energies

MDPI

Article

A Data-Based Hybrid Chemistry Acceleration Framework for the Low-Temperature Oxidation of Complex Fuels

Sultan Alqahtani [1,2], Kevin M. Gitushi [1] and Tarek Echekki [1,*]

[1] Department of Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, NC 27695, USA; ssqahtani@kku.edu.sa (S.A.); kmgitush@ncsu.edu (K.M.G.)
[2] Department of Mechanical Engineering, King Khalid University, Abha 61421, Saudi Arabia
* Correspondence: techekk@ncsu.edu

Abstract: The oxidation of complex hydrocarbons is a computationally expensive process involving detailed mechanisms with hundreds of chemical species and thousands of reactions. For low-temperature oxidation, an accurate account of the fuel-specific species is required to correctly describe the pyrolysis stage of oxidation. In this study, we develop a hybrid chemistry framework to model and accelerate the low-temperature oxidation of complex hydrocarbon fuels. The framework is based on a selection of representative species that capture the different stages of ignition, heat release, and final products. These species are selected using a two-step principal component analysis of the reaction rates of simulation data. Artificial neural networks (ANNs) are used to model the source terms of the representative species during the pyrolysis stage up to the transition time. This ANN-based model is coupled with C_0–C_4 foundational chemistry, which is used to model the remaining species up to the transition time and all species beyond the transition time. Coupled with the USC II mechanism as foundational chemistry, this framework is demonstrated using simple reactor homogeneous chemistry and perfectly stirred reactor (PSR) calculations for n-heptane oxidation over a range of composition and thermodynamic conditions. The hybrid chemistry framework accurately captures correct physical behavior and reproduces the results obtained using detailed chemistry at a fraction of the computational cost.

Keywords: chemistry reduction; data-based hybrid chemistry model; artificial neural network; principal component analysis

Citation: Alqahtani, S.; Gitushi, K.M.; Echekki, T. A Data-Based Hybrid Chemistry Acceleration Framework for the Low-Temperature Oxidation of Complex Fuels. *Energies* **2024**, *17*, 734. https://doi.org/10.3390/en17030734

Academic Editors: Alexandre M. Afonso, Pedro Resende and Mohsen Ayoobi

Received: 3 January 2024
Revised: 26 January 2024
Accepted: 30 January 2024
Published: 4 February 2024

1. Introduction

Establishing a mechanistic separation between faster pyrolysis and subsequent slower oxidation in complex hydrocarbon fuels forms the basis for the HyChem approach to high-temperature oxidation (HTO) [1–3]. The pyrolysis stage features a common set of fuel fragments, whose global chemistry model is coupled with simpler C_0–C_4 foundational chemistry [1]. In contrast to HTO, the low-temperature oxidation (LTO) of these fuels is characterized by the initial formation of fuel-specific intermediates [4] (e.g., for alkanes, these include alkyl, alkylperoxy, and alkyl hydroperoxide radicals). The modeling of the chemistry of these intermediates represents an important challenge in the development of new detailed mechanisms for complex fuels.

In a recent study, we proposed a data-based hybrid chemistry approach for the chemistry acceleration of complex fuels in HTO [5]. In this approach, a set of representative species, instead of pre-selected fuel fragments, are used to model the pyrolysis stage. These species "shadow" the progress of the fuel fragments and can be modeled beyond the pyrolysis stage using simpler foundational C_0–C_2 or C_0–C_4 chemistry [5–7]. The representative species are selected using principal component analysis (PCA), a traditional tool for chemistry reduction [8], and their hybrid chemistry is modeled using artificial neural networks (ANNs) during pyrolysis and simpler foundational chemistry during oxidation.

Recently, Alqahtani [9] proposed an extension of the same approach for the LTO chemistry of complex hydrocarbon fuels. He showed that a set of representative species can be selected for different fuels. One of the key strategies in the development of HyChem [1–3,10–14] and the data-based hybrid chemistry approach [5,7] is based on modeling fuel and fragments during the pyrolysis stage, during which fuel consumption occurs. The fast pyrolysis followed by the slower oxidation of fragments during HTO in complex hydrocarbon fuels was observed by You et al. [15]. Unlike HTO, in LTO, fuel is consumed primarily by oxidation and this consumption occurs in two stages. The first stage, which corresponds to an exponential rise in OH and HO_2 radicals, is characterized by a build-up of fuel-specific intermediates, as mentioned above, a moderate increase in temperature, and oxygenated intermediates like CO, CH_2O, CH_3CHO (acetaldehyde), CH_2CO (ketene), and other heavier carbonyls. The slow oxidation of these intermediates is then followed by a second-stage ignition [16]. Similarly to HTO [5], where PCA is used to identify fragments in the context of representative species, key intermediates that track both stages of fuel consumption can be selected as representative species using PCA. These representative species can then be modeled using ANNs. ANNs have been successfully used in combustion for the tabulation of chemical reactions and chemistry reduction [17–21]. Additionally, as a regression tool for species and temperature chemical source terms, ANNs have been used to accelerate the evaluation of chemistry [22–27].

This study extends the earlier work of Alqahtani in several ways. First, we validate the hybrid chemistry model from data generation to a posteriori simulations by constructing ANN models for the reactions of the representative species. Second, the hybrid model is demonstrated using two canonical reactor models that are commonly used for chemistry reduction. They include ignition with homogeneous chemistry and the use of the perfectly stirred reactor (PSR). Data clustering is also implemented to couple these two reactor models during the training of the ANN. The present study's ultimate goal is to extend the hybrid chemistry strategy for complex fuel HTO [5] to complex fuel LTO. The resulting approach addresses the specific requirements for implementing hybrid chemistry for LTO and implements strategies for combining training data from different reactor models. We explore this feasibility using n-heptane, which was one of the three fuels considered in Ref. [5].

2. Methodology

To demonstrate the proposed hybrid chemistry approach, we contrast the temporal evolutions of fuel-specific intermediates and other species to temperature during the homogeneous ignition of n-heptane during LTO. They are obtained from the simulations of a simple reactor model based on homogeneous chemistry at constant volume, using Senkin [28] and the detailed mechanism proposed by Mehl et al. [29]. Figure 1 shows the temperature and species mole fraction profiles during n-heptane LTO based on the initial conditions of 700 K, 20 atm, and a mixture equivalence ratio of 0.5. The species shown in Figure 1 account for some of the higher species concentrations during all phases of LTO. Figure 1d shows the fuel-derived intermediates, while Figure 1b,c show other intermediate species with lower chemical complexity, including the usual species found in the fragment list for high-temperature chemistry combined with CH_2O, HO_2, and H_2O_2. These species exhibit the mechanistic phases of the ignition process, including the two stages of ignition. More importantly, the simpler species account for the bulk of the composition within the mixture, while the fuel-derived intermediates (e.g., $C_7H_{15}O_2$, C_7H_{14}, $C_7H_{14}O$, and C_7H_1OOH for n-heptane) account for much smaller fractions of the composition of the original fuel at their peaks between the two ignition stages. Toward the start of the second stage, the fuel-derived intermediates are depleted at a faster rate than other species, such as CO, H_2, CH_2O, and C_2H_4. These results support our proposed hybrid chemistry approach. Tracking the initial stages of fuel oxidation using a common set of simpler hydrocarbons and other species, which also track the bulk of the mixture, can potentially delineate the various stages of fuel oxidation and account for the bulk of the atoms in the mixture.

