Supplementary material for D. K. Dalakoti, A. Wehrfritz, B. Savard, M. S. Day, J.B. Bell and E. R. Hawkes, An *a priori* evaluation of a principal component and artificial neural network based combustion model in diesel engine conditions, Proc. Combust. Inst. 38, 2020

1 ANN architecture

The architecture of the neural networks used in this paper is listed in Table 1. A rectified linear unit activation function [1] was used for all layers except the output layer where a linear activation function is used. A mean squared error cost function was used.

Quantity	Input (η)	Hidden	Output
Y_i, T	5	110×110×110	54
ω_η	5	20×20	1
C_p, λ, μ and D_η	5	10×10	1
$\overline{W}$	5	10×10	1

Table 1: ANN architecture for various quantities. Note that separate ANNs are used for each ω_η.

2 Table look-up

For the results presented in Section 3.2, tables with a uniformly-spaced grid were considered. Linear interpolation was used for table look-up. A sensitivity analysis of the number of grid points (N) on the error incurred by the lookup and interpolation of the tabulated heat release rate was carried out. Figure S1 presents the prediction error in heat release rate for Tab_{ξ,Y_c} and Tab_{PCA} with different N. The figure shows that the error converges for both tables as a function of N.

3 Prediction of thermophysical properties

Table 2 lists the error incurred in the prediction of the thermophysical properties by ANNs trained using the three canonical configurations. The results for the optimal estimator are also presented. The results show that

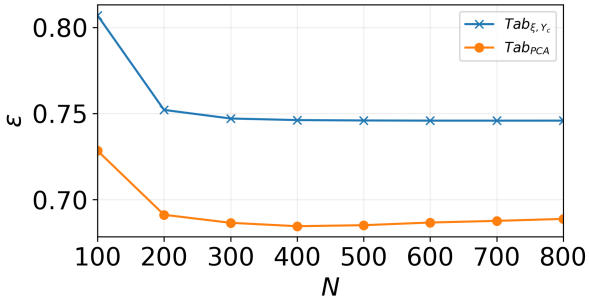

Figure S1: Prediction error in heat release rate for the two tables as a function of N.

the 2D flame incurs the least error followed by 1D flamelets and 0D reactors. Moreover, the error from the 2D flame is comparable to the optimal estimator.

Quantity	ϵ_{opt}	ϵ_{2D}	ϵ_{1D}	ϵ_{0D}
C_p	3.1%	3.8%	14%	15%
λ	3.8%	5%	6.32%	10.9%
μ	3.8%	3.85%	7%	16%
$\overline{W}$	3.2%	3.4%	30%	31%
D_η	1.4%	3.8%	-	-

Table 2: Error in the ANN prediction of the transport coefficients, C_p and $\overline{W}$ for the optimal estimator (ϵ_{opt}), 2D flame (ϵ_{2D}), 1D flamelet (ϵ_{1D}) and 0D reactor (ϵ_{0D}) case. Principal components from 2D flame are used for calculation of optimal estimators.

4 Prediction of thermochemical basis

The error in the prediction of the species mass fraction incurred by the ANN trained using data from 2D flame (ϵ_{2D}), 1D flamelet at scalar dissipation rate, $\chi = 20\,\mathrm{s}^{-1}$ (ϵ_{1D}) and 0D reactors (ϵ_{0D}) is presented in Table 3. Also presented are the results obtained with an ANN trained using 1D flamelets with a parametric variation of scalar dissipation rate (ϵ_{1D_χ}). Eight flamelets with χ in the range $1\,\mathrm{s}^{-1}$ to $35\,\mathrm{s}^{-1}$ were considered. This range was selected to represent the χ values observed at the flame base in the turbulent flame. The results with the optimal estimator (ϵ_{opt}) are also included for reference.

	ϵ_{opt}	ϵ_{2D}	ϵ_{1D}	ϵ_{1D_χ}	ϵ_{0D}
Mean	16%	20%	25%	20%	98%
Median	10%	17%	19%	18%	85%

Table 3: Error in the prediction of species mass fractions from ANNs trained using data from various canonical configurations.

The best performance is observed with the 2D flame database. Homogeneous reactors exhibit the worst performance. Consistent with the observation made in the paper, using data with a parametric variation of scalar dissipation rate (ϵ_{1D_χ}) to train the ANN improves accuracy.

5 Prediction of reaction rates

The error in the prediction of reaction rates incurred by ANNs trained using data from 2D flame (ϵ_{2D}), 1D flamelet at $\chi = 20\,\mathrm{s}^{-1}$ (ϵ_{1D}) and 0D reactors (ϵ_{0D}) is presented in Table 3. Also presented is the result obtained with a parametric variation of χ for 1D flamelets (ϵ_{1D_χ}). Consistent with the observations in Section 4, the best performance is obtained with the 2D flame followed by 1D flamelets. Errors reduce when using data with a parametric variation of χ.

ω_η	ϵ_{opt}	ϵ_{2D}	ϵ_{1D}	ϵ_{1D_χ}	ϵ_{0D}
ω_{η_1}	25%	42%	69%	79%	288%
ω_{η_2}	30%	50%	54%	36%	228%
ω_{η_3}	26%	46%	44%	31%	338%
ω_{η_4}	28%	56%	90%	29%	370%
ω_{η_5}	10%	85%	45%	42%	364%

Table 4: Error in the prediction of ω_η from ANNs trained using data from various canonical configurations.

6 R^2 scores

A normalised mean squared error metric is used to report the performance of models in the paper. Here we present the model performance in terms of the

coefficient of determinant (R^2 score). The R^2 score is computed according to equation 1 as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{p,i} - y_{DNS,i})^2}{\sum_{i=1}^{n}(y_{DNS,i} - \overline{y}_{DNS})^2}, \tag{1}$$

where y_p are the values predicted by the model, y_{DNS} are the values from the DNS and n is the number of samples. The mean and median R^2 scores of the ANN prediction of species mass fractions are 0.965 and 0.937, respectively. The R^2 scores for the 5 species which incur the highest error are listed in Table 5 for the ANN (R^2_{ANN}) and the optimal estimator (R^2_{opt}).

Species	R^2_{ANN}	R^2_{opt}
$CH_2{}^*$	0.685	0.774
CH_2	0.692	0.768
C_5H_9	0.707	0.824
pC_4H_9	0.793	0.901
$OC_{12}H_{23}OOH$	0.800	0.860

Table 5: R^2 scores for the 5 species with the highest error.

The R^2 scores for the ANN prediction of reaction rates is presented in Table 6 along with the results for the optimal estimator. Also presented are the results obtained with a training dataset containing two 2D flames with different inlet χ.

ω_η	R^2_{ANN}	R^2_{opt}	R^2_χ
ω_{η_1}	0.824	0.937	0.851
ω_{η_2}	0.734	0.909	0.869
ω_{η_3}	0.790	0.931	0.891
ω_{η_4}	0.693	0.920	0.867
ω_{η_5}	0.298	0.989	0.663

Table 6: R^2 scores for ω_η.

References

[1] R. H. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, H. S. Seung, Nature 405 (2000) 947.