Contents lists available at ScienceDirect





Combustion and Flame

journal homepage: www.sciencedirect.com/journal/combustion-and-flame

Deep mechanism reduction (DeePMR) method for fuel chemical kinetics

Zhiwei Wang ^a, Yaoyu Zhang ^{a,b}, Pengxiao Lin ^a, Enhan Zhao ^{e,f}, Weinan E ^{c,d}, Tianhan Zhang ^{e,*}, Zhi-Qin John Xu ^{a,**}

^a Institute of Natural Sciences, School of Mathematical Sciences, MOE-LSC and Qing Yuan Research Institute, Shanghai Jiao Tong

University, Shanghai, 200240, China

^b Shanghai Center for Brain Science and Brain-Inspired Technology, Shanghai, 200240, China

^c AI for Science Institute, Beijing, 100080, China

^d Center for Machine Learning Research, School of Mathematical Sciences, Peking University, Beijing, 100871, China

^e Department of Mechanics and Aerospace Engineering, Southern University of Science and Technology, Guangdong, 518055, China

^f School of Electronics Engineering and Computer Science, Peking University, Beijing, 100871, China

ARTICLE INFO

Keywords: Skeletal mechanism Mechanism reduction Machine learning Deep neural network Chemical kinetics

ABSTRACT

Fuel chemistry represents a typical complex system involving thousands of intermediate species and elementary reactions. Traditional mechanism reduction methods, such as sensitivity analysis and graph-based approaches, fail to explore global correlations of the sub-systems, thereby compromising their efficiency and accuracy. A novel machine learning-based approach called deep mechanism reduction (DeePMR) has been developed to address this issue. The current method transforms mechanism reduction into an optimization problem in the combinatorial space of chemical species while mitigating the curse of dimensionality inherent in the high-dimensional space. We propose an iterative sampling-training-predicting strategy combining deep neural networks with genetic algorithms to learn the landscape of the combinatorial space and locate the targeted subspace. Applying DeePMR to fuel chemistry mechanisms has led to much more compact mechanisms than traditional methods, including directed relation graph (DRG) or path flux analysis (PFA) methods, with three to four orders of magnitude acceleration in numerical simulation. In addition, reduced mechanisms, providing a straightforward yet effective alternative to hierarchy-based construction methods. The DeePMR method provides a general framework for model reduction across various fields.

1. Novelty and significance statement

We propose a deep mechanism reduction (DeePMR) method, transforming the mechanism reduction of a chemical reaction kinetics system into an optimization problem in the combinatorial space of species. DeePMR is a top-down scheme that utilizes a sampling-trainingpredicting strategy by combining deep neural networks with genetic algorithms. This allows it to explore high-dimensional combinatorial space and global species correlations effectively and efficiently. Applying the DeePMR method to fuel chemistry mechanisms yields state-of-the-art reduction, proving a stronger reduction ability than traditional methods.

2. Introduction

Accurate and efficient modeling of chemical reactions within dynamic systems represents a substantial challenge, from combustion in aerospace engines [1], chemical vapor deposition [2], greenhouse gas and pollutants formation in the atmosphere [3] to organic reactions in the interstellar medium [4]. The complexity arises from the vast intermediate products and elementary reactions [5-10] as well as its coupling with the spatiotemporal system [11–13]. A prevailing solution is to develop a full-scope mechanism of all available species and reactions' properties, complete with their thermal, transportive, and kinetic parameters [14-17] via a hierarchy-based approach, i.e., the modeling of simple molecules serves as building blocks for increasingly complex compounds through the sequential addition of new intermediate species and elementary reactions [18]. A typical reaction network development history is illustrated in Fig. 1 based on a gasoline chemistry mechanism [19]. While ab initio calculations and high-quality experiments are useful for certain molecules and reactions [20-22], a significant portion of the parameters is still approximated heuristically. The uncertainty resulting from estimated parameters sometimes, if

E-mail addresses: zhangth@sustech.edu.cn (T. Zhang), xuzhiqin@sjtu.edu.cn (Z.-Q.J. Xu).

https://doi.org/10.1016/j.combustflame.2023.113286

^{*} Corresponding author at: Department of Mechanics and Aerospace Engineering, Southern University of Science and Technology, Guangdong, 518055, China. ** Corresponding author.

Received 25 June 2023; Received in revised form 26 December 2023; Accepted 27 December 2023 0010-2180/ \odot 2023 The Combustion Institute. Published by Elsevier Inc. All rights reserved.



Fig. 1. A history of hierarchy kinetic mechanisms for gasoline surrogate (LLNL) [19]. Circles represent species, and gray dots represent reactions. Species are colored based on the mechanism that includes the species first and are sized based on vortex degree.

not often, goes unnoticed until revealed by later contradictory experiments. Recent studies have suggested that reducing the mechanism's dimensionality can enhance the prediction generalization capability and avoid the occurrence of local minima during optimization [23]. In addition, the cost grows superlinearly with the mechanism size due to the nonlinearity and stiffness. As a result, it is necessary to develop a reduced model with satisfying accuracy and efficiency.

