



A novel machine learning based lumping approach for the reduction of large kinetic mechanisms for plasma-assisted combustion applications

Georgios Rekkas-Ventiris^{a,b,*}, Alfredo Duarte Gomez^c, Nicholas Deak^{c,e}, Nicholas Kincaid^d, Perrine Pepiot^d, Fabrizio Bisetti^c, Aurélie Bellemans^{a,b}

^a Faculty of Engineering, Thermo and Fluid dynamics (FLOW), Vrije Universiteit Brussel (VUB), 1050 Brussels, Belgium

^b Brussels Institute for Thermal-fluid systems and clean Energy (BRITE), Vrije Universiteit Brussel (VUB) and Université Libre de Bruxelles (ULB), Belgium

^c Department of Aerospace Engineering and Engineering Mechanics, University of Texas at Austin, Austin, TX 78712, USA

^d Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY 14853, USA

^e National Renewable Energy Laboratory (NREL), Golden, CO 80401, USA

ARTICLE INFO

Keywords:

Plasma-assisted combustion

Kinetics reduction

P-DRGEP

Isomer lumping

Machine learning

ABSTRACT

The development of skeletal mechanisms has become essential for multi-dimensional simulations of plasma-assisted combustion (PAC). However, reduction tools developed for traditional combustion applications are not always applicable to PAC, due to the complex interplay between non-equilibrium plasma and combustion kinetics. Plasma direct relation graph with error propagation (P-DRGEP) is a recent plasma-specific reduction method developed in order to incorporate plasma energy branching in the reduction. In the first part of this work, the applicability of P-DRGEP to large kinetic mechanisms is investigated. A detailed isooctane/air plasma mechanism containing 2805 species and 18457 reactions is reduced to 415 species and 4716 reactions, keeping errors on ignition time within 3% for a wide range of initial conditions: from 750 K to 1200 K, 10 atm and equivalence ratios from 0.75 to 1.50. The second part focuses on isomer lumping, which is another reduction technique widely used in combustion. When applied to PAC, it is shown that the resulting lumped mechanism produces poor results. A novel plasma-specific isomer lumping strategy using machine learning is proposed instead. With the supervised algorithm of gradient boosting, predictive regression models are generated, which describe rate coefficients of lumped reactions adequately. These models are trained with simulation data. Leveraging this newly proposed lumping approach on the reduced mechanism, allows for an additional 28% reduction in the number of species and 19% reduction in the number of reactions. Two different versions are presented: in the first one the models are trained using one input feature (1D), while in the second one, two input features are selected (2D). The resulting lumped mechanism is shown to produce accurate predictions of PAC over the entire parameter space of interest, while significantly decreasing the computational time. Indicatively, with the 1D version the maximum error on ignition time in this range of conditions is 6%. The 2D approach produces even lower errors, which do not exceed 3%.

Novelty and significance statement

In this work, a novel approach for isomer lumping, in plasma-assisted combustion mechanisms, is demonstrated. This plasma-specific approach, uses predictive machine learning regression models to describe the complex evolution of lumped reaction rate coefficients. Combining it with the plasma direct relation graph with error propagation, a powerful reduction framework is created, which is successfully demonstrated on a detailed isooctane/air plasma kinetic mechanism, via zero-dimensional ignition simulations. This framework constitutes a useful tool towards the creation of highly accurate skeletal mechanisms, which significantly reduce the computational costs of simulations.

* Corresponding author at: Faculty of Engineering, Thermo and Fluid dynamics (FLOW), Vrije Universiteit Brussel (VUB), 1050 Brussels, Belgium.

E-mail address: georgios.rekkas.ventiris@vub.be (G. Rekkas-Ventiris).

1. Introduction

The enhancement of combustion with various forms of plasma offers significant advantages related to the shortening of ignition time, increase of flame speed and extension of extinction and flammability limits [1–5]. Detailed overviews of applications, which range from internal combustion to hypersonic ramjet engines, can be found in [6, 7].

Non-equilibrium (or low-temperature) plasma in particular, where the temperature of electrons T_e is much higher than the temperature of heavy particles (atoms, molecules, ions), has been a topic of great interest. Its use for robust ignition and stable flames has been proved to be promising [8].

Significant efforts have been made to gain a better understanding on the complex processes by which non-equilibrium plasmas affect combustion, using experiments [9–11] and numerical modeling [12, 13]. Numerical simulations facilitate the exploration of plasma-assisted combustion (PAC) regimes in various initial conditions, that would be difficult and expensive to analyze experimentally. However, simulations using detailed reaction kinetic mechanisms are computationally demanding, due to the increased number of species and reactions brought by the need to describe plasma kinetics in addition to oxidation chemistry [14].