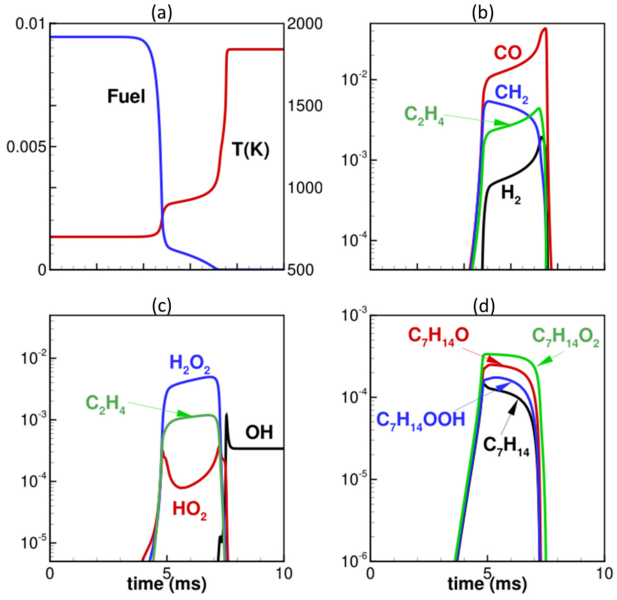

Figure 1. Temperature and species mole fraction profiles during the two-stage LTO of n-heptane, where $\phi = 0.5$, $T = 700$ K, and $p = 20$ atm. (**a**) temperature and fuel mole fraction, (**b**) CO, H_2, C_2H_4 and CH_2 mole fractions, (**c**) OH, HO_2, H_2O_2 and C_2H_4 mole fractions, (**d**) C_7H_{14}, $C_7H_{14}O$, $C_7H_{14}O_2$ and $C_7H_{14}OOH$ mole fractions.

As an extension of our recent work on HTO [5], we seek to couple a foundational chemistry model with a chemical model for complex fuel LTO. Using ANNs, we model the reaction rates of representative species, which are used to track the stages of fuel oxidation. These species tend to consist of simpler hydrocarbons and other intermediates or oxidation species. The remaining species are modeled entirely by foundational chemistry, using the species equations described in Section 2.1. As a data-based framework, the proposed hybrid model is designed to be adaptive to a set of conditions that is determined by the data on which the model is trained.

2.1. Database Generation

We generate a database of chemical data using the simple reactor model of homogeneous chemistry at constant pressure and steady PSR. The mixture conditions used for these reactors consist of equivalence ratios of 0.8, 1, and 1.2 and initial temperatures of 700 to 860 K in increments of 20 K. Additionally, 50 residence times that are evenly spaced on a log scale from -9 to 0 are used for the steady PSR. The lumped mechanism by Brunialti et al. [30] is used as the detailed chemistry mechanism, which consists of 538 species and 2824 reactions. The governing equations for the simple reactor are listed below.

- The energy equation

$$\frac{dT}{dt} = -\frac{\sum_{k=1}^{N} h_l \, \hat{\omega}_k \, \hat{W}_k}{\rho \bar{c}_p} \tag{1}$$

- The species equation for $k = 1, \ldots, N$

$$\frac{dY_k}{dt} = \frac{\hat{\omega}_k \hat{W}_k}{\rho}, \tag{2}$$

where T is the temperature, h_k, $\hat{\omega}_k$, and $\hat{W}_k$ are the enthalpy, reaction rate, and molecular weight of the kth species, respectively, ρ is the mixture density, $\bar{c}_p$ is the mixture average specific heat at constant pressure, and Y_k is the concentration of the kth species.

The governing equations for the steady PSR are listed below.

- The enthalpy equation

$$0 = \frac{h_{in} - h}{\tau} \tag{3}$$

- The species equation for $k = 1, \ldots, N$

$$0 = \frac{Y_{k,in} - Y_k}{\tau} + \frac{\hat{\omega}_k \hat{W}_k}{\rho}, \tag{4}$$

where h is the total enthalpy (chemical and sensible) of the mixture, τ is the residence time, and the subscript *in* denotes the inlet conditions to the PSR.

2.2. Representative Species Selection

Representative species are selected using a principal component analysis (PCA) of the simulation data [5], using the chemical reaction rates of all species in detailed or skeletal mechanisms. PCA is a powerful technique for identifying the importance of the chemical reactions involved in a mechanism [8].

PCA is implemented as an eigendecomposition of the covariance matrices of the species reaction rate data. It results in a vector of principal components (PCs), $\mathbf{\Psi}$, which represents a linear transformation of the centered species reaction rate vector $\mathbf{\Omega}$.

$$\mathbf{\Psi} = \mathbf{Q}^{\mathrm{T}} \ \mathbf{\Omega}. \tag{5}$$

where $\mathbf{Q}$ is the eigenvector of the species reaction rate covariance matrix. The jth PC can be expressed in terms of species reaction rate as follows:

$$\Psi_j = \sum_{k=1}^{N} q_{k,j} \, \hat{\omega}_k^* \tag{6}$$

where N is the total number of species/reaction rates, $\hat{\omega}_k^*$ is the centered (i.e., yielding a zero mean across training data) kth species reaction rate, and $q_{k,j}$ are the (k, j) elements of $\mathbf{Q}$.

The PCs are then ordered based on their eigenvalues, such that the bulk of the data variance is represented by the leading PCs. For a given PC, we also order the species reaction rates in Equation (6), based on the magnitudes of their coefficients $q_{k,j}$. From this re-ordering of the PCs and the species reaction rates for each PC, we identify the corresponding representative species. In Ref. [5], as well as here, we implement PCA in two steps. The two-step procedure is adopted to progressively select the representative species. A set of conditions is imposed on the retained principal components (PCs). The first condition is satisfied by accounting for more than 99% of the data variance. The second condition, i.e., the cutoff condition, is set at 90% of the data variance. The representative species are selected from the list of species that are part of the C_0–C_4 foundational chemistry. Once the representative species are selected, a model for their chemistry is constructed based on the available data.

2.3. Data Clustering

The database of chemical reactions used is a multidimensional and multiscale system with complex nonlinear behavior. Although it is theoretically possible to employ a single ANN to deal with such a complex problem, it is not a good choice because of poor computational efficiency and accuracy. And while a very deep and wide network can be used to approximate the entire dynamics, this may lead to higher computational costs. As a result, it is necessary to cluster data to improve accuracy and lower computational costs [17,18,31,32]. Self-organizing maps (SOMs) [33] and K-means [34] are well-established clustering techniques for clustering chemical data [17–19]. Local principal component analysis (LPA) [35] is another approach used to cluster data with similar chemical compositions.

While clustering techniques are not within the scope of this study, we use the clustering approach implemented in the PCAfold 2.0 software [36]. This software implements a zero-neighborhood clustering technique, which clusters data by separating close-to-zero observations into one or two clusters. This feature of clustering close-to-zero data is very important since the mass fractions and reaction rates of most representative species are zero, if not close to zero, during the early stages of oxidation. To account for the contributions of all representative species, zero-neighborhood clustering is carried out in the principal component (PC) spaces of the reaction rates of both the homogeneous reactor and PSR, where zero PCs correspond to regions where mass fractions and reaction rates are close to zero.

Before applying zero-neighborhood clustering, we separate the data into two categories based on the first and second stages of ignition. The first stage of ignition in LTO can be identified by the "plateau" of the slope of the CO time history [16]. In this study, this plateau occurs around the time when the CH_2O mass fraction reaches its maximum peak. After clustering the data into the two stages of ignition, we apply zero-neighborhood clustering on each stage of ignition.