Previous mechanism reduction methods for chemical reaction systems include sensitivity analysis [24,25], chemical lumping method [26,27], computational singularity perturbation [28,29], principal component analysis (PCA) [30,31], time-scale analysis [32,33], manifoldtype method [34], and graph-based methods [35-37]. This field has seen a proliferation of interdisciplinary methods. However, previous methods usually rely on timescale separation assumption (e.g., CSP methods), linear approximation (e.g., principal component analysis), or hypothetical error propagation mode inside reaction network [38]. To what extent those assumptions are necessary to perform model reduction is an open question. For example, the directed relation graph (DRG) method models the conversion flux in the reaction network. The flux conservation helps improve the reduced mechanism's accuracy, but sometimes the reduced model can still have accurate performance without conversion flux conservation. Fig. 2 shows a specific counterexample where the graph-based methods fail to perform the reduction. Based on the error estimation methods mentioned above, there is little difference among the impacts of B_i 's on the flux to the target species C. As a result, B_i will either be removed or retained altogether. In contrast, it is readily seen that removing B_2 and B_3 will not affect the reactant's consumption rate and product's creation rate, which is an acceptable reduction. Therefore, it is challenging for graph-based methods to evaluate to what extent the path flux can accurately indicate the overall performance of the reduced mechanisms and where the assumption might fail. It will be beneficial if we can design a top-down scheme that prioritizes the overall performance of the reduced mechanism without strong assumptions of the complex kinetics.

Machine learning has shown tremendous success in various fields, such as deep learning [39] in predicting protein structure [40], natural language processing [41] etc. A common key feature of these challenging tasks is high dimensionality. Machine learning has also shown potential in combustion research [42,43], such as chemical reaction network construction and analysis [44], generating detailed chemistry mechanism [21], optimizing the parameters in the chemistry mechanism [45–47] and learning surrogate models for stiff chemical processes [48–52].

This study introduces a novel approach called deep mechanism reduction (DeePMR). DeePMR focuses on constructing a mapping function between reduced mechanisms and their overall performance using



Fig. 2. An illustration for a reaction pathway. *A*, B_i , *C* are the reactant, intermediate species, and product, respectively. The inter-conversion rates among B_1 , B_2 , and B_3 are faster than other reaction rates. Consequently, B_i has similar concentrations and contributions to the target species *C*.

deep neural networks, so that DeePMR can gradually eliminate species and associated reactions while ensuring accurate functional preservation. The input to the network is a reduced mechanism representation, modeled as stochastic Boolean networks inspired by systems biology [53]. The output is the performance prediction for a group of benchmark tests. For such high-dimensional mappings, DNN effectively overcomes the 'curse of dimensionality' problem compared with conventional approximation techniques, such as polynomial fitting. In addition, the data sampling, DNN training, and new mechanism labeling proceed iteratively, which enables DNN to explore combinatory space effectively, resulting in a state-of-the-art reduction method.

The DeePMR method has been applied to both alkane-fueled and alcohol-fueled mechanisms, reducing them to mechanisms comprised of approximately 20 to 40 species, including a shared principal species and a limited number of intermediate species,¹ i.e., a principal-satellite formulation (PSF). This approach offers a straightforward yet effective alternative to hierarchy-based construction methods. Even though the current work focuses on fuel chemistry, the DeePMR method constitutes a general framework for model reduction problems across various research fields.

3. Methods

The DeePMR method proposed in this work simplifies a detailed mechanism with N_s species by removing unnecessary species to find

¹ The reduced mechanisms can be found at https://github.com/intelligentalgorithm-team/intelligent-combustion



Fig. 3. Schematic diagram of the DeePMR method. (a) Flowchart of the DeePMR method. A large population of the child I is generated from parents through perturbation and then is screened by fitness function I (neural network). The fitness labels of the retained child II are calculated through the fitness function 2 (Cantera). Child II with high fitness will be used as the parents for the new iteration. In addition, all child II and their fitness labels are also used to reinforce the neural network training dataset to enhance the predictive ability of the neural network. (b) Perturb the complex chemical reaction mechanism. The perturbation means the random removal of some species and the reactions they participate in. The reduced mechanism's ignition, extinction, and propagation characteristics are calculated by numerical simulation (Denoted as Y₁). (c) The neural networks are used to train and predict the ignition, extinction, and propagation characteristics. The entire process has a one-to-one correspondence with (b).

an optimally reduced mechanism, which retains high accuracy on M benchmark quantities, such as ignition time delay (IDT), equilibrium flame temperature, laminar flame speed, and extinction curve in perfectly stirred reactors (PSR) with various initial conditions (temperature, pressure, equivalence ratio). The following are detailed descriptions for each part in the DeePMR method:

State vector To represent a reduced mechanism, an N_s -dimensional Boolean vector $x \in \{0, 1\}^{N_s}$ is employed, with each entry indicating the status of a species: "1" signifies remaining, while "0" denotes removing. Once certain species are removed, their related reactions are reduced.

Perturbation In the perturbation process, we randomly remove *k* species to generate more compact reduced mechanisms, where *k* is a hyper-parameter. In this work, we set $k \in [20, 40]$ for start and $k \in [2, 6]$ for final iteration.

Deep neural network (DNN) The deep neural network directly takes the state vector \mathbf{x} as input, and the output $\mathbf{y} := \{\tilde{Y}_i\}_{i=1}^M$ is the prediction of accuracy on M benchmark quantities. For different types of benchmarks, we adopt separate DNNs to handle. We use a net of ReLU activation function and three hidden layers with 3000, 2000, and 1000 nodes, respectively. This data-driven approach can potentially extract high-order relations among species in order to predict interested quantities accurately. Compared with the graph-based method, which could only utilize low-order correlations, the deep neural network (DNN) could be more efficient for identifying important species groups.

