488.
[24] H. Mirgolbabaei, T. Echekki, *Combust. Flame* 161 (1) (2014) 118–126.
[25] A. Moreau, O. Teytaud, J.-P. Bertoglio, *Phy. Fluids* 18 (10) (2006) 105101.
[26] B. Savard, H. Wang, A. Wehrfritz, E.R. Hawkes, *Proc. Combust. Inst.* 37 (4) (2019) 4655–4662.
[27] D. Dalakoti, *Direct numerical simulation of lifted flames in diesel engine conditions*, UNSW, 2019 Ph.D. thesis.
[28] A. Krisman, E.R. Hawkes, M. Talei, A. Bhagatwala, J.H. Chen, *Proc. Combust. Inst.* 36 (3) (2017) 3567–3575.