The parameters of zero-neighborhood clustering are the offset percentage from zero, the number of clusters, and the option to group data close to zero into one or two different clusters. In this work, the number of clusters is set to four for both stages of ignition and data close to zero are clustered into two different bins. This results in a total of eight clusters. The offset percentage from zero at which splits are performed is set to 2.5 and 1 for the first and second stages of ignition, respectively. The clustering is performed as follows [36]:

(a) Using the offset percentage from zero ϵ_0, we obtain the offset ε

$$\varepsilon = \frac{\epsilon_0 \cdot |\max(PC_{obs}) - \min(PC_{obs})|}{100}, \tag{7}$$

where PC_{obs} is the reaction rate PCs;

(b) The cluster borders are obtained using the offset ε as

$$BC^- = \begin{cases} \mathcal{F}(\min(PC_{obs}), & -\varepsilon, & \frac{n_{clusters}+1}{2}), & \text{if } |\min(PC_{obs})| > |\max(PC_{obs})| \\ \mathcal{F}(\min(PC_{obs}), & -\varepsilon, & n_{clusters}/2), & \text{otherwise} \end{cases}, \tag{8}$$

$$BC^+ = \begin{cases} \mathcal{F}(\varepsilon, & \max(PC_{obs}), & n_{clusters}/2), & \text{if } |\min(PC_{obs})| > |\max(PC_{obs})| \\ \mathcal{F}(\varepsilon, & \max(PC_{obs}), & \frac{n_{clusters}+1}{2}), & \text{otherwise} \end{cases}, \tag{9}$$

where BC^- and BC^+ are the negative and positive borders of a cluster, $n_{clusters}$ is the set number of clusters, $||$ is the absolute value operator, and $\mathcal{F}$ is the function to evenly space numbers over a specified interval (*linspace*);

(c) Data are then clustered based on where they fall along the negative or positive borders.

The sensitivities of the offset percentage and the number of clusters are not in the scope of this study; however, these could be explored in future works.

2.4. Artificial Neural Network Architecture

In the present study, an ANN-based regression is developed to determine the reaction rates of the representative species in terms of the temperature and mass fractions of the fuel: O_2, H_2O_2, and CH_2O. These scalars, which are part of the representative species, are adequate markers of each of the stages of ignition and they have high correlations with the remaining representative species. The temperature and these five species are the inputs to the ANNs. ANNs are developed for each cluster.

For each cluster, two sets of ANNs are developed. The first set of ANNs is only used to predict each of the inputs. This is conducted to further improve the accuracy of the hybrid scheme during the a posteriori time integration of the governing equations. Since the ANN models are coupled with foundational chemistry, small errors in the predictions of

temperature and species mass fractions propagate over time and become larger, resulting in the inaccurate prediction of all representative species and the chemical system. By training an ANN for each input, we eliminate the complexity associated with multidimensional outputs and independently optimize each ANN's prediction accuracy and computational efficiency. The second set consists of one ANN for predicting the reaction rates of all remaining species. The data are scaled for training using the equations below.

- Input scaling

$$TS^* = \frac{TS_n^\lambda - 1}{\lambda} \tag{10}$$

where

$$TS_n = \frac{TS - \min(TS)}{\max(TS) - \min(TS)} \quad and \quad \lambda = 0.1 \tag{11}$$

- Reaction rate scaling

$$\hat{\dot{\omega}}_n = 2\frac{\hat{\dot{\omega}} - \min(\hat{\dot{\omega}})}{\max(\hat{\dot{\omega}}) - \min(\hat{\dot{\omega}})} - 1 \tag{12}$$

Each input is modeled by a single ANN comprising two hidden layers with 20 neurons, while the remaining representative species are modeled by a single ANN comprising three hidden layers with 20 neurons. Figure 2 shows an example of the network architectures for CH_2O and remaining species prediction. The transfer function connecting a given neuron to neurons in a previous layer is the hyperbolic tangent function. The ANN weights and biases are updated during the training process by minimizing the selected loss function, which is the mean squared error (MSE). For each input ANN, loss function minimization is optimized by a Levenberg–Marquardt backpropagation algorithm in Matlab [37] and the training is completed over 5000 epochs. The ANN modeling of the remaining species is optimized using the Adam optimizer in PyTorch [38] over 100,000 epochs. During the first 20,000 epochs, a learning rate of 0.005 is used with a weight decay of 10^{-6}, after which the learning rate and weight decay are decreased to 10^{-5} and 10^{-9}, respectively. The training data used to build the hybrid model are divided randomly into three sets: a training set, a test set, and a validation set with the proportions of 70/20/10%, respectively. The training is carried out on AMD Ryzen Threadripper ROI 5975WX CPU for a total time of about 1.5 h, of which about 1 h accounts for training the input ANNs and 30 min account for training the remaining representative species ANNs. A K-nearest neighbor (K-NN) algorithm [39] is used to classify each of the clusters and identify which ANN to use for the prediction of the representative species. This training is carried out in less than 2 min. The classifier uses the same inputs as the ANNs.

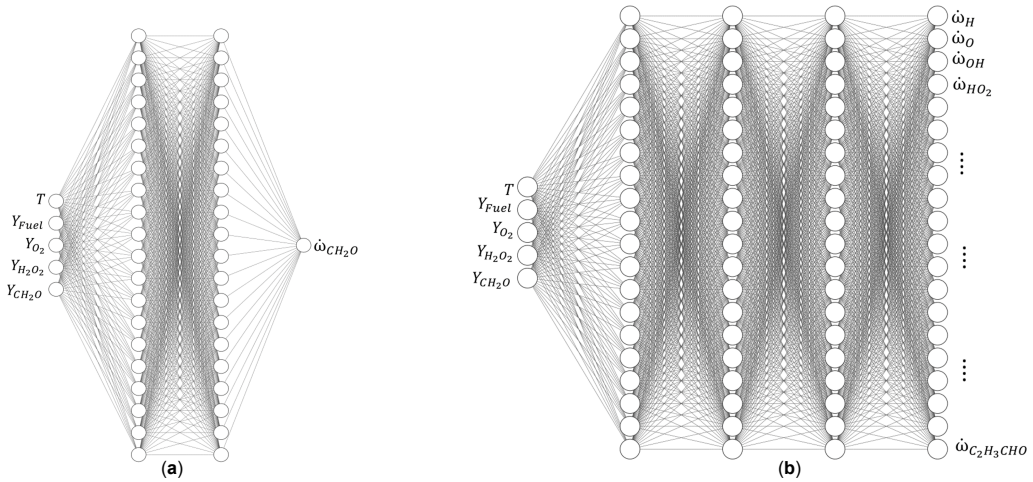

Figure 2. Illustration of the ANN architectures: (**a**) ANN architecture for CH_2O prediction; (**b**) ANN architecture for the remaining reaction rate prediction.

2.5. Summary of the Hybrid Chemistry Model

After generating the data, selecting the representative species, clustering the data, and modeling the reaction rates using ANNs, we couple the ANN-regression models with foundational chemistry, where the ANNs model the representative species and foundational chemistry models all other remaining species. To illustrate the implementation of this hybrid scheme, we use the simple reactor model of homogeneous chemistry at constant pressure, as described in Section 2.1. The governing equations for this system of N species in foundational chemistry and a subset of N_r representative species are listed below.