DNN screening We adopt DNN to select promising mechanism candidates. More specifically, for ignition delay time, the sample needs to satisfy $|\log_{10} \tau_{\text{reduced}}^{\text{predict}} / \tau_{\text{detailed}}^{\text{Cantera}}| < \max\{0.3 * 0.98^i, 0.18\}$, with $\tau_{\text{reduced}}^{\text{predict}}$ and $\tau_{\text{detailed}}^{\text{Cantera}}$ representing the IDTs by DNN prediction of the reduced mechanism and Cantera computation of the detailed mechanism, respectively, and *i* being the number of iterations. In each iteration, the screening will continue until we collect enough child samples. In this work, the child sample number is set to be 25,600.

Dataset The newly-reduced mechanisms and their labels constitute the dataset of the current iteration, denoted as $S = \{(x_j, y_j)\}_j$. This dataset serves two purposes: to update the candidate pool for the next



Fig. 4. Iterative sampling process. (a) Top: the error of the best-reduced mechanism (right) and its species number (left) in each iteration; The red solid line is the sum of average relative errors for ignition, extinction, and propagation for the reduced mechanism; The blue-shaded area is the range of species numbers for the reduced mechanism in each iteration. Middle: the retaining ratio of all species in counted iterations; The retaining ratio for the *j*-th species is defined as the number of reduced mechanisms containing the *j*-th species divided by the number of all sampled mechanisms. Bottom: The left and right sides of the error distribution graph are the distribution of $\tau_{reduced}^{Cantera} - \tau_{detailed}^{Cantera}$ and $\tau_{reduced}^{predict} - \tau_{detailed}^{Cantera}$ respectively; The bottom part is the species that were removed from the detailed mechanism. (b) The training process of neural networks. New sparser data are continuously fed to the neural network during the iterative training process, such as the training points marked by asterisks. (c) The comparison between DNN prediction and Cantera results of a reduced mechanism.

iteration and to enlarge the training set for the DNN. The top 5% best mechanisms of the dataset will be selected as the parent samples for the next iteration. The performance indicator used here is the average relative error of benchmark quantities.

DNN training As mentioned above, the neural network utilizes training set *S*. The optimizer is Stochastic Gradient Descent (SGD) with batch size 128. The mean square loss function is defined as

$$\text{Loss} = \frac{1}{N} \sum_{j=1}^{N} \|\boldsymbol{y}_j - \tilde{\boldsymbol{y}}_j\|_2^2,$$

where \tilde{y}_j is the DNN prediction of *j*th reduced mechanism and *N* is the size of dataset.

Simulation As illustrated in Fig. 3b, the benchmark quantities are computed through the open-source software package Cantera, denoted as $y_j := \{Y_{ij}\}_{i=1}^M \in \mathbb{R}^M$.

The method originates from a simple idea, that is brutal force search. Without strategic searching and DNN's efficient screening, the current formulation becomes a typical brutal force search problem. After exhaustively enumerating every possibility with and without certain species and using Cantera to evaluate, it is guaranteed to obtain an optimal reduced mechanism. However, the key challenge is the prohibitive cost to enumerate. For the LLNL n-heptane mechanism (648 species), the total number of all possible reduced mechanisms is $2^{648} \approx 1E195$. It is our current strategy that compresses the number of the mechanism candidates from 1E195 to approximately 1E9. Compared with a classical genetic algorithm method, the current DeePMR method includes boarder distributions of the mechanism candidates. This modification is due to the repeated observations when we obtain a good reduced mechanism by removing several species from poor reduced mechanisms. For further improving efficiency, DNN plays a critical role in fast screening. The DNN screening only takes a few minutes to complete. In the end, the total calls of Cantera are reduced



Fig. 5. Performance of iso-octane skeletal mechanism with 46 species generated by DeePMR. (a) Comparison of the average ignition error relation with the number of remaining species between the DRG method and the DeePMR simplification process. The DRG result comes from [37]. (b) Comparison of iso-octane reduced mechanism generated by DRG [35], DRGEP [37], DRGASA [54,55], DRGEPSA [56], and DeePMR methods. All experimental results are from [56]. (c, d) Comparison of laminar flame speed under 1 atm up to 40 atm between detailed mechanism and skeletal mechanism at 500 K in (c) and 298 K in (d). (e) Computation cost comparison in different reactors. We compare the simulation time between the detailed mechanism and the reduced mechanism in various reactors under different initial temperature, pressure, and equivalence ratio conditions. The colors red, green, and blue represent the equivalence ratios of 0.5, 1, and 1.5, respectively. The histogram shows the average simulation time. All simulations are done using Intel Xeon ICX Platinum 8358. (f) Comparison of extinction curve under 500 K and different pressure and equivalence ratio. (g) Comparison of ignition delay time under different initial conditions.

to 1*E*5. Considering the training and inference cost of the DNN is only several minutes each time, the speedup ratio provided by the DNN is huge. From another perspective, the high-dimensional combinatory space of reduced mechanism, which is also known as the landscape of the mapping function, is extremely complex. DNN is a well-known to effectively approximate such high-dimensional function. Therefore, the actual role of DNN is not simply to accelerate screening. Instead, we use DNN to characterize the landscape of the high-dimensional combinatory space. At the same time, the genetic algorithm helps the DNN collect high-quality candidate data in the neighborhood of optimally reduced mechanisms. The underlying intuition is that if good reduced mechanism can be found nearby. In summary, the optimal solution comes from a better description of the high-dimensional space using DNN instead of combining several good candidates.

