Computers and Fluids 314 (2026) 107099

Contents lists available at ScienceDirect

Computers and Fluids

journal homepage: www.elsevier.com/locate/compfluid

An adaptive-learning framework for chemistry tabulation in turbulent reacting flows☆

Kai Niemietz [a,*], Clément Scherding [b], Walter Gemassmer [a], Taraneh Sayadi [a,c], Heinz Pitsch [a]

[a] Institute for Combustion Technology, RWTH Aachen University, Aachen, 52056, Germany
[b] Institut Jean le Rond d'Alembert, Sorbonne University, Paris, 75005, France
[c] Mathematical and Numerical Modelling Laboratory, Conservatoire National Arts et Metiers, Paris, 75003, France

ARTICLE INFO

Dataset link: https://github.com/cscherding/RONAALP

Keywords:
Reduced-order model
Machine learning
Adaptive learning
Turbulent combustion
Flame-wall interaction

ABSTRACT

Many applications in turbulent combustion rely on some form of reduced-order model. Flamelet-type models with pre-computed, tabulated chemistry are widely used and often perform well. However, the creation of an appropriate and accurate table requires considerable knowledge and time. Data-driven models have shown a lot of promise in reducing this workload. They are fast, simple to use, and can provide high accuracy. But the model is often a black box. Moreover, data-driven extrapolation to new, unknown data is not obvious; a feature required in most applications.

In this work, we implemented the RONAALP algorithm (Reduced Order Nonlinear Approximation with Active Learning Procedure) which combines offline training with an adaptive learning approach to handle extrapolation events on-the-fly. The offline training relies on a combination of nonlinear auto-encoders, community clustering, and radial basis function networks to develop an efficient and compact base model from limited training data. The active learning procedure expands the base model to consider extrapolation outside the original training data. The algorithm was developed and validated for non-equilibrium hypersonic flows. We have adapted the algorithm to the turbulent reacting wall-bounded flows typical of gas turbine combustion. The model is trained on inexpensive 1D laminar flame data and assessed using a turbulent 3D DNS database.

The base model performs as well as conventional chemistry tabulations methods. Due to the built-in extrapolation procedure, adaptive learning with additional training data from strained counterflow flamelets and head-on quenching flamelets improves the predictions.

1. Introduction

The design and optimization of next-generation combustion systems will heavily rely on numerical simulations. Among the many available numerical frameworks, direct numerical simulations (DNS) are often employed to provide fundamental physical insights [1]. However, DNS are computationally prohibitive for practical engineering applications, making lower-cost modeling strategies, such as reduced-order models (ROMs), a viable solution due to the balance between computational efficiency and physical accuracy. The ROM approach requires the definition of a set of control variables that account for the effects of local thermo-chemical state variations, along with models for turbulent combustion closure and representation of chemical kinetics. In this context, chemistry representation is often addressed through the use of chemistry manifolds and tabulation techniques. Chemistry manifolds are generated by solving one-dimensional flamelets with detailed chemistry and detailed transport, thereby considering the flame structure. These manifolds are produced prior to the actual computation, shifting the cost of detailed chemistry into a preprocessing step. Successful examples include the flamelet-progress variable (FPV) approach [2] and flamelet generated manifolds (FGM) [3], both of which have been applied across a variety of combustion configurations.

For more complex configurations, such as flame-wall interaction (FWI), multi-phase flows, pollutant formation, or turbulent strain, the tabulation becomes more challenging. Knudsen et al. [4] tabulated strained counterflow flamelets to include the effects of turbulent strain. Efimov et al. [5] developed the quenching flamelet-generated manifold

☆ This article is part of a Special issue entitled: 'fusing data and physics' published in Computers and Fluids.

* Corresponding author.
E-mail address: k.niemietz@itv.rwth-aachen.de (K. Niemietz).

https://doi.org/10.1016/j.compfluid.2026.107099
Received 1 June 2025; Received in revised form 13 January 2026; Accepted 17 April 2026
Available online 21 April 2026
0045-7930/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

(QFM), which uses head-on quenching (HOQ) flamelets to improve the modeling of FWI. The QFM model was extended by Steinhausen et al. [6] to include interactions with exhaust gases due to flame-vortex interactions. As multiple physical phenomena must be captured, the size of the chemistry manifold increases significantly, affecting the computational cost. Furthermore, as more physical parameters are introduced, the control variables may become correlated, making the table lookup more complex and less efficient [7].

In this work, a reduced-order chemistry model is applied to a turbulent 3D DNS database relevant to gas turbine combustion [8]. The database includes several features that complicate the modeling: the domain is enclosed by isothermal walls on two sides, causing FWI, turbulent strain affects the flame, and the fresh mixture interacts with burnt gases due to recirculation. All of these aspects affect the flame characteristics and the pollutant formation and must be considered in a ROM. Due to the lean mixtures and low temperatures, local quenching of the combustion process can occur, resulting in increased carbon monoxide (CO) emissions [9]. One of the mechanisms causing this behavior is linked to the fact that the oxidation from CO to CO_2 is disturbed by FWI [10]. The non-equilibrium CO is kinetically controlled, making predictions difficult and highly accurate models a necessity. While several conventional tabulation models exist, e.g., [5–7], the combination of many influences remains a challenge, which data-driven methods can help resolve.

Significant advances in modern deep learning (DL) frameworks, combined with the availability of comprehensive DNS datasets, have motivated the exploration of data-driven DL-based tabulation methods, which are particularly effective at capturing strongly nonlinear interactions [11]. Several studies have investigated manifold identification [12,13] and the representation of quantities of interest [14,15], particularly the source term of the progress variable, $\dot{\omega}_C$, in turbulent combustion systems. Bissantz et al. [15] developed a self-organizing artificial neural network (ANN) capable of mapping complex flamelet structures onto lower-dimensional latent spaces, including a manifold identification step. Kumar et al. [16] demonstrated through *a priori* testing that autoencoder-based ANNs can effectively learn low-dimensional combustion manifolds directly from high-fidelity DNS data, offering efficient tabulation strategies and enabling faster simulations. Similarly, Chi et al. [17] incorporated an ANN into a finite-rate chemistry DNS solver, where the network was trained online using thermochemical states generated by the ODE solver. As the simulation progressed, increasing portions of the chemical source term evaluation were offloaded to the ANN, leading to notable improvements in computational efficiency without compromising accuracy. While traditional ANN architectures are generally used for such tasks, convolutional neural networks (CNNs) – particularly those with encoder–decoder structures like U-Nets – have gained traction due to their ability to effectively model spatial structures inherent in combustion data [18]. For example, Arumapperuma et al. [19] applied a U-Net architecture to assess the extrapolation performance of CNN-based models, demonstrating strong generalization to out-of-distribution conditions across different Reynolds numbers and filter configurations. Despite these promising results, the application of CNNs to complex reacting flow configurations remains limited. This is largely due to challenges in integrating CNNs with unstructured meshes or complex boundary conditions typical in practical combustion systems [20]. Consequently, ANN-based models – especially those relying on global thermo-chemical states rather than spatial features – remain an important option in many high-fidelity combustion modeling applications, owing to their greater flexibility and ease of integration [21].

In addition to their flexibility, ANNs provide two key advantages over conventional tabulation methods and CNN-based approaches: reduced memory requirements and smooth function representation [21]. These benefits become particularly significant for high-dimensional tabulations, which are often required in complex chemistry and flow problems. Unlike tabulation methods that scale poorly with dimensionality, ANNs maintain a relatively small memory footprint while preserving accuracy across a wide range of operating conditions. This advantage was quantified by Ihme [21] as the knowledge density of ROMs, defined as the ratio of accuracy to memory requirement. Higher knowledge density values highlight the superior efficiency of ANNs in capturing essential combustion dynamics without incurring prohibitive memory costs. Moreover, the performance of ANN models is further enhanced when physical knowledge is embedded into the training process [11]. For instance, Bissantz et al. [15] enforced boundedness by correcting source terms at manifold boundaries, reducing extrapolation errors. Such physics-based constraints are easier to implement in ANN models due to their point-wise formulation, where each state is treated independently—allowing straightforward enforcement of physical laws directly in the output or loss function. A central open challenge of such modeling approaches is ensuring the extrapolation beyond their training distribution. While exhaustive training on large datasets can help mitigate extrapolation errors by covering a broad range of thermo-chemical states, the inherently unsteady and multiscale nature of turbulent combustion often leads to operating conditions not represented in the training data [22], reducing accuracy in regions underrepresented in the training data or of lesser physical significance [18,22]. At the same time, generating such comprehensive datasets quickly becomes unfeasible, especially for high-fidelity models. A possible workaround is to employ adaptive modeling techniques that update the model during runtime, allowing it to learn from new data as it arises. One example is the *in situ* adaptive tabulation (ISAT) algorithm from Pope [23], which dynamically augments its tabulated database based on an error criterion. Although promising, such on-the-fly training remains challenging due to balancing adaptation speed with stability, and defining robust error metrics.