The development of reduced order models has therefore become a necessity. Such models are less stiff and enable two- and three-dimensional computational fluid dynamics (CFD) simulations [15]. Several methods have been applied with success for the reduction of combustion mechanisms without plasma. They are often organized in the following three categories [16]:

- **Dimension reduction.** Computational singular perturbation (CSP) studies the dynamics of a system on the slow manifold and was originally developed for methane and hydrogen combustion [17]. It was later also used for n-heptane/air mixtures [18] and non-premixed flames [19]. Quasi-steady-state approximation (QSSA) is another powerful reduction method, in which the change of concentration in time is set equal to zero, for species whose consumption rate is much faster than their production rate. QSSA is often used in combination with other methods, as an extra reduction step [20,21]. Principal component analysis (PCA) projects the original state-space on a small set of variables. Indicatively, it has been used for synthetic gas combustion [22] and reacting flows modeling [23].
- **Elimination.** Here, unimportant species and reactions are removed from the detailed mechanism. Sensitivity [24] and flux analysis [25] can help identify negligible species and reactions. In graph-based methods, the chemical system can be seen as a graph with nodes and edges. The nodes represent species and the edges quantify the strength of the coupling between the nodes. Direct relation graph (DRG), based on production rate analysis, was firstly applied for the reduction of an ethylene oxidation mechanism [26]. DRG with error propagation (DRGEP) allows for a finer selection of species, by taking into account the damping of the influence of a node to the user-defined target, as it propagates along the graph. DRGEP capabilities were demonstrated by reducing an isooctane oxidation mechanism [27].
- **Lumping.** This category is based on linear or nonlinear transformations on either the mathematical equations or the chemistry governing a mechanism. In isomer lumping, species with similar composition are grouped together and they are represented by a new single pseudo-species. N-heptane and n-dodecane oxidation mechanisms have been reduced with the help of isomer lumping [28–30].

In the recent years, there have been efforts to apply these well-known methods to PAC. For example, PCA was used to reduce a zero-dimensional carbon dioxide plasma model [31], a collisional-radiative argon plasma model [32], and PAC mechanisms [33].

A plasma-specific extension to the DRGEP method, which we called plasma DRGEP (P-DRGEP), was developed by some of the authors [34]. In P-DRGEP, variables that track energy transfer are used in addition to traditional combustion targets. In this manner, plasma kinetics are taken into account during reduction and important plasma-specific pathways are retained in the reduced mechanisms. Its advantage over DRGEP, especially in preserving energy branching characteristics, was demonstrated by reducing an ethylene-air plasma mechanism: starting from 163 species and 1167 reactions, a skeletal mechanism of 54 species and 236 reactions was developed with low errors on the prediction of time to ignition and electron energy [34].

Firstly, the present work investigates the applicability of P-DRGEP to larger, more complex fuels that exhibit complex low temperature chemistry, with isooctane as an initial case study, in the framework of zero-dimensional simulations for PAC using nanosecond pulsed discharges (NSPD).

After the initial reduction with P-DRGEP, isomer lumping is pursued in order to reduced the size of the mechanism further. In this context, the limitations of conventional low-temperature isomer lumping when applied to PAC are characterized and found inadequate. Due to its inability to capture the dynamics of the majority of lumped reactions, a plasma-specific isomer lumping strategy is proposed. The novel approach is based on data-driven science and uses machine learning in order to create predictive regression models, which capture rate coefficients of lumped reactions accurately. This novel approach to lumping is used as part of a hierarchical framework after the initial reduction of the detailed plasma-combustion mechanism by P-DRGEP.

2. Isooctane/air plasma mechanism: physical modeling and detailed simulation

Isooctane (C_8H_{18} , 2,2,4-trimethylpentane) is chosen here as a case study, since it is widely used as a primary reference fuel for gasoline in combustion modeling [35].

In order to assemble the isooctane/air plasma mechanism, the following two mechanisms are combined: a detailed chemical kinetic model for isooctane combustion, which was recently updated and validated against experimental data [36] and a plasma reaction model [37]. The latter has also been used in zero-dimensional ignition studies for methane/air and ethylene/air plasma mixtures [38]. Once combined, the detailed plasma-combustion mechanism features 2805 species and 18457 reactions (S2805R18457), counting forward and reverse reactions separately.