- The energy equation

$$\frac{dT}{dt} = -\frac{\sum_{k=1}^{N} h_k \hat{\omega}_k \hat{W}_k}{\rho \bar{c}_p} \tag{13}$$

- The representative species equation for $k = 1, \ldots, N_r$

$$\frac{dY_k}{dt} = \frac{\hat{\omega}_k \hat{W}_k}{\rho} \text{ where } \hat{\omega}_k = \begin{cases} \hat{\omega}_{k,\mathrm{h}}, & \text{if } t < \tau_{\text{transition}} \\ \hat{\omega}_{k,\mathrm{f}}, & \text{otherwise} \end{cases}, \tag{14}$$

- Remaining species equation for $k = 1, \ldots, N - N_r$

$$\frac{dY_k}{dt} = \frac{\hat{\omega}_{k,\mathrm{f}} \hat{W}_k}{\rho}. \tag{15}$$

where T is the temperature, h_k, $\hat{\omega}_k$, and $\hat{W}_k$ are the enthalpy, reaction rate, and molecular weight of the kth species, respectively, ρ is the mixture density, $\bar{c}_p$ is the mixture average specific heat at constant pressure, and Y_k is the concentration of the kth species.

In the above equations, the subscript "h" corresponds to a global or ANN model for the representative species that is used up to a transition point and "f" corresponds to the foundational chemistry model. In the present study, that transition point also can be tagged as the transition time $\tau_{\text{transition}}$ during fuel oxidation. An adequate model may be needed up to the onset of the second ignition stage. Beyond that transition time, the representative species reaction rates are modeled with foundational chemistry. Meanwhile, all remaining species are modeled with foundational chemistry. The hybrid model for representative species chemistry attempts to capture the roles of the initial stages of fuel oxidation, including the formation of fuel-specific intermediates, which are not captured by foundational chemistry.

For more complex reactor models, a classifier may be used to decide which chemistry model to use for a given mixture condition. These classifiers, such as the pattern recognition network (PRN) used in Ref. [5], are also trained on detailed chemistry simulation data. Hybrid models for the reaction rates of the representative species can be determined via regression using ANNs, as shown in [5–7]. Alternatively, they can also be determined using global reaction rates, similar to the forms proposed for fragments in HyChem [1–3].

3. Results and Discussion

3.1. Selected Representative Species

Table 1 lists the representative species obtained using the two-step PCA approach. From 538 species in the detailed mechanism, 24 species are retained using the two-step PCA. The two-step PCA achieves a 95.5% reduction. The major products of combustion, CO_2 and H_2O, along with CO, are among the selected species. Low-temperature radicals, such as H_2O_2, HO_2, and CH_2O, are also retained. Other species that are dominant and highly reactive in HTO or the second stage of ignition, such as H_2 and HCCO, are also among the retained species. Additionally, the two-step PCA also selects the species H, O, OH, HO_2, H_2, H_2O, H_2O_2, CH_3, HCO, CH_2O, CO, CO_2, and CH_3CHO, which are formed in the oxidation of any fuel and can be used as universal markers of oxidation reactions [40], as well as OH, HO_2, CH_2O, and CO, which can be used to predict the IDTs of many fuels [41].

This demonstrates that the two-step PCA retains key species that adequately mark both stages of ignition during LTO. A total of 26 thermochemical scalars are modeled, including 24 representative species, temperature, and fuel.

Table 1. List of the selected representative species.

Fuel	**n-C_7H_{16}**
Selected Species (24)	H, O, OH, HO_2, H_2, H_2O, H_2O_2, O_2, CH_3, HCO, CH_2O, CH_3O, CO, CO_2, C_2H_3, C_2H_4, C_2H_5, HCCO, CH_2CO, CH_3CO, CH_2CHO, CH_3CHO, C_3H_6, and C_2H_3 CHO

3.2. Data Clusters

The results of the clustering technique are shown in Figures 3 and 4 for the homogeneous reactor and Figures 5 and 6 for the steady PSR. In Figure 3, the clustered mass fractions (top row) and the reaction rates (bottom row) are shown for the fuel, O_2, H_2O_2, and CH_2O. Figure 4 shows the clusters during both stages of ignition. We can observe that Clusters 2, 3, and 4 contain all zero and close-to-zero reaction rates in the first stage of ignition, while Cluster 1 mostly contains reaction rates with larger magnitudes. This illustrates the importance of the clustering approach. Since most of the close-to-zero reaction rates present during the slow and early stages of oxidation are modeled separately from the large reaction rates, better prediction accuracy is achieved for the early stages of ignition compared to modeling the entire first stage of ignition. Similarly, for the second ignition stage, Cluster 7 contains all zero reaction rates, corresponding to equilibrium. Cluster 8 contains the high-reactivity data of the high-temperature ignition.

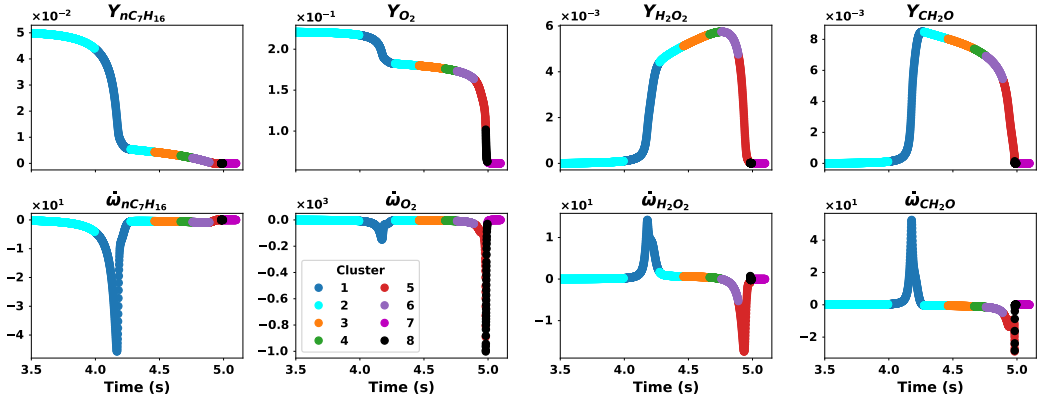

Figure 3. Homogeneous data clustering: species mass fraction and reaction rate temporal profiles during the oxidation of n-heptane, where ϕ = 0.8, T = 700 K, and p = 20 atm.

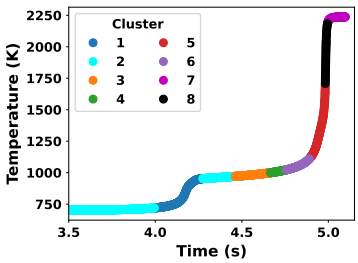

Figure 4. Homogeneous data clustering: temperature temporal profile during the oxidation of n-heptane, where ϕ = 0.8, T = 700 K, and p = 20 atm.

Figure 6 shows the clustered mass fractions (top row) and reaction rates (bottom row) for the fuel, O_2, H_2O_2, and CH_2O and Figure 6 shows the clustered temperature. Unlike the homogeneous reactors where eight clusters are identified, here, data are only found

in Clusters 2 and 7. This is due primarily to the initial separation of the data into two categories based on the stages of ignition, as described in Section 2.3. Additionally, all data after ignition are included in Cluster 7. The reaction rates of the PSRs are 1 to 2 orders of magnitude smaller than those of the homogeneous reactors; therefore, they are close to zero, which is why they fall into Cluster 7.