Another approach is training several ANNs for different aspects of the problem, as proposed by Chatzopoulos and Rigopoulos [24] and Franke et al. [25]. Small, lightweight ANNs can be trained quickly and capture the relevant features of their sub-problem with greater accuracy. This makes them particularly attractive for complex configurations like turbulent FWI, where different physical phenomena, such as heat transfer, chemical reactions, and turbulence, need to be correctly accounted for. This approach introduces challenges that are unexplored in the context of FWI modeling, as identifying and dividing the problem into suitable sub-domains is not always straightforward. Moreover, the number of necessary sub-models can grow rapidly as the complexity of the problem increases, leading to larger computational overhead.

In this work, the second type of approach is adopted, where multiple lightweight ANNs are trained to capture different features in the domain. However, the clustering into sub-domains is performed in a low-dimensional latent space, which reduces the necessary number of sub-models, cf. [12]. The RONAALP model (Reduced Order Nonlinear Approximation with Active Learning Procedure) was originally developed to accelerate chemistry calculations in hypersonic non-equilibrium boundary layer flows by replacing the chemistry solver with ML methods [26,27]. The overall accuracy was retained in both *a priori* and *a posteriori* investigations, while significantly reducing the time to solution. Here, we apply the RONAALP model to turbulent reactive flows. In contrast to the previous application and to most ANN approaches in the literature, the model's training data is not extracted from DNS. Instead, the model is trained on inexpensive, low-dimensional training data (1D flamelets). The trained model is then evaluated on a turbulent 3D DNS database relevant to gas turbine combustion [8].

2. DNS database and training data

2.1. DNS database

The DNS database considered in this work consists of two parallel turbulent premixed jet flames separated by a laminar pilot of burnt

Table 1
Simulation parameters, initial and boundary conditions. s_L is the laminar burning velocity and l_F is the thermal flame thickness in the laminar unstretched flame.

$L_x \times L_y \times L_z$ (h_j)	$100 \times 12 \times 6$
$\Delta_x \times \Delta_y \times \Delta_z$ (μm)	$20 \times 10 \times 20$
φ	0.5
p (bar)	4
T^u (K)	673
T^b (K)	1782
T_w (K)	1000
u_j (m/s)	73.5
h_j (mm)	1.2
Re_j	5500
s_L (cm/s)	27.37
l_F (μm)	205

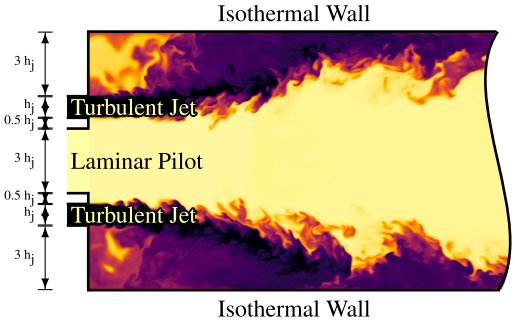


Fig. 1. Sketch of the simulation domain. Color indicates the temperature field. The inflow consists of two turbulent jets with a laminar pilot in the center. The outer borders of the inlet plane are modeled as adiabatic walls.
Source: Adapted from [8].

gases and enclosed by isothermal walls. A sketch of the domain is presented in Fig. 1. The database was designed to contain features relevant to gas turbine combustion. The fresh gas is a methane/air mixture with an equivalence ratio of $\varphi = 0.5$ at a temperature of $T = 673$ K and an elevated pressure of $p = 4$ atm. The bulk inlet velocity of the jets is $u_j = 73.5$ m/s and the jet width is $h_j = 1.2$ mm resulting in a jet Reynolds number ($Re_j = u_j h_j / \nu$) of 5500, where ν is the kinematic viscosity of the reactants. A summary of the relevant parameters is provided in Table 1. Finite-rate chemistry is employed, using a skeletal methane mechanism with 25 species [28], which was derived from the full mechanism of Cai et al. [29].

The crosswise direction is bounded by no-slip, isothermal walls with a temperature of $T_w = 1000$ K, and recirculation regions drive mixing and interactions between fresh and (cooled) burnt gases. The outer edges of the inlet are adiabatic walls, creating the spreading behavior observed in the jets. The instantaneous temperature field is shown in Fig. 2. The flames spread out along the cold walls with significant FWI. Downstream, cooled fluid from the walls is entrained back into the main flow. A detailed description of the DNS is found in [8].

2.2. Model input data

A caveat of machine-learning models is the necessity of large training data sets. The training data should be as close as possible to the final application to generate an accurate model. Thus, a typical workflow includes the generation of training data specific to the desired application. In the present case, we generate an ensemble of 1D adiabatic flamelet simulations using the code FlameMaster [30]. We use the progress variable C, defined via the major product species as

$$C = Y_{CO} + Y_{CO_2} + Y_{H_2} + Y_{H_2O}. \quad (1)$$

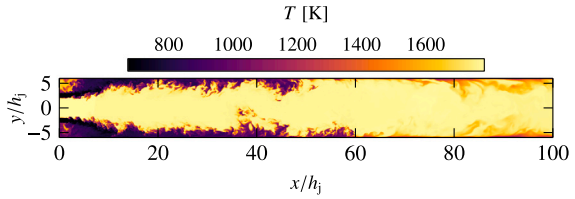


Fig. 2. Instantaneous temperature field from the DNS. The flames are elongated by the FWI. Cooled burnt gases near the walls are entrained into the main flow downstream of the flame.

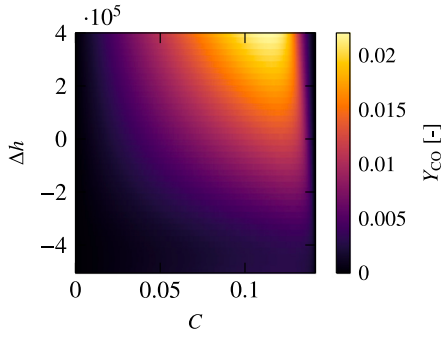


Fig. 3. Ensemble of flamelets for the input data. The x-axis is given by the combustion progress variable C and the y-axis by the enthalpy defect Δh. Coloring indicates Y_{CO}.

Additionally, we must account for the isothermal walls and the resulting heat loss, which is done by tracking the enthalpy in the fluid. This enthalpy loss must be considered in the model input data. The unburnt temperature of the flamelets is varied to capture the variations in enthalpy. The enthalpy is tabulated as the enthalpy difference between the various flamelets and the nominal enthalpy at the inlet conditions: $\Delta h = h_{FL} - h^u$.

In total, this results in about 80 flamelets with 1000 data points each as model input data. An example of the input data is shown in Fig. 3. The CO mass fraction Y_{CO} is parameterized by the progress variable C and the enthalpy defect Δh.

3. RONAALP: Deep-learning framework

The reduced-order model used in this investigation has been developed by Scherding et al. [26,27] for simulations of hypersonic non-equilibrium flows. Very high velocities and friction cause large temperature increases and a decomposition of the air molecules. The RONAALP model was implemented to replace the finite-rate chemistry solver, which was responsible for a large portion of the computational load. The overall accuracy was retained in both *a priori* and *a posteriori* investigations. Furthermore, the model did not produce spurious values or other errors and the solution remained stable after more than two flow-through times. In total, the computational time was reduced by 70 %. Only a short recap is provided here; further results, analysis, and discussions are available in [26,27].

The model is made up of three main components: First, an input/output encoder reduces the dimensions of the input data. Second, community clustering is applied to detect different clusters in the low-dimensional manifold. Third, individual surrogate models are trained for each cluster. All three components are necessary to produce high-quality output data and are analyzed individually in the following.

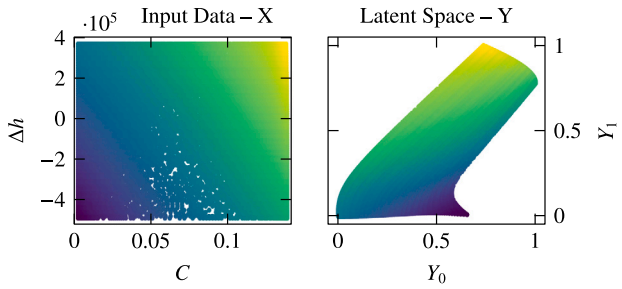


Fig. 4. Schematics of the temperature field in the input layer and the low-dimensional representation in the latent space.

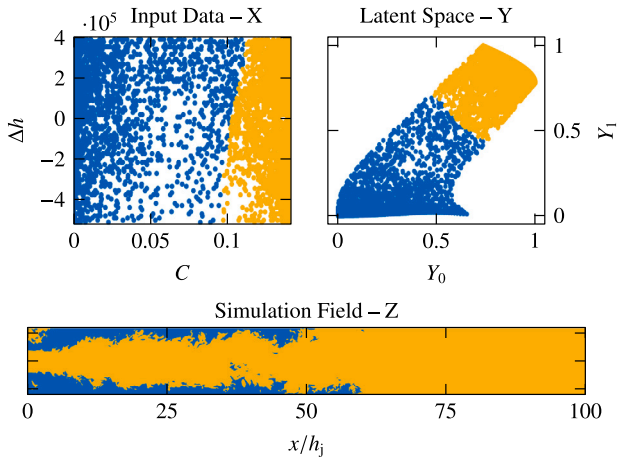


Fig. 5. Clustering in input layer, latent space, and output layer. Two clusters are detected: the cold unburnt region and the hot burnt region.