The ignition of the isooctane/air plasma mixture is simulated in a zero-dimensional isochoric and adiabatic reactor, thus neglecting any transport effects. For a more accurate representation of the ignition process, a more complex and higher order system should be investigated (e.g., perfectly stirred reactor) [39]. By representing electrons with the subscript e and all the other particles with p , the evolution of the molar concentration of each species in time is described by the following system of ordinary differential equations (ODEs):

$$\frac{dc_e}{dt} = \omega_e, \quad (1)$$

$$\frac{dc_i}{dt} = \omega_i, \quad i \neq e \quad (2)$$

In the above, ω is the molar production rate. Similarly, the evolution of the two temperatures present in the model can be computed via the ODEs for energy conservation:

$$C_{ve}c_e \frac{dT_e}{dt} = -\omega_e u_e + Q_{el} + Q_{ix} + Q_{re} + Q_E, \quad (3)$$

$$\sum_{i \neq e} C_{vi}c_i \frac{dT}{dt} = -\sum_{i \neq e} \omega_i u_i - Q_{el} - Q_{ix} - Q_{re}. \quad (4)$$

C_v is the specific heat and u is the molar internal energy. Q_{el} describes the elastic energy exchanges, Q_{ix} is the energy lost by the electrons due

Table 1

Discharge characteristics used for the ignition of the isooctane/air plasma mixture.

Number of pulses in each burst	8
Pulse frequency, f [kHz]	100
Single pulse peak power density, P [MW/cm ³]	20
FWHM of each pulse [ns]	30
Energy density per pulse, W [J/cm ³]	0.64

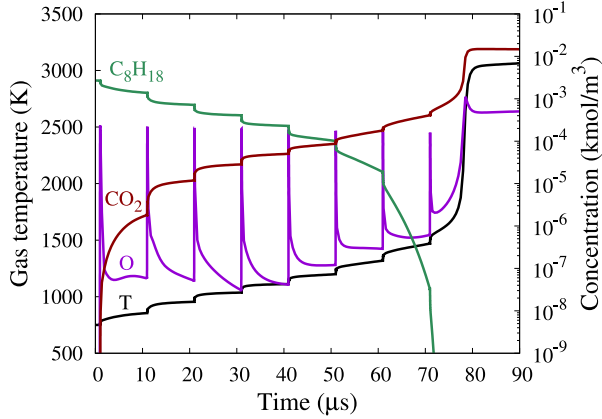


Fig. 1. Temporal evolution of gas temperature and various species for a C_8H_{18} /air plasma mixture. Simulation with the detailed mechanism (S2805R18457) at $T_0 = 750$ K, $p_0 = 10$ atm, $\phi = 1.0$.

to ionization, dissociation and excitation and Q_{re} is the energy lost by the electrons through recombination processes. Consecutive NSPD with pulse frequency f excite the mixture until ignition is reached. Each pulse is Gaussian and therefore the power deposited by the discharge per unit volume Q_E is described by the following equation:

$$Q_E(t) = \frac{E}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(t-\mu)^2}{\sigma^2}\right). \quad (5)$$

μ is the time of peak power, E is the energy density per pulse and σ is the pulse width related to the full-width-half-max FWHM = $2\sqrt{2\ln 2}\sigma$. The discharge characteristics of the pulses used in the simulations are summarized in Table 1.

An in-house computer program called PACMAN is used for the simulations [33,34,38]. PACMAN utilizes CVODE in order to solve the system of ODEs, as formulated in Eqs. (1)–(4) [40]. The thermochemical characteristics of the system are assessed and the necessary computations (e.g., reaction rate coefficients) are performed using the CHEMKIN library [41].

During a pulse, electrons reach a peak mean energy. This energy is then gradually lost through collisions with heavy particles. Consecutively, excited species are formed, mainly driven by electron impact reactions. The new pathways contribute to the production of primary combustion radicals O, OH and H, which are important for the fuel breakdown. Pulse after pulse, the gas temperature T increases, together with the concentration of CO and CO_2 . A more detailed description of the ignition is given in Refs. [38,42].

In order to provide an overview of a plasma-assisted ignition of the isooctane/air mixture, the reactor is initialized at temperature $T_0 = 750$ K, pressure $p_0 = 10$ atm, and equivalence ratio $\phi = 1.0$. The temporal evolution of key species, is shown in Fig. 1. At every pulse there is a rapid increase $\mathcal{O}(1-10)$ ns in the concentration of O, followed by a slower decay $\mathcal{O}(100-1000)$ ns. The same behavior is observed for the two other main combustion radicals, OH and H (not shown in Fig. 1). In the meantime, as C_8H_{18} is consumed, the concentration of CO_2 increases and so does the gas temperature T . The instant when the rate of change in the concentration of CO_2 is maximum designates the time of ignition t^* . Time to ignition τ_{ig} is then defined as the time interval between the first pulse t_1 and ignition t^* : $\tau_{ig} = t^* - t_1$. For the specific detailed

simulation under examination, it is evident that ignition is reached after 8 pulses at $t^* \approx 80 \mu s$.