4. Results

4.1. Mechanism reduction for various fuels

In this section, we demonstrate the efficacy of the DeePMR method (see the workflow in Fig. 3) in reducing detailed mechanisms while maintaining state-of-the-art accuracy. As an example, we consider the detailed iso-octane mechanism [57], which contains 857 species and 6480 reactions. A wide range of autoignition initial conditions are covered, including temperatures ranging from 600–1700 K, pressures from 1–40 atm, and equivalence ratios from 0.5–1.5.

Firstly, we study the reduction process of the iso-octane mechanism as the sampling iteration proceeds. In Fig. 4a top, with the iteration increasing, the number of kept species constantly decreases significantly while the average relative error of benchmark quantities only slightly increases. In every ten consecutive iterations (one colored block), we compute the retaining ratio distribution and IDT relative error distribution. The retaining ratio for the *j*th species is defined as the number of reduced mechanisms containing the *j*th species divided by the number of all sampled mechanisms in the counted iterations. As shown in Fig. 4a middle, more and more species stably have higher retaining ratio than others, i.e., important species are gradually found, such as fuel, O2, CO2, and H2O. To validate the accuracy of DNN, we examine distributions of the IDT relative error for $\tau_{reduced}^{Cantera} - \tau_{detailed}^{Cantera}$ and $-\tau_{detailed}^{Cantera}$ in each block. As shown in Fig. 4a bottom, these two distributions are symmetric after certain iterations, a consequence of the good prediction accuracy of DNN. The end-to-end approach enables



Fig. 6. Performance of n-heptane skeletal mechanism with 41 species generated by DeePMR. (a) Comparison of Laminar flame speed under 1 atm up to 50 atm and 298 K among detailed mechanism and skeletal mechanism. (b) Comparison of ignition delay time under different initial conditions. (c) Comparison of extinction curve under 298 K and different pressure and equivalence ratio. (d) Comparison of extinction curve under 500 K and different pressure and equivalence ratio. (e) Comparison of laminar flame speed under 1 atm and 298 K among skeletal mechanisms from Jacobian-aided DRGEP method and TSA-reduced skeletal mechanisms from various DRGEP-based methods. (f) Computation cost comparison among different mechanisms.

the DNN to use high-order correlations for good predictions potentially. We further show that the training and the test losses consistently decrease to small values in Fig. 4b. As an example, Fig. 4c shows that the IDT of a reduced mechanism computed by Cantera (red dot) or accurately predicted by DNN (green circle) are both consistent with the one of the detailed mechanisms computed by Cantera. At the bottom of Fig. 4, two snapshots of reduction are shown.

Then, we report that the DeePMR method has stronger simplification ability and stability compared with traditional DRG methods. As shown in Fig. 5a, the mechanism obtained by the DeePMR method has a smaller average error than DRG while retaining the same number of components. As shown in Fig. 5b, our DeePMR method obtains a reduced skeletal mechanism with 46 species and 301 reactions, far less than those of DRG, DRGASA, DRGEP, DRGEPSA [56] while keeping good accuracy. The following comparison experiments are performed to show good accuracy of the reduced mechanism: the laminar flame speed with initial conditions of pressure from 1 atm to 40 atm at 500 K in Fig. 5c and at 298 K in Fig. 5d, the extinction curve at 500 K in Fig. 5f and ignition delay time in Fig. 5g under different initial pressures (1, 5, 40 atm) and equivalence ratios (0.5, 1, 1.5). As the number of species decreases significantly, the detailed mechanism can save the computation cost of about two orders of magnitude compared with the detailed mechanism in various reactors under different initial temperature, pressure, and equivalence ratio conditions (Fig. 5e).

The same strategy is used for an n-heptane mechanism reduction task. Fig. 6 shows the performance of the reduced mechanism with 41 species generated by DeePMR. The laminar flame speed, ignition delay time, and extinction curves comparison are shown for initial pressure P = 1,20 and 50 atm in Fig. 6a–d, respectively. Fig. 6e compares the DeePMR's result of the laminar flame speeds under 1 atm and 298 K with other skeletal mechanisms using Jacobian-aided DRGEP method (Dijkstra_JAC in [61]) and subsequent TSA-reduced skeletal models from various DRGEP-based methods [61] (All reference data for other methods come from [62]). The comparison shows the DeePMR method can reduce the detailed mechanisms into a much more compact size but maintain equally high accuracy. Compared

with the detailed mechanism, the simplified mechanism achieves two orders of magnitude computational acceleration (Fig. 6f). It is worth mentioning that less time cost is not equivalent to weaker stiffness. The reduced mechanisms obtained can exhibit potentially significant numerical stiffness at some low-temperature conditions. For example, a smaller timestep size is required in some autoignition simulations using a reduced iso-octane model compared with the detailed mechanisms. (full details can be found in the supplementary material).