3.1. Dimensionality reduction

The first step in the RONAALP methodology is a dimensionality reduction from a typically high-dimensional input space to a low-dimensional normalized latent space. Many algorithms exist to accomplish this, e.g., principal component analysis (PCA) or auto-encoders (AEs). Here, we are using a modified variant of an AE, called input/output-encoder (IO-E) [26]. AEs are non-linear, parametric models (i.e. deep neural networks) which typically enable higher compression rates than linear methods.

An AE encodes a high-dimensional input $\mathbf{X}$ onto a low-dimensional latent space $\mathbf{Y}$ and then decodes it back to a representation of the input $\hat{\mathbf{X}}$. The model weights are trained by minimizing the L_2 error between input and representation: $min\left(\|\mathbf{X}-\hat{\mathbf{X}}\|_2\right)$. However, since we are interested in a model that best predicts the output $\mathbf{Z}$ for a given input $\mathbf{X}$, it is convenient for the low-dimensional representation to contain information from the model output prediction $\hat{\mathbf{Z}}$. Instead of training an AE via the input space, the IO-E is trained via back-propagation of the reconstruction error of the output ($\|\mathbf{Z}-\hat{\mathbf{Z}}\|_2$), as a function of the input ($\|f(\mathbf{X})-\hat{\mathbf{Z}}\|_2$). For a full description of the IO-E and a comparison with other dimensionality reduction techniques the reader is directed to Scherding et al. [26]. As with AEs, this method creates a two-part model: the encoder and the decoder. Only the encoder is used in the final model. Although the IO-E could be used to predict the output, the decoder accuracy is not optimal. Instead, the clustering and surrogate model presented in the next sections are used. An example of the dimensionality reduction is given in Fig. 4. The left image shows the temperature field in the input layer with progress variable C and enthalpy defect Δh. The right image shows the corresponding latent space representation with the two parameters Y_0 and Y_1. In the current case, the input data is already two dimensional and remains so. Nonetheless, the IO-E compresses the representation and normalizes the field. This step becomes more relevant with higher dimensional inputs.

3.2. Community clustering

The second component of the RONAALP framework is community clustering. In turbulent combustion, many different physical effects must be considered, and the leading-order physics can strongly differ throughout the domain, e.g., the cold unburnt fuel/air mixture, the hot exhaust gases, or the chemistry-dominated flame front. By clustering the data into distinct communities, individual detailed surrogate models can be trained for each of the communities instead of one global model that must consider many distinct physical phenomena. The clustering is performed in the latent space with Newman's spectral algorithm [31]. The algorithm clusters the datapoints by their relative Euclidean distance for a given threshold ϵ. The threshold is typically defined as a fraction of the global mean distance. The major advantage of Newman's algorithm in this case over, for instance, k-means is that the number of clusters is not defined *a priori*. In complex turbulent flows, it is very difficult, if not impossible, to define the relevant number of clusters that best represent the physical phenomena *a priori*.

The result of the clustering is presented in Fig. 5. The figure shows two clusters detected by the algorithm and their representation in input ($\mathbf{X}$), latent ($\mathbf{Y}$), and output ($\mathbf{Z}$) space. Clustering is based on the latent space representation, input and output are included to give physical meaning to the clusters. The input data shows the two clusters formed mainly as a function of the progress variable C with a slight gradient from the enthalpy levels. As expected, in the latent space, where the clustering is performed, both clusters form distinct, compact regions. In the output space, the clusters follow the flame surface closely: one cluster in the unburnt region and one cluster in the burnt region.

3.3. Surrogate model

Finally, surrogate models are constructed using radial basis function (RBF) networks [32]. One individual network is trained for each of the clusters, which allows a higher fidelity representation of the relevant physics for each region. RBF networks are a special class of 3-layer networks. They enable sequential learning [33,34] and growing and pruning [35], making them convenient for adaptive learning algorithms. Many other algorithms could be used to build such surrogate models. RBF provide a very good trade-off between accuracy and performance as well as an easy training step. The network has one hidden layer with N_R centers $\boldsymbol{x}^c$ and N_R weights a_i. The RBF interpolant g is given by

$$g(\boldsymbol{x},\phi)=\sum_{i=1}^{N_R} a_i\phi\left(\|\boldsymbol{x}-\boldsymbol{x}_i^c\|\right), \tag{2}$$

with a kernel function ϕ. The N_R centers $\boldsymbol{x}_i^c$ are determined with k-means of the composite input/output space and the optimal weights a_i are found through the direct solution of a linear system, instead of iterative methods.

4. Results

The model must predict several fluid and chemistry quantities, such as temperature, density, viscosity, etc. Here, we evaluate two quantities that are particularly relevant to the model's performance. First, the source term of the progress variable $\dot{\omega}_C$, which is unclosed in the transport equation for the progress variable. An accurate model for $\dot{\omega}_C$ is essential for the correct representation of the flame location and the model's overall stability. Systematic errors in the prediction of $\dot{\omega}_C$ may lead to an erroneous flame surface, while nonphysical values will

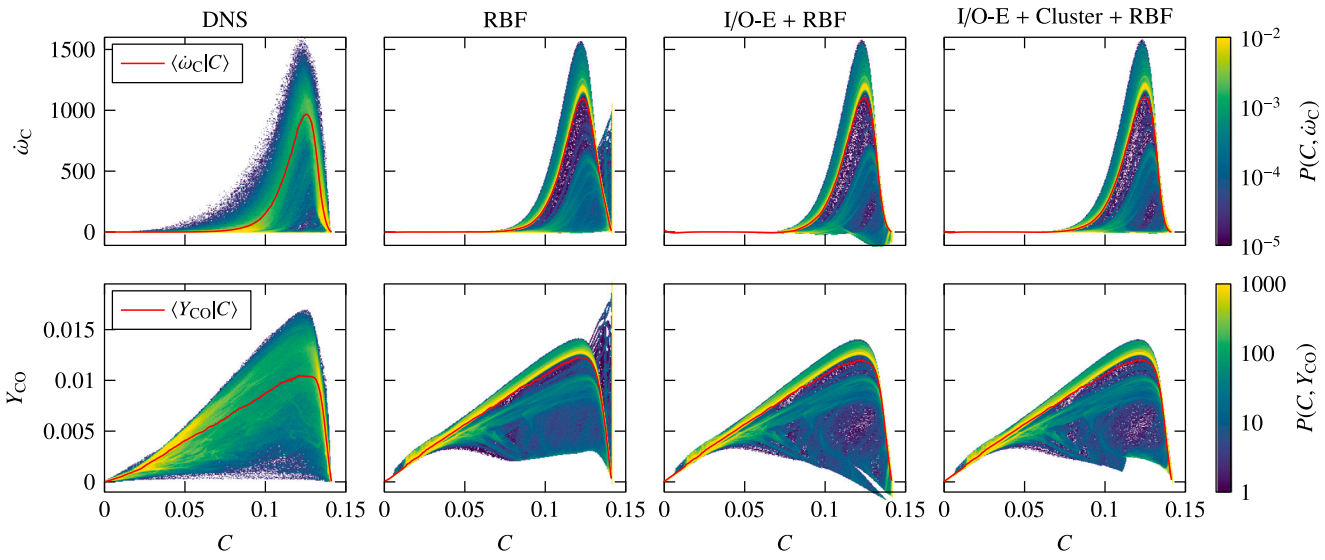


Fig. 6. Joint PDF of the progress variable source term ($\dot{\omega}_C$, top row) and the CO mass fraction (Y_{CO}, bottom row) over the progress variable C. Effects of the three model components are compared DNS data. Left to right: DNS data; model prediction using only the surrogate model (RBF); model predictions including the dimensionality reduction with input/output encoder and the surrogate model (I/O-E + RBF); and the full model with dimensionality reduction, clustering, and one surrogate model per cluster (I/O-E + Cluster + RBF).

reduce the model's stability. The second output investigated here is the mass fraction of CO, Y_{CO}, as it is the main quantity of interest. Further results for other fields, i.e., temperature and major species, are available in the Appendix.