3. Isomer lumping using machine learning: motivation and methodology

During oxidation of large hydrocarbon fuel molecules, several reactions lead to the formation of functional isomers: species with the same composition (same number of atoms of each element), but different arrangement of atoms. In lumping, they are grouped together into a lumped group and they are thereafter represented by a single pseudo-species.

The relative distribution α of each isomer m with respect to its lumped group l is defined as:

$$\alpha_{m,l} = \frac{c_m}{c_l}, \quad (6)$$

where c denotes molar concentration. For a lumped group consisting of a total number of isomers N_m , the concentration of the new lumped species is equal to the sum of concentrations of isomers in the group:

$$c_l = \sum_{m=1}^{N_m} c_m. \quad (7)$$

After substituting all isomers to be lumped in the mechanism with their corresponding pseudo-species, groups of reactions with identical equations will be formed. These reactions are lumped together as well. For a group of N_r elementary reactions, the forward reaction rate of the newly created lumped reaction r_L has to be equal to the sum of the forward reaction rates of the individual reactions to be lumped r_j :

$$r_L = \sum_{j=1}^{N_r} r_j. \quad (8)$$

For elementary reactions, the concentrations of reactants that are not isomers cancel out, so Eq. (8) becomes:

$$k_L c_l^{v_l} = \sum_{j=1}^{N_r} (k_j \prod_{m=1}^{N_m} c_m^{v_{m,j}}), \quad (9)$$

where k_j is the reaction rate coefficient of reaction j , $v_{m,j}$ is the stoichiometric coefficient of isomer m in reaction j and v_l is the stoichiometric coefficient of the lumped isomer in the lumped reaction ($v_{m,j} = v_l$). Solving Eq. (9) for the reaction rate coefficient of the lumped reaction k_L , using Eq. (7) :

$$k_L = \sum_{j=1}^{N_r} (k_j \prod_{m=1}^{N_m} \alpha_{m,l}^{v_{m,j}}), \quad (10)$$

In conventional combustion, various isomer lumping approaches differ on the manner in which $\alpha_{m,l}$ and therefore k_L are approximated [28,30,43,44]. Some of these approaches were compared in Ref. [45], with respect to the accuracy in the prediction of the ignition delay time of an isooctane oxidation mechanism. In the same work [45], a more accurate isomer lumping strategy was proposed. Specifically, data for the isomer distributions $\alpha_{m,l}$ were gathered from simulations of the detailed mechanism and it was assumed that k_L can be parameterized as a temperature-dependent analytical function. Using least-square regression, values of k_L , calculated as in Eq. (10), were fitted to an Arrhenius-like expression:

$$f(T) = AT^b e^{-\frac{E}{RT}}. \quad (11)$$

In the above, R is the universal gas constant. After fitting, the resulting values for the unknowns A , b and E will be used as the new pre-exponential factor, temperature exponent and activation energy, respectively.

In PAC however, not all lumped reaction rate coefficients are well approximated by an exponential function. In the isooctane/air plasma

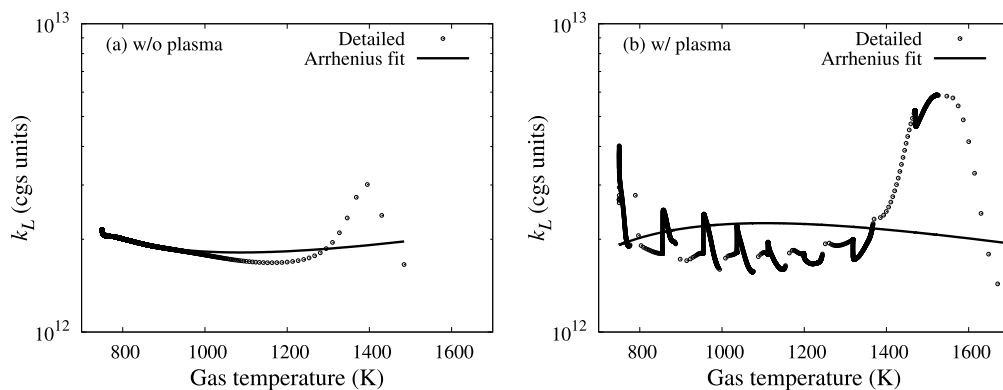


Fig. 2. Evolution and fit Eq. (11) of reaction rate coefficient of lumped reaction (16), as a function of gas temperature. Detailed simulations at $T_0 = 750$ K, $p_0 = 10$ atm, $\phi = 1.0$ (a) without plasma and (b) with plasma generated by pulsed discharges.