As demonstrated in Table 1, the performance of the DeePMR method is further verified through a number of mechanism reduction tasks. For example, in addition to the iso-octane mechanism mentioned above, we also reduced the 648-species LLNL Gasoline surrogate mechanism [19] and obtained the $C_1 \sim C_7$ alkanes fueled combustion reduced mechanisms; Moreover, the 284-species LLNL butanol isomers mechanism [59] fueled by methanol, ethanol or butanol, 156-species n-propanol mechanism [60], and 341-species JKL n-butane mechanism [58] are reduced to about 20 species by DeePMR, which fully demonstrates that DeePMR is capable of reducing the combustion mechanism for various fuels. The results of reduced mechanisms for $C_1 \sim C_6$ alkanes fueled LLNL mechanism will be further discussed in Section 4.2.

4.2. Reconcile the hierarchical construction of mechanism and the principal-satellite formulation

The development of chemical kinetic mechanisms for hydrocarbon fuels is typically carried out using a hierarchical formulation. In this approach, larger fuel molecules inherit submechanisms from smaller molecules and incorporate additional elementary reactions specific to the targeted fuel. Consequently, the size of the detailed mechanism grows with increasing carbon number and complexity of the fuel. For instance, a mechanism for n-hexadecane ($C_{16}H_{34}$) encompasses a total of 8130 reversible elementary reactions among 2116 chemical species, as reported by [63].

Although the hierarchical approach has been proven useful for modeling various fuels, our research indicates a more effective alternative

Table 1

Reduction results using the DeePMR method for GMIT iso-octane [57], LLNL Gasoline surrogate mechanism [19], JKL $C_0 \sim C_4$ Skeletal [58], LLNL butanol isomers [59], Propanol combustion [60] with various fuels. Ignition delay time and equilibrium flame temperature are common indicators for all mechanisms. The T_{ini} , P_{ini} , and Φ_{ini} are the initial ignition conditions. The abbreviations "D" and "R" after "ns/nr" represent detailed and reduced mechanisms, respectively. The tick symbol " \checkmark " indicates that this indicator is considered in the reduction process, which means the reduced mechanism is applicable to this reactor. The "Error" is the sum of the mean relative errors of all considered indicators.

Mechanism	Fuel	T _{ini} (K)	P _{ini} (atm)	$\Phi_{ m ini}$	PSR	Flame	ns/nr (D)	ns/nr (R)	Err
GMIT iso-octane	iC_8H_{18}	600–1800	1–40	0.5–1.5	1	1	857/6480	46/301	14.3%
LLNL Gasoline surrogate	nC ₇ H ₁₆	650-1800	1–50	0.5–2	1	1		41/292	19.1%
	nC ₆ H ₁₄	1200-1700	1-20	0.5-2				24/162	14.7%
	nC ₅ H ₁₂	1200-1700	1-20	0.5-2				24/165	17.8%
	nC_4H_{10}	1200-1700	1-20	0.5-2			648/4846	21/138	13.3%
	C_3H_8	1200-1700	1-20	0.5-2				20/132	18.7%
	C_2H_6	1200-1700	1-20	0.5-2				19/120	10.2%
	CH_4	1200-1700	1–20	0.5–2				20/146	11.3%
LLNL butanol isomers	CH ₃ OH	1250-1700	1–40	0.5–1.5	1			20/98	3.2%
	C ₂ H ₅ OH	1250-1700	1-40	0.5-1.5	1		284/1900	22/90	5.6%
	nC_4H_9OH	1250-1700	1-40	0.5–1.5	1			24/100	10.7%
Propanol combustion	nC ₃ H ₇ OH	1250-1700	1–40	0.5–1.5	1		156/1414	24/169	7.5%
JKL $C_0 \sim C_4$ Skeletal	nC_4H_{10}	1000-1400	1–40	0.5–2			341/1977	27/149	14.8%

for constructing skeletal fuel mechanisms. As the current method remains elementary reaction rates the same as the detailed one, a detailed mechanism may start from the skeletal reaction pathways instead of adopting a hierarchical formulation. Specifically, we present reduced mechanisms for alkanes and alcohols constrained by ignition delay time (IDT) and equilibrium flame temperature, and reduced mechanisms for alcohols further constrained by the extinction curve in PSRs. Fig. 7a. provides detailed information on these mechanisms. The horizontal row of points shows the inclusion relationship between the mechanism and species. Fig. 7b. demonstrates that these reduced mechanisms accurately predict IDTs in various initial conditions. Notably, as shown in the left panel of Fig. 7a, a group of species shared by all reduced mechanisms is identified as the principal species, while others are contingent on the targeted fuel type so we call them satellite species, that is,

Principal-Satellite Formulation (PSF): a mechanism consists of two parts: one is the principal species shared by all hydrocarbons, and the satellite species represents reaction pathways from the fuel to the principal part.

This principal-satellite formulation is a conceptual analogy with city development, as different fuels share a principal city for modeling high-temperature reactions and possess satellite cities for fuelspecific decomposition and oxidation reactions. In detail, these reduced mechanisms have two characteristics:

- Principal species: alkane-fueled mechanisms all have 13 common species: H₂, O₂, H₂O, CO₂, O, H, OH, H₂O₂, HO₂, CO, HCO, CH₃, and CH₂O. We call them the alkane-principal species. And alcohol-fueled mechanisms have only two more common species: CH₃O and CH₂OH.
- Satellite species: apart from the principal species, the remaining species in the reduced mechanisms can be attributed to the fuel, corresponding free radicals formed through dehydrogenation reactions, and small molecular species generated through various bond dissociation pathways.