As a first step in evaluating the model results, the three model components are evaluated individually. This allows us to assess their respective contributions and interactions. Joint probability density functions (jPDFs) of $\dot{\omega}_C$ and Y_{CO} over C are presented in Fig. 6, along with their conditional means. The jPDF from DNS data is compared to the three model components. "RBF" indicates that only a single surrogate model was used, without data preprocessing. "I/O-E + RBF" includes the dimensionality reduction, but no clustering. "I/O-E + Cluster + RBF" is the full model with all components. The model inputs for all model variants are the progress variable C and the enthalpy defect Δh. Comparing the DNS data to RBF, the overall shape of the distribution is recreated, though the values are more focused in a narrow band. This narrow band corresponds to the laminar flamelet solution at inlet conditions, which is overrepresented in the result. The concentration of values influences the conditional mean which is slightly overpredicted by the model. While the prediction is not perfect, it is not expected to negatively impact the solution stability of the model. A greater challenge to the stability are the spurious values at high progress variable values. A very similar behavior is observed in the CO mass fraction. Two differences occur in the CO prediction that were not present in the $\dot{\omega}_C$ data. First, the model in unable to reproduce the height of the intermediate CO peak. Second, in the DNS data, the formation of CO is suppressed early in the combustion ($C < 0.05$) which leads to a reduced conditional mean throughout the flame. This reduction is not present in the model prediction which remains closer to flamelet conditions. The spurious values are removed by including the I/O-E. The more compact representation in latent space reduces aliasing errors. However, negative values appear in the progress variable source term and the CO mass fraction, which may negatively impact the model's stability. The conditional means remain essentially unchanged to the previous model formulation. With the full model both the spurious values and the negative values are removed. The classification into two clusters and the training of separate surrogate models for each cluster allows the model to capture the relevant physical aspects more accurately.

As a first test case, the model performs reasonably well. While the accuracy of the model should be improved, overall trends are well captured, and the global flame behavior can be predicted. Most importantly, for ML-models, no unexpected or nonphysical values were observed with the full model.

4.1. RONAALP vs. Conventional tabulation

The training data for the model is composed of freely propagating flamelets. Turbulence effects, such as strain, stretch, or turbulent mixing, are not included in the training data. Accordingly, a perfect representation should not be expected. As a measure to evaluate the RONAALP model, the results are compared to a conventional chemistry tabulation with the same table parameters [7]. Fig. 7 shows the jPDF of $\dot{\omega}_C$ and Y_{CO} for the RONAALP model and the conventional chemistry tabulation model (ChemTable) with almost identical results. In the conditional means, the same behavior as before is observed. The peak value of $\langle\dot{\omega}_C|C\rangle$ is overpredicted by both models, though the RONAALP model performs slightly better than the conventional tabulation. This is caused by the distribution of intermediate values, which is a little lower in the RONAALP model than in the ChemTable. Both models overestimate the production of CO, while the consumption is captured more accurately. While both models perform similarly well, the additional advantage of the ML approach is its ease of use. The tabulation of many flamelets can be time-consuming and complex, particularly if flamelets in different configurations are used, e.g. freely propagating and counterflow flamelets.

4.2. Turbulent strain

Turbulent strain influences the flame front and global flame characteristics. It also has a pronounced impact on combustion chemistry and particularly CO reacts sensitively to strain [36]. Accordingly, strain effects must be considered when modeling CO in turbulent flames [8].

One challenge with turbulent strain in an LES context is how to calculate strain at the sub-grid level, where the velocity fluctuations are not resolved. Knudsen et al. [4] have shown that both the H and the OH radical are able to parameterize strain; both react sensitively and monotonically to increasing strain. In this investigation, OH is more relevant because it is also the main reaction partner for CO oxidation:

$$CO + OH \rightleftharpoons CO_2 + H. \quad (3)$$

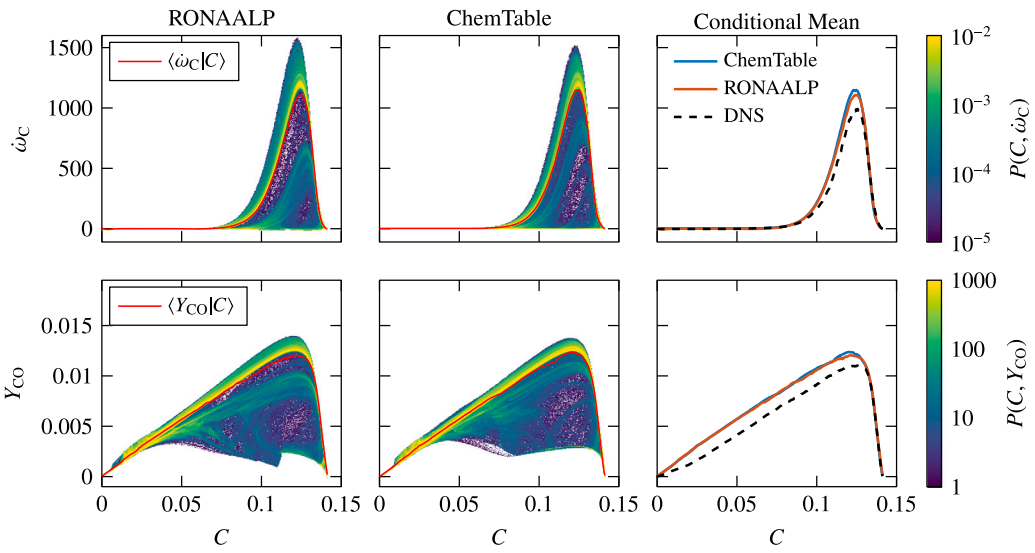


Fig. 7. Joint PDF of the progress variable source term ($\dot{\omega}_C$, top row) and the CO mass fraction (Y_{CO}, bottom row) over the progress variable C. RONAALP model compared to a conventional ChemTable chemistry tabulation. Included are the conditional means of all jPDF.

The OH mass fraction Y_{OH} is included as a third input parameter by calculating additional strained counterflow flamelets with varying strainrates. The strainrate range is chosen to correspond to the turbulent strain calculated from the DNS data; details are available in [8]. The counterflow fresh-to-burnt configuration was recommended by Hawkes and Chen [36] as the most relevant to turbulent premixed combustion.

Instead of combining the unstretched freely propagating flamelets and the strained counterflow flamelets into a single training data set, the previous unstrained model is updated with the new counterflow flamelet data. This adaptive learning approach has shown better stability when adding new data.

With the additional input of Y_{OH}, the clustering step introduces a third cluster. The three-cluster representations of input, latent, and output space are presented in Fig. 8. Additionally, a close-up of the flame surface near the inlet is provided for a more detailed view. The new cluster is shown in red. In the input data, the third cluster includes the low enthalpy, high progress variable data—the cold, burnt regions. The cluster also extends into higher enthalpy regions, though not towards the fully burnt, forming a boundary between the previous clusters (unburnt and burnt). The latent space representation again shows compact clusters. The latent space remains two dimensional for the 3D input. In the output data, the third cluster captures two distinct regions: a thin area covering the flame surface, and low temperature areas downstream, where cold fluid from the near-wall region is entrained back into the main flow (cf. temperature distribution in Fig. 2). In both regions, the OH concentration is different from the surrounding fluid: the high OH peak in the flame front and the low OH concentration in the cold burnt gases [37].

Results from the model version with Y_{OH} input are shown in Fig. 9. DNS data is compared to the 2D model version with unstretched flamelets and the 3D model extension with strained flamelets. The model with strained flamelets is better able to capture the wide variations in the values than the unstrained model, both for $\dot{\omega}_C$ and Y_{CO}. For the source term, the effect is more pronounced, and a good agreement of the conditional means is observed. In the CO mass fraction, the wider variation around the flamelet solution has less influence on the conditional mean which remains very similar to the previous results, with only a small improvement around the peak value.

4.3. CO as input parameter

Since the underlying goal of this research is the accurate prediction of CO emissions, it would be convenient to use Y_{CO} as a marker for

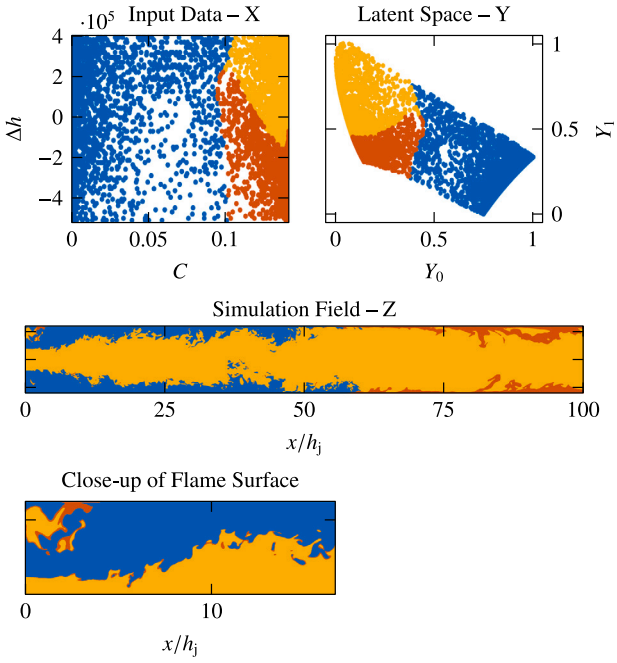


Fig. 8. Clusters when the OH mass fraction Y_{OH} is included as an input parameter. The additional input introduces a third cluster which comprises two distinct regions: the flame front and the downstream low-temperature regions. The bottom image is a close-up view of the upper inlet corner and flame surface.

turbulent strain instead of Y_{OH}. CO is also sensitive to strain, though less so than OH; but if another transported control variable is needed, CO offers additional benefits. In this case, Y_{CO} is transported and the CO source term $\dot{\omega}_{CO}$ is an unclosed term, that must be retrieved from the model. However, CO does not perform well as an input parameter. The joint PDF of the progress variable C and the source term $\dot{\omega}_C$ is presented in Fig. 10. The peak source term values are significantly overpredicted while the conditional mean is underpredicted (cf. Fig. 9).