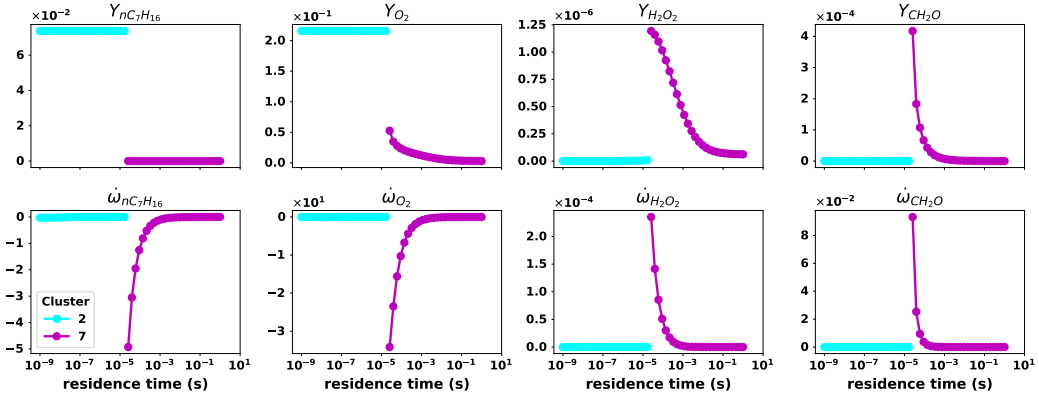

Figure 5. PSR data clustering: species mass fraction and reaction rate temporal profiles, where $\phi = 1.2$, $T = 780$ K, and $p = 20$ atm.

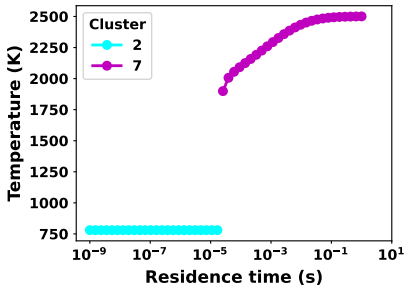

Figure 6. PSR data clustering: temperature temporal profile, where $\phi = 1.2$, $T = 780$ K, and $p = 20$ atm.

3.3. ANN Training and Reaction Rate Prediction

We investigate the generalization of the ANNs by predicting test and training datasets. The scatter plot in Figure 7 shows good correlation between the actual Cantera and ANN-predicted reaction rates. Here, the models are trained using data obtained by solving Equations (1)–(4) with the initial temperatures of 700, 740, 780, 820, and 860 K for equivalence ratios of 0.8, 1, and 1.2 and tested using the initial temperatures of 720, 760, 800, and 840 K for the same equivalence ratios. Figure 8 shows the evolution of the loss functions as a function of the epochs or iterations for the first four clusters. Training for Clusters 5 to 8 also exhibits similar behavior. At around 20,000 iterations, the loss function becomes somewhat "steady" and no longer decreases. Decreasing the learning rate and weight decay improves the network's training and decreases the loss function. Table 2 shows the training losses.

The homogeneous reactor ANN predictions of reaction rates are shown in Figure 9, with an initial temperature of 700 K and an equivalence ratio of 1. The PSR ANN predictions of reaction rates are shown in Figures 10 and 11, with initial temperatures of 740 K and 700 K and equivalence ratios of 1 and 0.8. The ANN predictions accurately match the Cantera solutions. Despite the reactors displaying two different behaviors, the ANNs in each cluster accurately learn and predict each reactor's behavior.

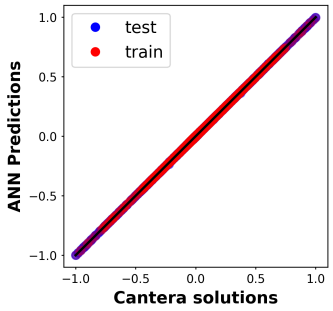

Figure 7. Training (red symbols) and test (blue symbols) data scatter plot of Cantera and ANN-predicted reaction rates. The black line represents the line of best fit.

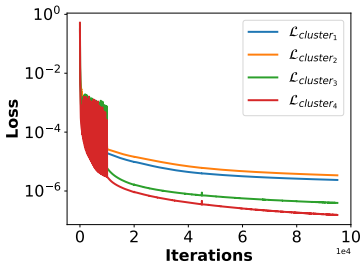

Figure 8. Training losses of clusters for the first stage of ignition.

Table 2. Values of the loss functions of the obtained models for the training dataset.

Cluster	Loss
1	2.4×10^{-6}
2	3.4×10^{-6}
3	3.9×10^{-7}
4	1.5×10^{-7}
5	2.6×10^{-6}
6	2.1×10^{-7}
7	4.6×10^{-6}
8	1.3×10^{-6}

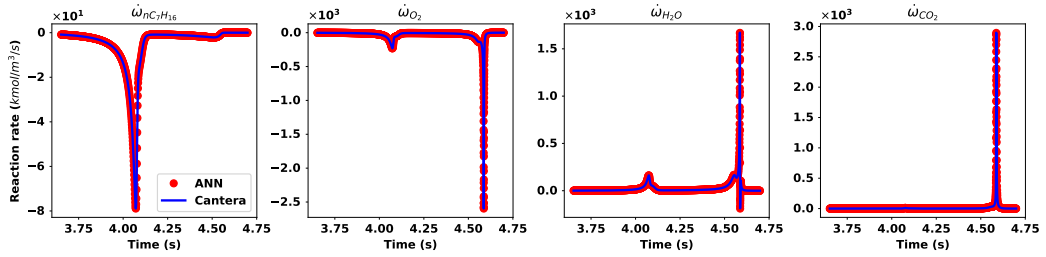

Figure 9. Comparison of the ANN-predicted and Cantera reaction rates for the homogeneous reactor, with an initial temperature of 700 K and an equivalence ratio of 1.

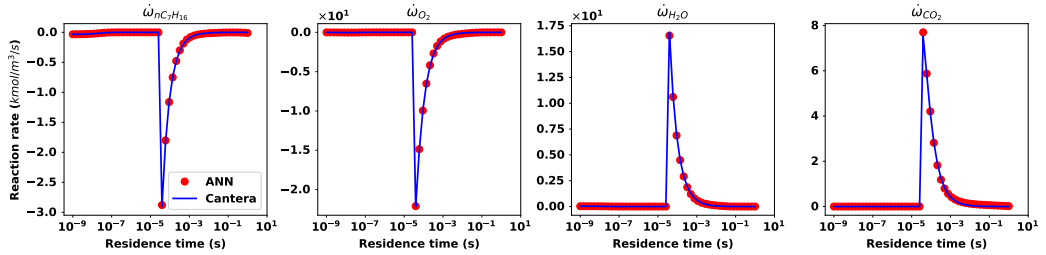

Figure 10. Comparison of the ANN-predicted and Cantera reaction rates for the PSR, with an initial temperature of 740 K and an equivalence ratio of 1.

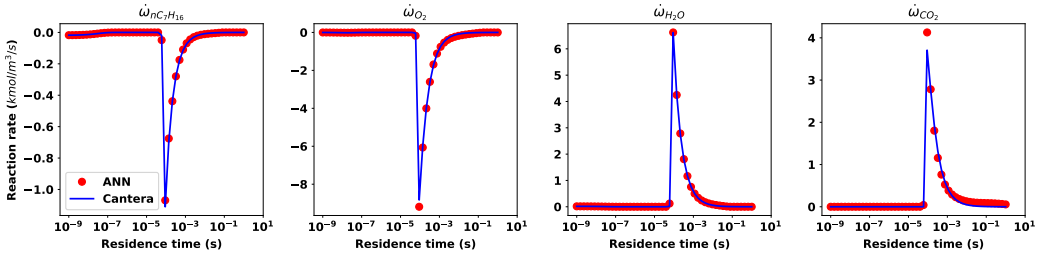

Figure 11. Comparison of the ANN-predicted and Cantera reaction rates for the PSR, with an initial temperature of 700 K and an equivalence ratio of 0.8.