Even though the PSF proves valid for the current selections of fuels, it remains an open question whether this structure can be preserved in a more general case. An even more challenging question is how to automatically identify satellite species for different fuels without performing model reduction tediously. In this regard, besides the area's long-time endeavors in chemical kinetic constructions, some pioneering work, including the Hy-Chem model [17], and reaction mechanism generator (RMG) [64], can be helpful for future investigations.

5. Conclusions

This study presents the DeePMR method, which utilizes machine learning techniques to explore high-order correlations among detailed chemical combustion mechanisms within a vast sample space. Through the application of DeePMR, we can extract skeletal mechanisms that accurately reproduce essential benchmark quantities while retaining a minimal number of species at the state-of-the-art level. These skeletal mechanisms serve as effective models for comprehending the dynamics of combustion. Furthermore, the DeePMR method introduces a principal-satellite formulation (PSF) for constructing mechanisms, which offers an alternative to the traditional hierarchy-based approach. Overall, the DeePMR method proposes a machine-learning-based strategy for simplifying complex systems while preserving the accuracy of crucial benchmark quantities.

CRediT authorship contribution statement

Zhiwei Wang: Performed research, Discuss ideas, Analyzed data, Wrote the paper. Yaoyu Zhang: Discuss ideas. Pengxiao Lin: Discuss ideas. Enhan Zhao: Discuss ideas, Analyzed data. Weinan E: Discuss ideas. Tianhan Zhang: Designed research, Performed research, Discuss ideas, Analyzed data, Wrote the paper. Zhi-Qin John Xu: Designed research, Performed research, Discuss ideas, Analyzed data, Wrote the paper.

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

This work is supported by the National Key R&D Program of China Grant No. 2019YFA0709503 (Z.X.), the Natural Science Foundation of Shanghai Grant No. 20ZR1429000 (Z.X.), the National Natural Science Foundation of China Grant No. 12371511 (Z.X.), 12101402 (Y.Z.), Shanghai Municipal of Science and Technology Project Grant No. 20JC1419500 (Y.Z.), the Lingang Laboratory Grant No. LG-QS-202202-08 (Y.Z.), Shanghai Municipal of Science and Technology Major Project No. 2021SHZDZX0102 (Z.X., Y.Z.), and the HPC of School of Mathematical Sciences (Z.X., Y.Z.) and the Student Innovation Center (Z.X., Y.Z.), and the Siyuan-1 cluster supported by the Center for High Performance Computing (Z.X., Y.Z.) at Shanghai Jiao Tong University, and AI for Science Institute, Beijing.



Fig. 7. The retaining species analysis and performance of various mechanisms. (a) Retaining species analysis of the LLNL reduced mechanisms fueled by C_1 to C_6 alkanes and the reduced alcohol mechanisms fueled by C_1 to C_4 alcohols. The common inert gas N_2 is not shown in this figure. Beyond that, all mechanisms share 13 common species, H_2 , O_2 , H_2O , CO_2 , O, H, OH, H_2O_2 , HO_2 , HO_2 , CO, HCO, CO_3 , $OHCO_2$

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.combustflame.2023.113286.

References

- T. Lu, C.K. Law, Toward accommodating realistic fuel chemistry in large-scale computations, Prog. Energy Combust. Sci. 35 (2009) 192–215.
- [2] K. Choy, Chemical vapour deposition of coatings, Prog. Mater. Sci. 48 (2003) 57–170.
- [3] S.M. Saunders, M.E. Jenkin, R. Derwent, M. Pilling, Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of non-aromatic volatile organic compounds, Atmospheric Chem. Phys. 3 (2003) 161–180.
- [4] E. Herbst, W. Klemperer, The formation and depletion of molecules in dense interstellar clouds, Astrophys. J. 185 (1973) 505–534.

- [5] C.A. Taatjes, N. Hansen, A. McIlroy, J.A. Miller, J.P. Senosiain, S.J. Klippenstein, F. Qi, L. Sheng, Y. Zhang, T.A. Cool, et al., Enols are common intermediates in hydrocarbon oxidation, Science 308 (2005) 1887–1889.
- [6] M.P. Burke, S.J. Klippenstein, Ephemeral collision complexes mediate chemically termolecular transformations that affect system chemistry, Nature Chem. 9 (2017) 1078–1082.
- [7] T. Zhang, W. Sun, L. Wang, Y. Ju, Effects of low-temperature chemistry and turbulent transport on knocking formation for stratified dimethyl ether/air mixtures, Combust. Flame 200 (2019) 342–353.
- [8] T. Zhang, Y. Ju, Structures and propagation speeds of autoignition-assisted premixed n-heptane/air cool and warm flames at elevated temperatures and pressures, Combust. Flame 211 (2020) 8–17.
- [9] T. Zhang, A.J. Susa, R.K. Hanson, Y. Ju, Studies of the dynamics of autoignition assisted outwardly propagating spherical cool and double flames under shock-tube conditions, Proc. Combust. Inst. 38 (2) (2021) 2275–2283.
- [10] T. Zhang, A.J. Susa, R.K. Hanson, Y. Ju, Two-dimensional simulation of cool and double flame formation induced by the laser ignition under shock-tube conditions, Proc. Combust. Inst. 39 (2) (2023) 2017–2025.