The prediction of the CO source term by the various model versions is evaluated in Fig. 11. The DNS is compared to the unstretched baseline

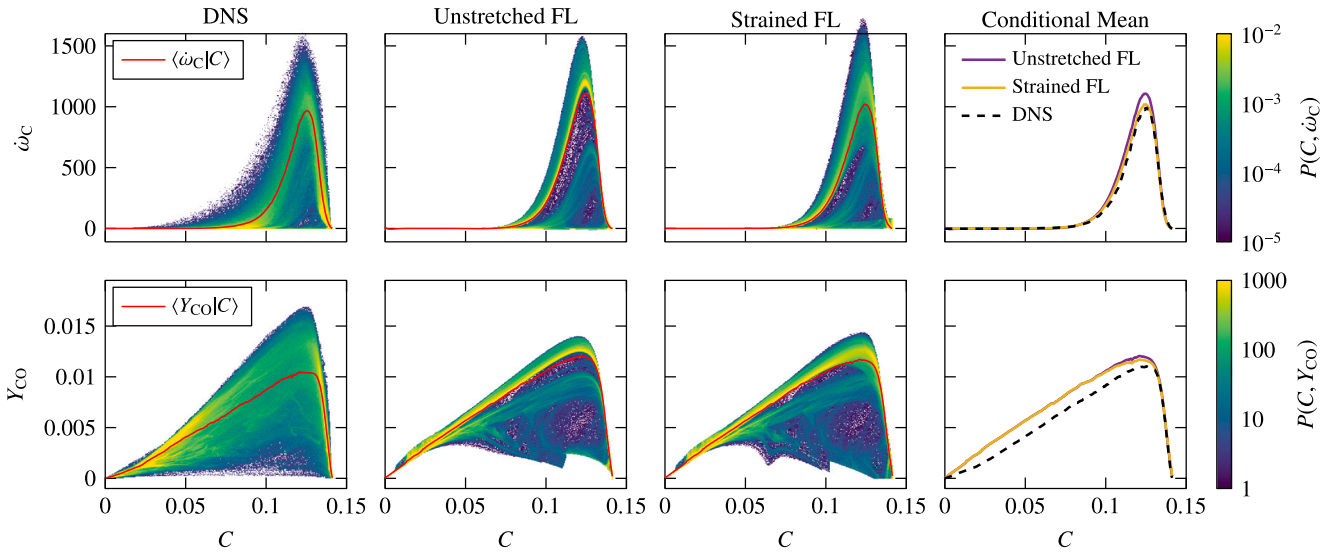


Fig. 9. Joint PDF of the progress variable source term ($\dot{\omega}_C$, top row) and the CO mass fraction (Y_{CO}, bottom row) over the progress variable C. DNS data compared to the base model with two inputs and neglecting turbulent strain effects and to the updated model with three inputs including strained flamelets. Included are the conditional means of all jPDF.

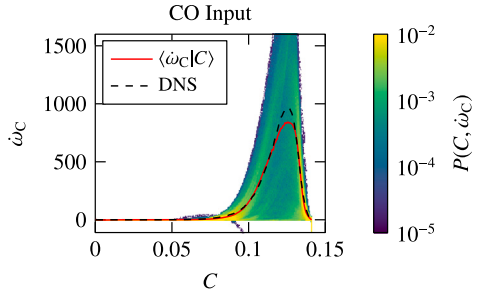


Fig. 10. Joint PDF of $\dot{\omega}_C$ over C when using the CO mass fraction as an input parameter. The conditional means of the model and the DNS are included.

model, the strained model, and the CO input model. Results are similar to the progress variable. While the baseline model and strained model tend to underpredict more extreme values, the results remain within the boundaries of the DNS data. With the addition of Y_{CO} to the input parameters, the CO peak is overpredicted and the region of CO consumption (negative source term) becomes much wider.

The conditional means for all model versions are compiled in Fig. 12; on the left are the conditional means of the progress variable source term $\langle\dot{\omega}_C|C\rangle$ and on the right the CO source term $\langle\dot{\omega}_{CO}|C\rangle$. In both cases, the peak values are underpredicted by the model with CO input, while the unstretched model slightly overpredicts both source terms. The expectation of the strongest correlation between the CO mass fraction and source term does not hold. Instead the strained model with OH predicts both $\dot{\omega}_C$ and $\dot{\omega}_{CO}$ more accurately.

Only two clusters are found when CO is included in the input (Fig. 13), compared to three clusters with OH. This decreases the accuracy of the individual surrogate models, as more different physical effects must be described by a single model. Additionally, the two clusters that are detected are less compact than the clusters of both the baseline and the strained model, indicating that the underlying physical phenomena are less well distinguished. In particular, the "unburnt" (blue) cluster covers regions of cold, but burnt gas. These regions are found downstream of the flame along the cold walls and highlighted in the close-up. While it is possible to force the model to use more clusters, this was not done to avoid *a priori* decisions on the underlying physics; particularly since excessive clustering can cause overfitting [26].

While the CO mass fraction would have been a convenient parameter to track, the physical effects described by it are not sufficiently unique [4] and the algorithm cannot detect and classify them correctly. Accordingly, OH is retained as the parameter describing turbulent strain and CO is not included in the input parameters for further investigations.

4.4. Extrapolation and update procedure

The question of extrapolation must be considered in almost any type of reduced-order model. In the application, thermo-chemical states can occur that were not included in the training data. In conventional tabulation approaches, the detection of extrapolation events is trivial and the extrapolated values can be calculated in a defined manner. Typically, either the closest value or a linear extrapolation are used [3], though more complex methods are possible. In ML-models the extrapolation behavior is more problematic since neither the data boundaries, nor the extrapolation method are well-defined. As many ML-models are non-linear, extrapolation is not necessarily bounded, or even detected [11].

The RONAALP methodology tracks extrapolation events, which enables the user to inspect the consistency of the values. Furthermore, the model can be updated with additional training data to include the thermo-chemical states requiring extrapolation. The first step is extrapolation detection. The data points are evaluated in latent space and their Euclidean distances to the closest RBF center are computed. Data close to a center lies within the original training dataset and can be interpolated, while data far from the closest center is flagged for extrapolation. The threshold depends on the geometry of the latent space and the extrapolation threshold is computed from the average distances of the original data to the RBF centers. The extrapolation detection is adapted from the procedure by Lohninger [38]. Second, the new data must be included in the model. If a new data point has a significant contribution to the model performance a new center is added to the model; if not, the parameters in the vicinity are updated to reflect the new data. As in the original training step, new centers are generated by a k-means algorithm, though in the update step, a sequential k-means algorithm is used. Because the decision regarding new RBF centers is based on the extrapolation threshold, the number of new centers is automatically determined by the algorithm. Finally, the updated RBFs for each cluster are retrained. Further details on the extrapolation and update procedure are provided in [27].

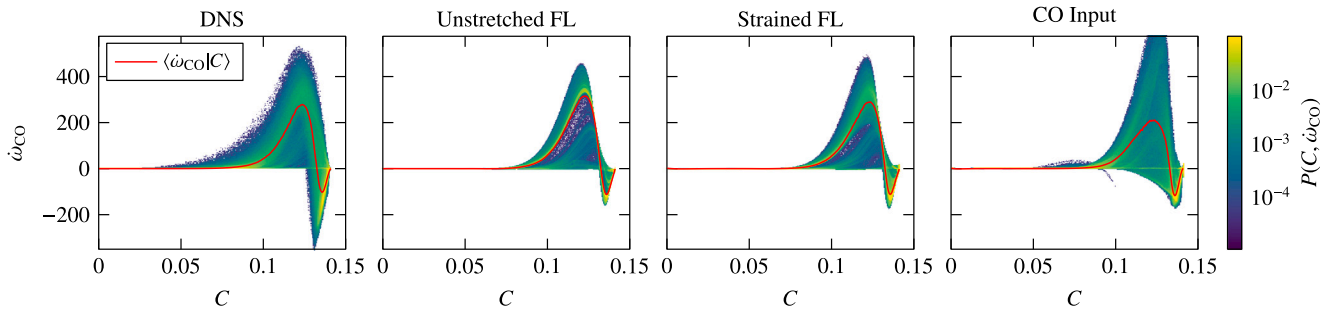


Fig. 11. Joint PDF of the CO source term $\dot{\omega}_{CO}$ over the progress variable C. DNS data compared to the different model formulations: unstretched baseline model, strained model, and the model with CO input.