3.4. Hybrid Chemistry Model

We demonstrate the hybrid chemistry approach by comparing a posteriori simulations to detailed chemistry data using the homogeneous reactor. The foundational chemistry is based on the C_0–C_4 USC Mech-II mechanism [42], with 111 species and 784 reactions. The governing Equations (13)–(15) are integrated using Cantera [43]. To demonstrate the hybrid chemistry scheme, we show the results of simulations with initial temperatures of 740 K and 820 K and an equivalence ratio of 1. These two sets of conditions are included in the training data. We also show results with an initial temperature of 720 K and an equivalence ratio of 0.8 to demonstrate the model's ability to interpolate. Figures 12–14 show comparisons of the temporal profiles of the detailed mechanism calculations (solid black) and those of the hybrid scheme (dotted red) for the temperature and select species mass fractions under these initial conditions.

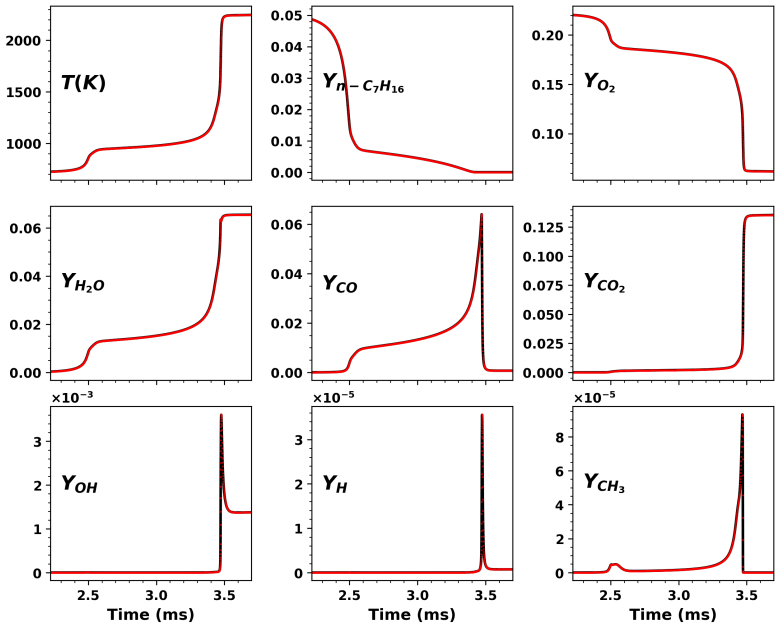

Figure 12. Comparison of temperature and species mass fraction profiles during the oxidation of n-heptane, where ϕ = 0.8, T = 720 K, and p = 20 atm. The hybrid model results are shown by the dotted red lines and those of the detailed mechanism are shown by the solid black lines.

The figures show that the timings of the two stages of ignition are accurately captured by the hybrid chemistry scheme. However, we do not consider fuel-specific intermediates nor the very initial stages of fuel oxidation, which include H abstraction and some of the chain propagation reactions that also involve other key radicals, such as OH and CH_3 [4]. Nonetheless, the profiles of these radicals are well predicted and match the detailed chemistry calculations. Regardless, if the goal is to capture the ignition delay, heat release, and products, the hybrid scheme's performance is very good. More importantly, the scheme results in a computation speed increase of up to 9 times that of detailed chemistry

calculations. Additionally, there is a potential to save memory since the hybrid model tracks fewer species (111) and reactions (784) compared to the detailed mechanism (538 species and 2824 reactions). Further acceleration may be achieved with a further reduction in the number of representative species and foundational chemistry as well.

Overall, the proposed hybrid scheme can capture ignition delay times (IDTs) and the extent of temperature increases (i.e., the rates of heat release) at each stage and can be used to determine the final products of combustion.

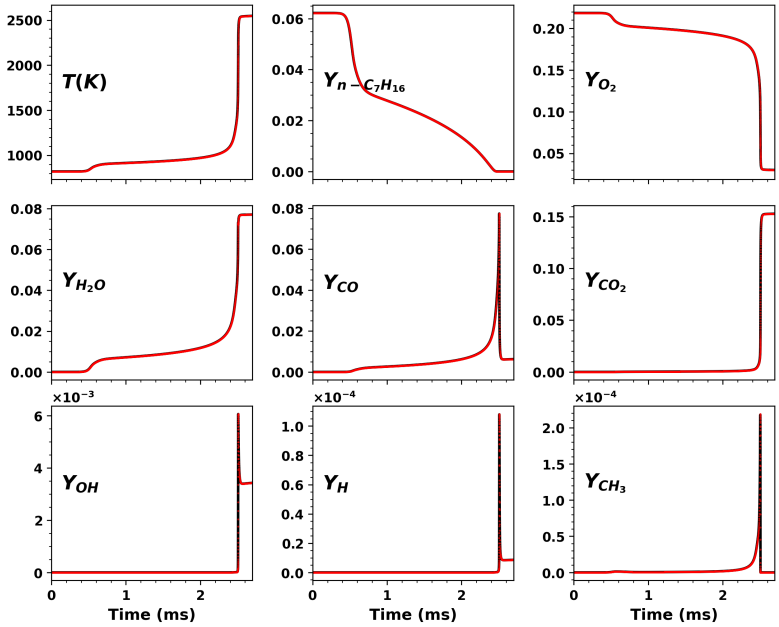

Figure 13. Comparison of temperature and species mass fraction profiles during the oxidation of n-heptane, where ϕ = 1, T = 820 K, and p = 20 atm. The hybrid model results are shown by the dotted red lines and those of the detailed mechanism are shown by the solid black lines.

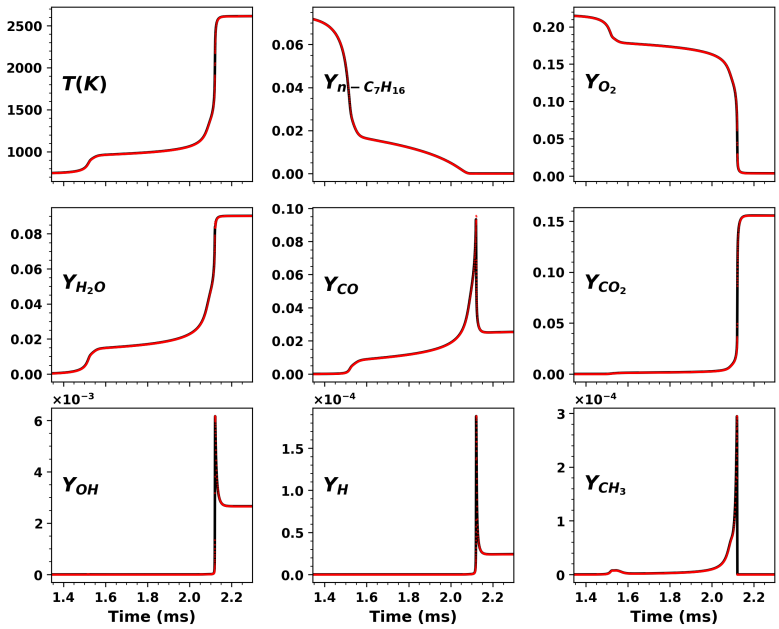

Figure 14. Comparison of temperature and species mass fraction profiles during the oxidation of n-heptane, where ϕ = 1.2, T = 740 K, and p = 20 atm. The hybrid model results are shown by the dotted red lines and those of the detailed mechanism are shown by the solid black lines.