- [11] R.J. Kee, M.E. Coltrin, P. Glarborg, Chemically Reacting Flow: Theory and Practice, John Wiley & Sons, 2005.
- [12] T. Zhang, W. Sun, Y. Ju, Multi-scale modeling of detonation formation with concentration and temperature gradients in n-heptane/air mixtures, Proc. Combust. Inst. 36 (1) (2017) 1539–1547.
- [13] Y. Wang, X. Guan, S. Xie, M. Zhou, Z. Zhang, Z. Chen, T. Zhang, Numerical studies on the ignition and propagation for spherically expanding premixed cool flames under gravitational conditions, Combust. Flame 259 (2024) 113194.
- [14] E.W. Lemmon, M.L. Huber, M.O. McLinden, et al., NIST standard reference database 23, 2010, REFPROP. version 9.
- [15] G.P. Smith, GRI-mech 3.0, 1999, http://www.me.berkley.edu/gri_mech/.
- [16] D. McElroy, C. Walsh, A. Markwick, M. Cordiner, K. Smith, T. Millar, The UMIST database for astrochemistry 2012, Astron. Astrophys. 550 (2013) A36.
- [17] H. Wang, R. Xu, K. Wang, C.T. Bowman, R.K. Hanson, D.F. Davidson, K. Brezinsky, F.N. Egolfopoulos, A physics-based approach to modeling real-fuel combustion chemistry-I. Evidence from experiments, and thermodynamic, chemical kinetic and statistical considerations, Combust. Flame 193 (2018) 502–519.
- [18] C.K. Westbrook, F.L. Dryer, Chemical kinetic modeling of hydrocarbon combustion, Prog. Energy Combust. Sci. 10 (1984) 1–57.
- [19] M. Mehl, W.J. Pitz, C.K. Westbrook, H.J. Curran, Kinetic modeling of gasoline surrogate components and mixtures under engine conditions, Proc. Combust. Inst. 33 (2011) 193–200.
- [20] D.G. Truhlar, B.C. Garrett, S.J. Klippenstein, Current status of transition-state theory, J. Phys. Chem. 100 (1996) 12771–12800.
- [21] J. Zeng, L. Cao, M. Xu, T. Zhu, J.Z. Zhang, Complex reaction processes in combustion unraveled by neural network-based molecular dynamics simulation, Nature Commun. 11 (2020) 5713.
- [22] O. Welz, J.D. Savee, D.L. Osborn, S.S. Vasu, C.J. Percival, D.E. Shallcross, C.A. Taatjes, Direct kinetic measurements of criegee intermediate (CH200) formed by reaction of CH2i with O2, Science 335 (2012) 204–207.
- [23] T. Isomura, T. Toyoizumi, Dimensionality reduction to maximize prediction generalization capability, Nat. Mach. Intell. 3 (2021) 434-446.
- [24] H. Rabitz, M. Kramer, D. Dacol, Sensitivity Analysis in Chemical Kinetics, Annu. Rev. Phys. Chem. 34 (1983) 419–461.
- [25] T. Turányi, Sensitivity analysis of complex kinetic systems. Tools and applications, J. Math. Chem. 5 (1990) 203–248.
- [26] R.A. Zaveri, L.K. Peters, A new lumped structure photochemical mechanism for large-scale applications, J. Geophys. Res.: Atmos. 104 (1999) 30387–30415.
- [27] P. Pepiot-Desjardins, H. Pitsch, An automatic chemical lumping method for the reduction of large chemical kinetic mechanisms, Combust. Theory Model. 12 (2008) 1089–1108.
- [28] S. Lam, D. Goussis, Understanding complex chemical kinetics with computational singular perturbation, Symp. (Int.) Combust. 22 (1989) 931–941.
- [29] M. Valorani, F. Creta, D.A. Goussis, J.C. Lee, H.N. Najm, An automatic procedure for the simplification of chemical kinetic mechanisms based on CSP, Combust. Flame 146 (2006) 29–51.
- [30] S. Vajda, P. Valko, T. Turanyi, Principal component analysis of kinetic models, Int. J. Chem. Kinet. 17 (1985) 55–81.
- [31] J.C. Sutherland, A. Parente, Combustion modeling using principal component analysis, Proc. Combust. Inst. 32 (2009) 1563–1570.
- [32] N. Peters, Numerical and asymptotic analysis of systematically reduced reaction schemes for hydrocarbon flames, in: Numerical Simulation of Combustion Phenomena: Proceedings of the Symposium, INRIA Sophia-Antipolis, France May 21–24, 1985, 2005, pp. 90–109.
- [33] T. Løvås, D. Nilsson, F. Mauss, Automatic reduction procedure for chemical mechanisms applied to premixed methane/air flames, Proc. Combust. Inst. 28 (2000) 1809–1815.
- [34] U. Maas, S.B. Pope, Simplifying chemical kinetics: intrinsic low-dimensional manifolds in composition space, Combust. Flame 88 (1992) 239–264.
- [35] T. Lu, C.K. Law, A directed relation graph method for mechanism reduction, Proc. Combust. Inst. 30 (2005) 1333–1341.
- [36] W. Sun, Z. Chen, X. Gou, Y. Ju, A path flux analysis method for the reduction of detailed chemical kinetic mechanisms, Combust. Flame 157 (2010) 1298–1307.
- [37] P. Pepiot-Desjardins, H. Pitsch, An efficient error-propagation-based reduction method for large chemical kinetic mechanisms, Combust. Flame 154 (2008) 67–81.
- [38] Y. Wu, Y. Liu, T. Lu, A linearized error propagation method for skeletal mechanism reduction, Combust. Flame 211 (2020) 303–311.
- [39] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436-444.