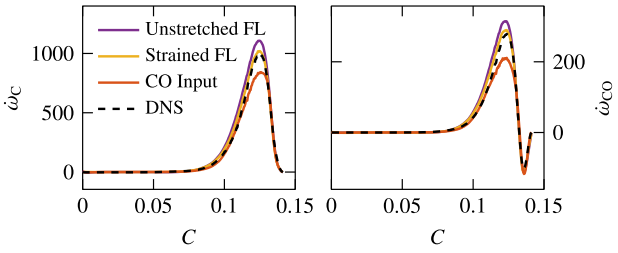


Fig. 12. Conditional means of different RONAALP model formulations, compared to the DNS data. Shown are the progress variable source term $\langle\dot{\omega}_C|C\rangle$ and the CO source term $\langle\dot{\omega}_{CO}|C\rangle$ conditioned on the progress variable.

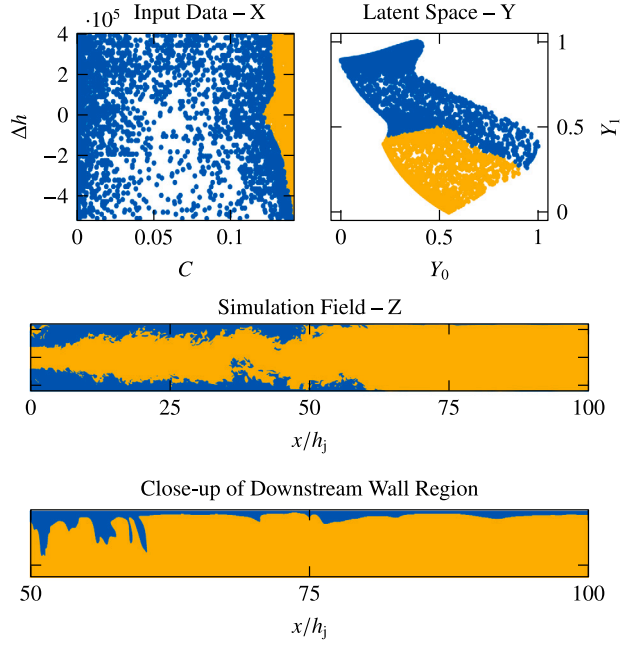


Fig. 13. Clusters when the CO mass fraction Y_{CO} is included as an input parameter. With CO, only two clusters are detected and the clusters are less compact. The bottom image shows a close-up of the near-wall region downstream of the flame. The cold burnt gas is clustered with the unburnt gas.

In this investigation, the main areas of extrapolation are regions of low enthalpy. The training data of adiabatic flamelets at various unburnt temperatures does not reach sufficiently low enthalpy values to capture the cooled burnt gases downstream. An example is presented in Fig. 15, which shows the joint PDF of temperature over enthalpy. Data that was detected as extrapolation is colored in gray.

Several methods exist to compute 1D flames with high heat loss. Burner stabilized flames (e.g. [39,40]) and head-on quenching flames (e.g. [5,41]) are often used. Here, we follow the method of Efimov et al. [5] and compute a 1D flame propagating towards and quenching against an isothermal wall. The boundary conditions of the flame are chosen to match the conditions in the DNS data. The thermo-chemical state of the flame is saved at various times and the model is updated with this additional training data, which contains significantly lower enthalpy levels. A comparison between the DNS data and the model predictions before and after the update is presented in Fig. 14. The figure again shows the joint PDF of temperature and enthalpy. The prediction in the area of lowest enthalpy is clearly improved after the update with data from HOQ flames.

5. Conclusions

A novel machine-learning framework for chemistry modeling has been investigated. The RONAALP model offers a lightweight, computationally inexpensive framework for reduced-order chemistry models. The model is made up of three main components: a dimensionality reduction step, community clustering of similar data, and the training of surrogate models for each cluster. The RONAALP model was applied to turbulent, reacting, wall-bounded flows. Compared to the original application in non-equilibrium, hypersonic flows, this problem included increased chemical complexity while providing a reduced set of training data. In contrast to many ANN applications in the literature, where training is performed with DNS datasets, the training data in this work consisted of one-dimensional flamelets.

The model was evaluated on a DNS database of premixed turbulent methane jet flames with FWI, designed to mimic gas turbine combustion, and yielded promising results. The base version of the ML-model performed slightly better than a conventional chemistry tabulation model. Extending the model with additional strained counterflow flame data further improved the model's performance. The advantage of the ML-model increases as the tabulation input data becomes higher-dimensional and more complex. Due to the adaptive learning procedure, the model is able to detect extrapolation events and can be sequentially retrained with additional data, appending the necessary thermo-chemical states.

The current investigation has demonstrated the feasibility of using this type of light-weight machine learning framework to model the chemistry in complex turbulent flows. Because of the nature of *a priori* investigations, speed-up and further performance metrics could not be calculated for this case. Scherding et al. [27] demonstrated a

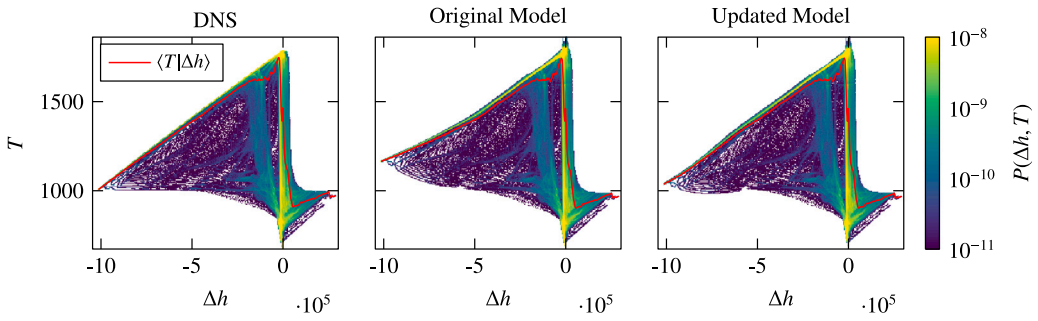


Fig. 14. Joint PDF of the temperature prediction over enthalpy before and after the model update with HOQ flamelets are compared to the DNS data.

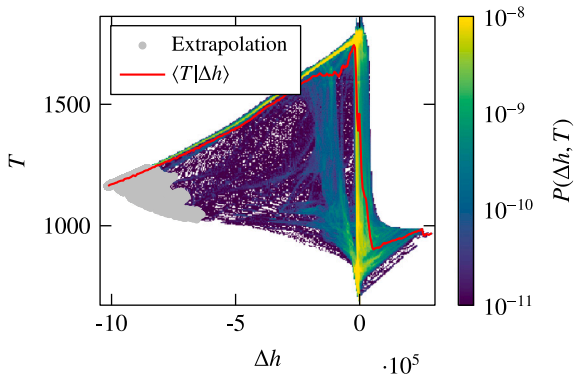


Fig. 15. Joint PDF of temperature over the enthalpy defect showing model predictions with the extrapolated data points highlighted.

speed-up of approximately 4 with RONAALP compared to finite rate chemistry for hypersonic flows in non-equilibrium air with five species. For more complex combustion chemistry mechanisms, higher speed-ups are expected. *A posteriori* model performance and computational metrics such as speed-up and memory requirements will be investigated in future work.

The model was evaluated on data relevant to gas turbine combustion, which is typically statistically stationary. Further topics of interest could be the application to other combustion systems with additional phenomena such as ignition and unsteady combustion effects. One example could be to follow the methodology of Mittal and Pitsch [42], who employed pressure-expansion terms, evaluated from flamelets at various pressures, to consider the chemistry's pressure-dependence in internal combustion engines. The model framework itself is agnostic as long as adequate training data is provided, as was demonstrated by the application to both hypersonic flows [27] and turbulent combustion.

CRediT authorship contribution statement

Kai Niemietz: Writing – original draft, Validation, Methodology, Investigation, Conceptualization. **Clément Scherding:** Validation, Software, Methodology, Formal analysis. **Walter Gemassmer:** Writing – review & editing, Visualization, Software, Investigation. **Taraneh Sayadi:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Heinz Pitsch:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

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

Generous support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant number 434872808 and the Research Association for Combustion Engines (FVV), Germany under grant number 601504 is gratefully acknowledged. Computing time was provided by the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) on the GCS Supercomputer SuperMUC-NG at Leibniz Supercomputing Centre (www.lrz.de) and by the NHR Center NHR4CES at RWTH Aachen University (project number p0020415). This is funded by the Federal Ministry of Education and Research, and the state governments participating on the basis of the resolutions of the GWK for national high performance computing at universities (www.nhr-verein.de/unsere-partner).