4. Conclusions

In the present study, we demonstrate the feasibility of a data-based hybrid chemistry scheme for LTO, based on modeling the chemistry of representative species and coupling this chemistry with simpler foundational chemistry. In total, 24 representative species are selected using two-step PCA carried on reaction rate data obtained from simulations using detailed chemistry with 538 species and 2824 reactions. Among the 24 representative species are 14 species that are formed during the oxidation of any fuel [40] and 4 species (OH, HO_2, CH_2O, and CO) that can be used for IDT prediction of many fuels [41]. ANNs are used to model the representative species reaction rates and are coupled with smaller C_0–C_4 foundational chemistry with 111 species and 784 reactions to model the remaining species. To improve the computational efficiency and prediction accuracy of the ANNs, the reaction rate data are clustered into eight clusters such that data close to zero are in the same cluster. A validation of this approach via comparisons to detailed chemistry calculations for n-heptane shows that the model results match well with detailed chemistry calculations. The model also correctly captures the timings of the stages of ignition, as well as IDT, heat release, and major species and radicals. With a calculation speed increase of up to 9 times that of detailed chemistry, the hybrid scheme can accelerate the chemistry integration process. Similarly, a speed increase of 1 order of magnitude is achieved for HTO [5].

Future work will aim to achieve aggressive reductions in the numbers of representative species by refining the PCA-based procedure or using other reduction methods, such as directed related graphs (DRGs) [44–46] or global pathway selection (GPS). Ref. [47] identified a smaller number of representative species in a hybrid model based on desired criteria for accuracy in prediction chemistry. The sensitivity of the hybrid scheme to the choice of foundational chemistry will also be investigated to establish prediction accuracy and computational acceleration across different foundational chemistry sizes. We will also extend the scheme using neural ordinary differential equations (NODEs) as these have been shown to have better prediction accuracy and generalization [48–51]. Additionally, improving the criteria for determining the transition from ANN regression to foundational chemistry could further enhance the model and these criteria could be used to implement PRN classifiers. The choice of transition can determine how accurate the final products are because, beyond this point, only foundational chemistry is used; thus, having accurate mixture compositions before the transition is important.

Author Contributions: Conceptualization, S.A., K.M.G. and T.E.; methodology, K.M.G., S.A. and T.E.; software, K.M.G. and S.A.; validation, K.M.G. and S.A.; formal analysis, K.M.G. and S.A.; writing—original draft preparation, K.M.G., S.A. and T.E.; writing—review and editing, K.M.G., S.A. and T.E.; supervision, T.E.; project administration, T.E.; funding acquisition, T.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by King Abdullah University of Science and Technology, under grant number 4351-CRG9.

Data Availability Statement: The data presented in this study are available upon reasonable request from the corresponding author.

Acknowledgments: Sultan Alqahtani would like to acknowledge the support of King Khalid University in Abha, Saudi Arabia.

Conflicts of Interest: The authors declare no conflicts of interest and that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Wang, H.; Xu, R.; Wang, K.; Bowman, C.; Hanson, R.; Davidson, D.; Brezinsky, K. A physics-based approach to modeling real-fuel combustion chemistry-I. evidence from experiments, and thermodynamic, chemical kinetic and statistical considerations. *Combust. Flame* **2018**, *193*, 502–519. [CrossRef]

2. Xu, R.; Wang, K.; Banerjee, S.; Shao, J.; Parise, T.; Zhu, Y.; Wang, S.; Movaghar, A.; Lee, D.; Zhao, R.; et al. A physics-based approach to modeling real-fuel combustion chemistry-II. reaction kinetic models of jet and rocket fuels. *Combust. Flame* **2018**, *193*, 520–537. [CrossRef]
3. Tao, Y.; Xu, R.; Wang, K.; Shao, J.; Johnson, S.; Movaghar, A.; Han, X.; Park, J.-W.; Lu, T.; Brezinsky, K.; et al. A physics-based approach to modeling real-fuel combustion chemistry-III. reaction kinetic model for JP10. *Combust. Flame* **2018**, *198*, 466–476. [CrossRef]
4. Zádor, J.; Taatjes, C.; Fernandes, R. Kinetics of elementary reactions in low-temperature autoignition chemistry. *Prog. Energy Combust. Sci.* **2011**, *37* 371–421. [CrossRef]
5. Alqahtani, S.; Echekki, T. A data-based hybrid model for complex fuel chemistry acceleration at high temperatures. *Combust. Flame* **2021**, *223*, 142–152. [CrossRef]
6. Ranade, R.; Alqahtani, S.; Farooq, A.; Echekki, T. An ANN based hybrid chemistry framework for complex Fuels. *Fuel* **2019**, *241*, 625–636. [CrossRef]
7. Ranade, R.; Echekki, T. An extended hybrid chemistry framework for complex hydrocarbon fuels. *Fuel* **2019**, *251*, 4276–4284. [CrossRef]
8. Brown, N.; Li, G.; Koszykowski, M. Mechanism reduction via principal component analysis. *Int. J. Chem. Kin.* **1996**, *29*, 393–414. [CrossRef]
9. Alqahtani, S. Machine Learning Methods for Chemistry Reduction in Combustion. Ph.D. Thesis, North Carolina State University, Department of Mechanical and Aerospace Engineering, Raleigh, NC, USA, 2020.
10. Xu, R.; Wang, H.; Davidson, D.F.; Hanson, R.K.; Bowman, C.T.; Egolfopoulos, F.N. Evidence supporting a simplified approach to modeling high-temperature combustion chemistry. In Proceedings of the 10th US National Meeting on Combustion, College Park, MD, USA, 23–24 April 2017.
11. Xu, R.; Chen, D.; Wang, K.; Tao, Y.; Shao, J.K.; Parise, T.; Zhu, Y.; Wang, S.; Zhao, R.; Lee, D.J.; Egolfopoulos, F.N. HyChem model: Application to petroleum-derived jet fuels. In Proceedings of the 10th US National Meeting on Combustion, College Park, MD, USA, 23–24 April 2017; Volume 69, pp. 70–77.
12. Xu, R.; Chen, D.; Wang, K.; Wang, H. A comparative study of combustion chemistry of conventional and alternative jet fuels with hybrid chemistry approach. In Proceedings of the 55th AIAA Aerospace Sciences Meeting, Grapevine, TX, USA, 9–13 January 2017; p. 0607.
13. Wang, K.; Xu, R.; Parise, T.; Shao, J.K.; Lee, D.J.; Movaghar, A.; Davidson, D.F.; Hanson, R.K.; Wang, H.; Bowman, C.T.; et al. Combustion kinetics of conventional and alternative jet fuels using a hybrid chemistry (HyChem) approach. In Proceedings of the 10th US National Combustion Institute Meeting, the Combustion Institute, College Park, MD, USA, 23–24 April 2017.
14. Wang, K.; Xu, R.; Parise, T.; Shao, J.K.; Davidson, D.F.; Hanson, R.K.; Wang, H.; Bowman, C.T. Evaluation of a hybrid chemistry approach for combustion blended petroleum and bio-derived jet fuels. In Proceedings of the 10th US National Combustion Institute Meeting, the Combustion Institute, College Park, MD, USA, 23–24 April 2017.
15. You, X.; Egolfopoulos, F.N.; Wang, H. Detailed and simplified kinetic models for n-dodecane oxidation: The role of fuel cracking in aliphatic hydrocarbon combustion. *Proc. Combust. Inst.* **2009**, *32*, 403–410 [CrossRef]
16. Choudhary, R.; Boddapati, V.; Clees, S.; Biswas, P.; Shao, J.; Davidson, D.F.; Hanson, R.K. Towards HyChem modeling of kinetics of distillate fuels in the NTC regime. In Proceedings of the Spring Technical Meeting of the Western States Section of the Combustion Institute, Stanford, CA, USA, 21–22 March 2022.
17. An, J.; He, G.; Luo, K.; Qin, F.; Liu, B. Artificial neural network based chemical mechanisms for computationally efficient modeling of hydrogen/carbon monoxide/kerosene combustion. *Int. J. Hydrogen Energy* **2020**, *45*, 29594–29605. [CrossRef]
18. Haghshenas, M.; Mitra, P.; Santo, N.D.; Schmidt, D.P. Acceleration of chemical kinetics computation with the learned intelligent tabulation (LIT) method. *Energies* **2021**, *14*, 7851. [CrossRef]
19. Nguyen, H.-T.; Domingo, P.; Vervisch, L.; Nguyen, P.-D. Machine learning for integrating combustion chemistry in numerical simulations. *Energy AI* **2021**, *5*, 100082. [CrossRef]
20. Zhou, L.; Song, Y.; Ji, W.; Wei, H. Machine learning for combustion. *Energy AI* **2022**, *7*, 100128. [CrossRef]
21. Chen, J-Y.; Blasco, J. A.; Fueyo, N.; Dopazo, C. An economical strategy for storage of chemical kinetics: Fitting in situ adaptive tabulation with artificial neural networks. *Proc. Combust. Inst.* **2000**, *28*, 115–121. [CrossRef]
22. Blasco, J.A.; Fueyo, N.; Dopazo, C.; Ballester, J. Modelling the temporal evolution of a reduced combustion chemical system with an artificial neural network. *Combust. Flame* **1998**, *113*, 38–52. [CrossRef]
23. Blasco, J.A.; Fueyo, N.; Larroya, J.C.; Dopazo, C.; Chen, Y.J. A single-step time-integrator of a methane–air chemical system using artificial neural networks. *Comput. Chem. Eng.* **1999**, *23*, 1127–1133. [CrossRef]
24. Ihme, M.; Marsden, A.L; Pitsch, H. Generation of optimal artificial neural networks using a pattern search algorithm: Application to approximation of chemical systems. *Neural Comput.* **2008**, *20*, 573–601. [CrossRef]
25. Chatzopoulos, A.K.; 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*, 1465–1473. [CrossRef]
26. Franke, L.L.; Chatzopoulos, A.K.; 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*, 245–260. [CrossRef]