- [40] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al., Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021) 583–589.
- [41] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, et al., Training language models to follow instructions with human feedback, Adv. Neural Inf. Process. Syst. 35 (2022) 27730–27744.
- [42] L. Zhou, Y. Song, W. Ji, H. Wei, Machine learning for combustion, Energy AI 7 (2022) 100128.
- [43] V. Raman, M. Hassanaly, Emerging trends in numerical simulations of combustion systems, Proc. Combust. Inst. 37 (2) (2019) 2073–2089.
- [44] M. Wen, E.W.C. Spotte-Smith, S.M. Blau, M.J. McDermott, A.S. Krishnapriyan, K.A. Persson, Chemical reaction networks and opportunities for machine learning, Nat. Comput. Sci. (2023) 1–13.
- [45] J. Si, G. Wang, P. Li, J. Mi, A new skeletal mechanism for simulating MILD combustion optimized using Artificial Neural Network, Energy 237 (2021) 121603.
- [46] M. Kelly, S. Dooley, G. Bourque, Toward machine learned highly reduced kinetic models for methane/air combustion, in: ASME Turbo Expo 2021: Turbomachinery Technical Conference and Exposition, 2021, pp. GT2021–58476.
- [47] M. Cailler, N. Darabiha, B. Fiorina, Development of a virtual optimized chemistry method. Application to hydrocarbon/air combustion, Combust. Flame 211 (2020) 281–302.
- [48] F. Christo, A.R. Masri, E.M. Nebot, Artificial neural network implementation of chemistry with pdf simulation of H2/CO2 flames, Combust. Flame 106 (1996) 406–427.
- [49] T. Zhang, Y. Zhang, W. E, Y. Ju, DLODE: a deep learning-based ODE solver for chemistry kinetics, in: AIAA Scitech 2021 Forum, 2021, p. 1139.
- [50] C. Chi, G. Janiga, D. Thévenin, On-the-fly artificial neural network for chemical kinetics in direct numerical simulations of premixed combustion, Combust. Flame 226 (2021) 467–477.
- [51] T. Zhang, Y. Yi, Y. Xu, Z.X. Chen, Y. Zhang, E. Weinan, Z.-Q.J. Xu, A multiscale sampling method for accurate and robust deep neural network to predict combustion chemical kinetics, Combust. Flame 245 (2022) 112319.
- [52] S. Barwey, S. Prakash, M. Hassanaly, V. Raman, Data-driven classification and modeling of combustion regimes in detonation waves, Flow Turbul. Combust. 106 (2021) 1065–1089.
- [53] S.A. Kauffman, Metabolic stability and epigenesis in randomly constructed genetic nets, J. Theoret. Biol. 22 (1969) 437–467.
- [54] X. Zheng, T. Lu, C. Law, Experimental counterflow ignition temperatures and reaction mechanisms of 1, 3-butadiene, Proc. Combust. Inst. 31 (2007) 367–375.
- [55] T. Lu, C.K. Law, Strategies for mechanism reduction for large hydrocarbons: n-heptane, Combust. Flame 154 (2008) 153–163.
- [56] K.E. Niemeyer, C.-J. Sung, M.P. Raju, Skeletal mechanism generation for surrogate fuels using directed relation graph with error propagation and sensitivity analysis, Combust. Flame 157 (2010) 1760–1770.
- [57] H.J. Curran, P. Gaffuri, W.J. Pitz, C.K. Westbrook, A comprehensive modeling study of iso-octane oxidation, Combust. Flame 129 (2002) 253–280.
- [58] Q.-D. Wang, S. Panigrahy, S. Yang, S. Martinez, J. Liang, H.J. Curran, Development of multipurpose skeletal core combustion chemical kinetic mechanisms, Energy Fuels 35 (2021) 6921–6927.
- [59] S.M. Sarathy, S. Vranckx, K. Yasunaga, M. Mehl, P. Oßwald, W.K. Metcalfe, C.K. Westbrook, W.J. Pitz, K. Kohse-Höinghaus, R.X. Fernandes, et al., A comprehensive chemical kinetic combustion model for the four butanol isomers, Combust. Flame 159 (2012) 2028–2055.
- [60] W. Li, Y. Zhang, B. Mei, Y. Li, C. Cao, J. Zou, J. Yang, Z. Cheng, Experimental and kinetic modeling study of n-propanol and i-propanol combustion: Flow reactor pyrolysis and laminar flame propagation, Combust. Flame 207 (2019) 171–185.
- [61] Y. Chen, J.-Y. Chen, Application of Jacobian defined direct interaction coefficient in DRGEP-based chemical mechanism reduction methods using different graph search algorithms, Combust. Flame 174 (2016) 77–84.
- [62] Y. Chen, J.-Y. Chen, Towards improved automatic chemical kinetic model reduction regarding ignition delays and flame speeds, Combust. Flame 190 (2018) 293–301.
- [63] C.K. Westbrook, W.J. Pitz, O. Herbinet, H.J. Curran, E.J. Silke, A comprehensive detailed chemical kinetic reaction mechanism for combustion of n-alkane hydrocarbons from n-octane to n-hexadecane, Combust. Flame 156 (2009) 181–199.
- [64] C.W. Gao, J.W. Allen, W.H. Green, R.H. West, Reaction mechanism generator: Automatic construction of chemical kinetic mechanisms, Comput. Phys. Comm. 203 (2016) 212–225.