Appendix. Temperature field and major species

The focus of the paper has been the accurate prediction of the progress variable source term and the CO mass fraction. Further quantities of interest, i.e., temperature and major species, are presented in this appendix. The following figures (Figs. A.16–A.20) show the temperature field, product species (CO_2 and H_2O), and reactants (CH_4 and O_2) in the full simulation domain. The DNS database is compared to the values evaluated from the RONAALP ML-model. The fully trained model with strained flamelets (cf. Section 4.2) is used here. In addition to the visual comparison, two error metrics are presented: the difference between ML prediction and DNS (ϵ_Δ) and the difference normalized by the mean value of the DNS (ϵ_{norm}), defined as

$$\epsilon_\Delta = \Psi_{\text{ML}} - \Psi_{\text{DNS}} \quad \text{and} \quad \epsilon_{\text{norm}} = \frac{\Psi_{\text{ML}} - \Psi_{\text{DNS}}}{\langle \Psi_{\text{DNS}} \rangle} \tag{A.1}$$

for a scalar field Ψ. The ML predictions are reasonable, with normalized errors below 10 % for almost all fields. Due to the low concentration of methane, the normalized error increases in the unburnt region (Fig. A.19). Throughout all major species, higher errors are found in the unburnt than in the burnt regions. This may be caused by the cluster in the unburnt covering a larger distribution of data than the other clusters (cf. Fig. 8), thus limiting the accuracy of the surrogate model.

Data availability

The source code associated with the implementation of RONAALP described in this paper is available on GitHub: https://github.com/cscherding/RONAALP.

The DNS database is available upon request.

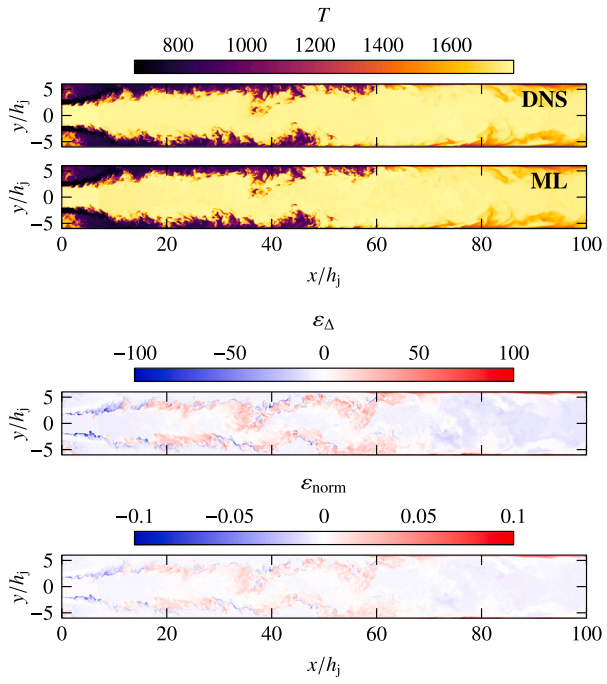


Fig. A.16. Temperature field from the DNS compared to the ML-model prediction. The difference between DNS and ML-model (ε_{Δ}) and the error normalized by the mean value of the DNS ($\varepsilon_{\text{norm}}$) are included.

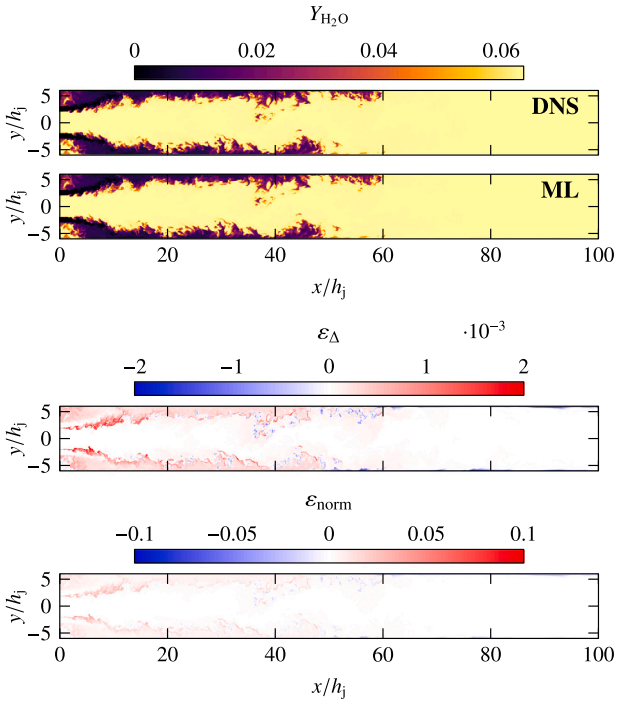


Fig. A.18. Mass fraction of water (H_2O) from the DNS compared to the ML-model prediction. The difference between DNS and ML-model (ε_{Δ}) and the error normalized by the mean value of the DNS ($\varepsilon_{\text{norm}}$) are included.

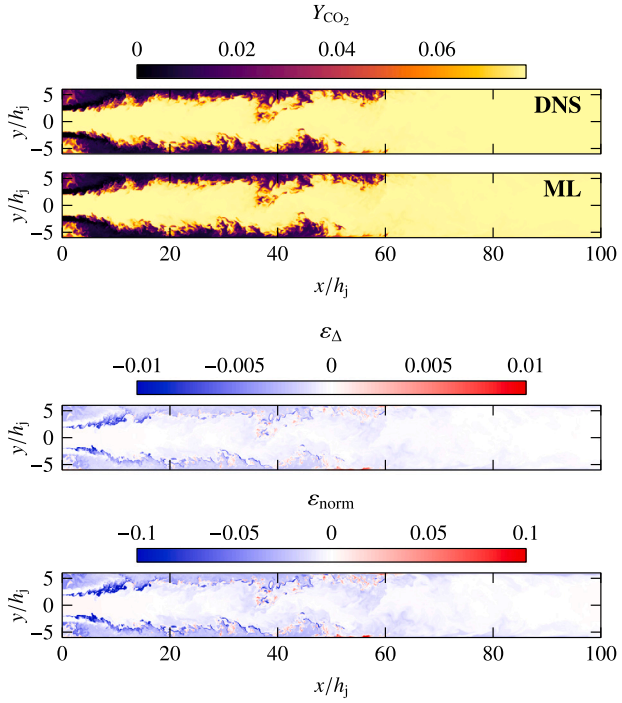


Fig. A.17. Mass fraction of carbon dioxide (CO_2) from the DNS compared to the ML-model prediction. The difference between DNS and ML-model (ε_{Δ}) and the error normalized by the mean value of the DNS ($\varepsilon_{\text{norm}}$) are included.

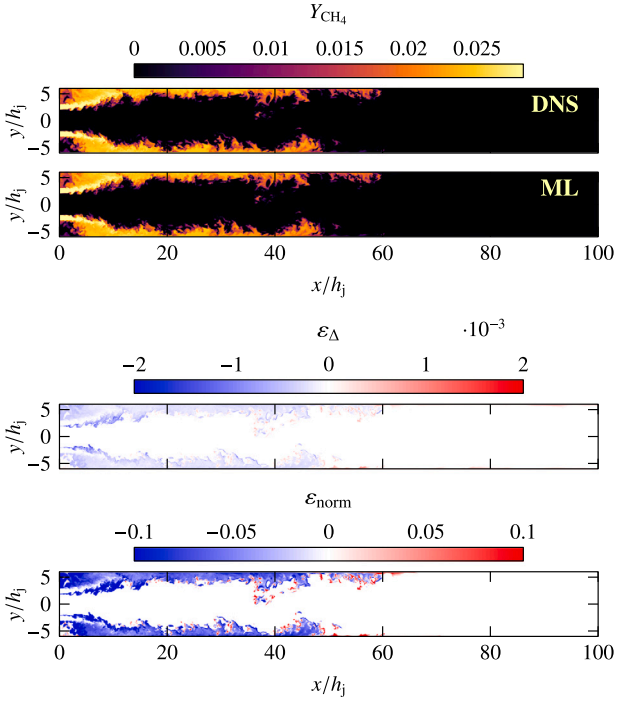


Fig. A.19. Fuel field (methane, CH_4) from the DNS compared to the ML-model prediction. The difference between DNS and ML-model (ε_{Δ}) and the error normalized by the mean value of the DNS ($\varepsilon_{\text{norm}}$) are included.

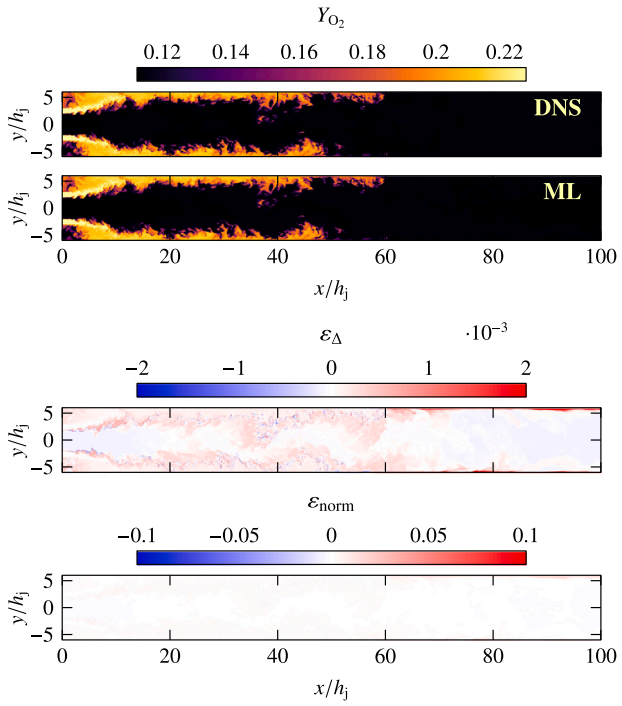


Fig. A.20. Oxygen (O_2) from the DNS compared to the ML-model prediction. The difference between DNS and ML-model (ε_Δ) and the error normalized by the mean value of the DNS (ε_{norm}) are included.