27. Sinaei, P.; Tabejamaat, S. Large eddy simulation of methane diffusion jet flame with representation of chemical kinetics using artificial neural network. *Proc. Inst. Mech. Eng. Part E J. Process Mech. Eng.* **2017**, *231*, 147–163. [CrossRef]
28. Lutz, A.E.; Kee, R.J.; Miller, J.A. *SENKIN: A FORTRAN Program for Predicting Homogeneous Gas Phase Chemical Kinetics with Sensitivity Analysis*; Sandia National Laboratories: Livermore, CA, USA, 1988.
29. Mehl, M.; Pitz, W.J.; Westbrook, C.K.; Curran, H.J. Kinetic modeling of gasoline surrogate components and mixtures under engine conditions. *Proc. Combust. Inst.* **2011**, *33*, 193–200. [CrossRef]
30. Brunialti, S.; Zhang, X.; Faravelli, T.; Frassoldati, A.; Sarathy, S.M. Automatically generated detailed and lumped reaction mechanisms for low- and high-temperature oxidation of alkanes. *Proc. Combust. Inst.* **2023**, *39*, 335-344. [CrossRef]
31. Barwey, S.; Prakash, S.; Hassanaly, M.; Raman, V. Data-driven classification and modeling of combustion regimes in detonation waves. *Flow Turbul. Combust.* **2021**, *106*, 1065–1089. [CrossRef]
32. Ranade, R.; Li, G.; Li, S.; Echekki, T. An efficient machine-learning approach for pdf tabulation in turbulent combustion closure. *Combust. Sci. Tech.* **2021**, *193*, 1258–1277. [CrossRef]
33. Kohonen, T.; Hynninen, J.; Kangas, J.; Laaksonen, J. SOM_PAK: The self-organizing map program package. *Tech. Rep. Citeseer* **1996**, *31*, 1996.
34. MacQueen, J. Some methods for classification and analysis of multivariate observations. In *Fifth Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press: Berkeley, CA, USA, 1967; Volume 1, pp. 281–297.
35. D'Alessio, G.; Parente, A.; Stagni, A.; Cuoci, A. Adaptive chemistry via pre-partitioning of composition space and mechanism reduction. *Combust. Flame* **2020**, *211*, 68–82. [CrossRef]
36. Zdybal, K.; Armstrong, E.; Parente, A.; Sutherland, J.C. *P*CAfold 2.0-Novel tools and algorithms for low-dimensional manifold assessment and optimization. *SoftwareX* **2023**, *23*, 101447. [CrossRef]
37. MathWorks Inc. Matlab Version: 9.13.0 (r2022b). 2022. Available online: https://www.mathworks.com (accessed on 1 December 2022).
38. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Kottayam, India, 2019; Volume 32, pp. 8024–8035.
39. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
40. Buras, Z.; Hansen, N.; Taatjes, C.A.; Sheps, L. Prospects and Limitations of Predicting Fuel Ignition Properties from Low-Temperature Speciation Data. *Energy Fuels* **2022**, *36*, 3229–3238. [CrossRef]
41. Buras, Z.J.; Safta, C.; Zádor, J.; Sheps, L. Simulated production of OH, HO_2, CH2O, and CO_2 during dilute fuel oxidation can predict 1st-stage ignition delays. *Combust. Flame* **2020**, *216*, 472–484. [CrossRef]
42. Wang, H.; You, X.; Joshi, A.; Davis, S.; Laskin, A.; Egolfopoulos, F.; Law, C.K. USC Mech Version II. High-Temperature Combustion Reaction Model of $H_2/CO/C_1$-C_4 Compounds, Tech. Rep. (May, 2007). Available online: https://ignis.usc.edu:80/Mechanisms/USC-Mech%20II/USC_Mech%20II.htm (accessed on 15 May 2021).
43. Goodwin, D. G.; Moffat, H. K.; Schoegl, I.; Speth, R. L.; Webber, B. W. Cantera: An Object-Oriented Software Toolkit for Chemical Kinetics, Thermodynamics, and Transport Processes. Version 2.6.0. 2022. Available online: https://www.cantera.org (accessed on 20 November 2023).
44. Lu, T.F.; Law, C.K. A directed relation graph method for mechanism reduction. *Proc. Combust. Inst.* **2005**, *30*, 1333–1341. [CrossRef]
45. Lu, T.F.; Law, C.K. Linear time reduction of large kinetic mechanisms with directed relation graph: N-heptane and iso-octane. *Combust. Flame* **2006**, *144*, 24–36. [CrossRef]
46. Lu, T.F.; Law, C.K. On the applicability of directed relation graphs to the reduction of reaction mechanisms. *Combust. Flame* **2006**, *146*, 472–483. [CrossRef]
47. Gao, X.; Yang, S.; Sun, W. A global pathway selection algorithm for the reduction of detailed chemical kinetic mechanisms. *Combust. Flame* **2016**, *167*, 238–247. [CrossRef]
48. Dikeman, H.E.; Zhang, H.; Yang, S. Stiffness-reduced neural ODE models for data-driven reduced-order modeling of combustion chemical kinetics. In Proceedings of the AIAA SCITECH 2022 Forum, San Diego, CA, USA, 3–7 January 2022; p. 0226.
49. Owoyele, O.; Pal, P. Chemnode: A neural ordinary differential equations framework for efficient chemical kinetic solvers. *Energy AI* **2022**, *7*, 100118. [CrossRef]
50. Lee, K.; Parish, E.J. Parameterized neural ordinary differential equations: Applications to computational physics problems. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2021**, *477*, 2253. [CrossRef]
51. Ji, W.; Richter, F.; Gollner, M.J.; Deng, S. Autonomous kinetic modeling of biomass pyrolysis using chemical reaction neural networks. *Combust. Flame* **2022**, *240*, 111992. [CrossRef]