References

[1] Pitsch H. The transition to sustainable combustion: Hydrogen- and carbon-based future fuels and methods for dealing with their challenges. Proc Combust Inst 2024;40:105638.

[2] Pierce CD, Moin P. Progress-variable approach for large-eddy simulation of non-premixed turbulent combustion. J Fluid Mech 2004;504:73–97.

[3] van Oijen JA, de Goey LPH. Modelling of premixed laminar flames using flamelet-generated manifolds. Combust Sci Technol 2000;161:113–37.

[4] Knudsen E, Kolla H, Hawkes ER, Pitsch H. LES of a premixed jet flame DNS using a strained flamelet model. Combust Flame 2013;160:2911–27.

[5] Efimov DV, de Goey P, van Oijen JA. QFM: Quenching flamelet-generated manifold for modelling of flame–wall interactions. Combust Theory Model 2020;24:72–104.

[6] Steinhausen M, Zirwes T, Ferraro F, Scholtissek A, Bockhorn H, Hasse C. Flame-vortex interaction during turbulent side-wall quenching and its implications for flamelet manifolds. Proc Combust Inst 2023;39:2149–58.

[7] Niemietz K, Steinhausen M, Hasse C, Attili A, Pitsch H. A-priori analysis of CO modeling approaches for premixed turbulent jet flames with flame-wall interaction. Combust Flame 2025;278:114242.

[8] Niemietz K, Berger L, Huth M, Attili A, Pitsch H. Direct numerical simulation of flame-wall interaction at gas turbine relevant conditions. Proc Combust Inst 2023;39:2209–18.

[9] Andreini A, Facchini B. Gas turbines design and off-design performance analysis with emissions evaluation. J Eng Gas Turbines Power 2004;126:83–91.

[10] Dreizler A, Böhm B. Advanced laser diagnostics for an improved understanding of premixed flame-wall interactions. Proc Combust Inst 2015;35:37–64.

[11] Ihme M, Chung WT, Mishra AA. Combustion machine learning: Principles, progress and prospects. Prog Energy Combust Sci 2022;91:101010.

[12] Parente A, Sutherland J, Tognotti L, Smith P. Identification of low-dimensional manifolds in turbulent flames. Proc Combust Inst 2009;32:1579–86.

[13] Sutherland JC, Parente A. Combustion modeling using principal component analysis. Proc Combust Inst 2009;32:1563–70.

[14] Bode M, Collier N, Bisetti F, Pitsch H. Adaptive chemistry lookup tables for combustion simulations using optimal B-spline interpolants. Combust Theory Model 2019;23:674–99.

[15] Bissantz J, Karpowski J, Steinhausen M, Luo Y, Ferraro F, Scholtissek A, Hasse C, Vervisch L. Application of dense neural networks for manifold-based modeling of flame-wall interactions. Appl Energy Combust Sci 2023;13:100113.

[16] Kumar T, Almeldein A, Pal P, Kabil I. Data-driven learning of unsteady flamelet progress variable manifolds via hierarchical clustering and grouped multi-target artificial neural networks. In: AIAA SCITECH 2024 forum. 2024.

[17] Chi C, Janiga G, Thévenin D. On-the-fly artificial neural network for chemical kinetics in direct numerical simulations of premixed combustion. Combust Flame 2021;226:467–77.

[18] Zhou L, Song Y, Ji W, Wei H. Machine learning for combustion. Energy AI 2022;7:100128.

[19] Arumapperuma G, Sorace N, Jansen M, Bladek O, Nista L, Sakhare S, Berger L, Pitsch H, Grenga T, Attili A. Extrapolation performance of convolutional neural network-based combustion models for large-eddy simulation: Influence of Reynolds number, filter kernel and filter size. Flow Turbul Combust 2025;1–30.

[20] Nista L, Schumann CD, Petkov P, Pavlov V, Grenga T, MacArt JF, Attili A, Markov S, Pitsch H. Parallel implementation and performance of super-resolution generative adversarial network turbulence models for large-eddy simulation. Comput & Fluids 2025;288:106498.

[21] Ihme M. Construction of optimal artificial neural network architectures for application to chemical systems: Comparison of generalized pattern search method and evolutionary algorithm. In: Hui CLP, editor. Artificial neural networks–application. Rijeka: IntechOpen; 2011, p. 125–50.

[22] Scherding C, Sayadi T, Williams C, Renzo M Di, Rigas G, Schmid PJ. Data-driven reduction of lookup libraries for the simulation of non-equilibrium effects in hypersonic flows. In: Center of Turbulence Research, Proceedings of the summer program. 2022, p. 57–66.

[23] Pope SB. Computationally efficient implementation of combustion chemistry using in situ adaptive tabulation. Combust Theory Model 1997;1:41–63.

[24] Chatzopoulos A, Rigopoulos S. A chemistry tabulation approach via rate-controlled constrained equilibrium (RCCE) and artificial neural networks (ANNs), with application to turbulent non-premixed $CH_4/H_2/N_2$ flames. Proc Combust Inst 2013;34:1465–73.

[25] Franke LL, Chatzopoulos AK, Rigopoulos S. Tabulation of combustion chemistry via artificial neural networks (ANNs): Methodology and application to LES-PDF simulation of sydney flame L. Combust Flame 2017;185:245–60.

[26] Scherding C, Rigas G, Sipp D, Schmid PJ, Sayadi T. Data-driven framework for input/output lookup tables reduction: Application to hypersonic flows in chemical nonequilibrium. Phys Rev Fluids 2023;8:023201.

[27] Scherding C, Rigas G, Sipp D, Schmid PJ, Sayadi T. An adaptive learning strategy for surrogate modeling of high-dimensional functions – application to unsteady hypersonic flows in chemical nonequilibrium. Comput Phys Comm 2025;307:109404.

[28] Niemietz K, Gauding M, Pitsch H. Modeling CO in large eddy simulations of premixed turbulent jet flames with flame-wall interactions. Combust Flame 2026;284:114655.

[29] Cai L, Ramalingam A, Minwegen H, Heufer KA, Pitsch H. Impact of exhaust gas recirculation on ignition delay times of gasoline fuel: An experimental and modeling study. Proc Combust Inst 2019;37:639–47.

[30] Pitsch H. A C++ computer program for 0D combustion and 1D laminar flame calculations – FlameMaster. 1993, URL: https://www.itv.rwth-aachen.de/downloads/flamemaster/.

[31] Newman MEJ. Modularity and community structure in networks. Proc Natl Acad Sci 2006;103:8577–82.

[32] Park J, Sandberg IW. Universal approximation using radial-basis-function networks. Neural Comput 1991;3:246–57.

[33] Platt J. A resource-allocating network for function interpolation. Neural Comput 1991;3:213–25.

[34] Kadirkamanathan V, Niranjan M. A function estimation approach to sequential learning with neural networks. Neural Comput 1993;5:954–75.

[35] Huang G-B, Saratchandran P, Sundararajan N. A generalized growing and pruning RBF (GGAP-RBF) neural network for function approximation. IEEE Trans Neural Netw 2005;16:57–67.

[36] Hawkes ER, Chen JH. Comparison of direct numerical simulation of lean premixed methane–air flames with strained laminar flame calculations. Combust Flame 2006;144:112–25.

[37] Creighton JR. Dependence of CO emissions on the rate of product cooling. Combust Flame 2000;123:402–11.

[38] Lohninger H. Evaluation of neural networks based on radial basis functions and their application to the prediction of boiling points from structural parameters. J Chem Inf Comput Sci 1993;33:736–44.

[39] Fiorina B, Baron R, Gicquel O, Thevenin D, Carpentier S, Darabiha N. Modelling non-adiabatic partially premixed flames using flame-prolongation of ILDM. Combust Theory Model 2003;7:449–70.

[40] Ketelheun A, Kuenne G, Janicka J. Heat transfer modeling in the context of large eddy simulation of premixed combustion with tabulated chemistry. Flow Turbul Combust 2013;91:867–93.

[41] Ganter S, Straßacker C, Kuenne G, Meier T, Heinrich A, Maas U, Janicka J. Laminar near-wall combustion: Analysis of tabulated chemistry simulations by means of detailed kinetics. Int J Heat Fluid Flow 2018;70:259–70.

[42] Mittal V, Pitsch H. A flamelet model for premixed combustion under variable pressure conditions. Proc Combust Inst 2013;34:2995–3003.