mechanism under investigation, for example, the following 4 reactions are difficult to be lumped and fit with exponentials:



with $k_1 \neq k_2 \neq k_3 \neq k_4$. They are lumped into a single reaction:



To highlight the difference in the evolution of k_L when it comes to PAC, two simulations are performed with the detailed isooctane/air plasma mechanism (S2805R18457): one using NSPD with the characteristics described in Table 1 and another one without NSPD. The theoretical values of k_L are calculated from the output data of the simulations. As apparent in Fig. 2(b), in the absence of plasma, the evolution of k_L as a function of gas temperature is adequately described by Eq. (11) in the low temperature region. The inaccurate fit at higher temperatures ($T > 1000$ K) has a negligible impact on the accuracy of the lumping method [45]. In the simulation with plasma however, there are peaks in the evolution of k_L , even at low temperatures and an Arrhenius equation is clearly inappropriate. These peaks, which are due to plasma pulses, appear in the evolution of the relative distributions of isomers $\alpha_{m,i}$ and also affect k_L , since these two quantities are related via Eq. (10).

Indeed, lumping the detailed mechanism with NSPD just by fitting all resulting lumped forward reaction rate coefficients to Eq. (11), produces a lumped mechanism with very low accuracy. This is evident in Fig. 3, where neither the gas temperature nor the fuel breakdown are approximated accurately by the lumped mechanism. Consequently, the error on the prediction of time to ignition in this case is above 12%.

The inadequacy of Eq. (11) in describing the dependence of k_L from temperature is the motivation behind the pursuit of a novel plasma-specific isomer lumping strategy.

To this end, machine learning is employed to build predictive regression models. In particular, we adopt gradient boosting, which is a well-known supervised machine learning method, widely used both in classification and regression problems. “Boosting” refers to the idea of modifying a weak learner for improved accuracy [46]. “Gradient” implies the use of gradient descent to minimize a loss function [47]. Gradient boosting [48,49], is considered an ensemble algorithm, as it uses multiple decision trees, which represent weak learners. Trees are added one at a time and gradient descent minimizes a loss function when doing so, until an acceptable accuracy is obtained. A strong advantage of this method is that it does not need exhaustive data

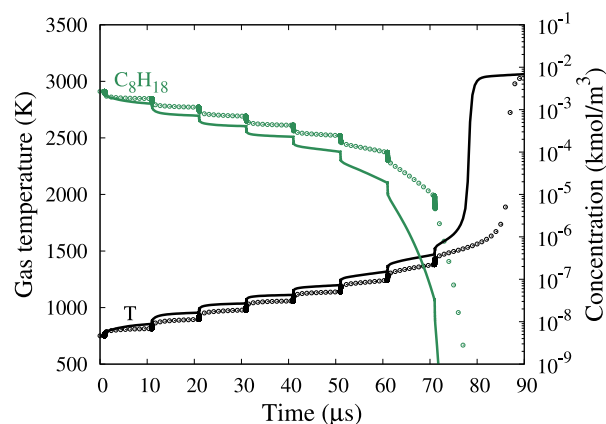


Fig. 3. Temporal evolution of gas temperature and fuel for a C_8H_{18} /air plasma mixture. Comparison between simulations with detailed S2805R18457 (lines) and lumped mechanism, using regular isomer lumping approach (circles) at $T_0 = 750$ K, $p_0 = 10$ atm, $\phi = 1.0$.

pre-processing. Feature scaling, for example, is not applied, as it has negligible impact on ensemble techniques like gradient boosting. This is because these algorithms build trees based on conditions and do not rely on the value range of features.

For this application, eXtreme Gradient Boosting (XGBoost) is used, which is an open-source library for the implementation of gradient boosting [50]. Specifically, we propose a novel isomer lumping approach leveraging XGBoost. The framework consists of the following steps:

1. Data collection of state variables in time (e.g., temperature, pressure, and species concentration) from a simulation with either the detailed mechanism or a very accurate reduced mechanism;
2. Determination of isomers to be lumped;
3. Reaction lumping: identification of reactions which cannot be lumped accurately with the traditional fitting approach presented above;
4. For each one of the lumped reactions identified in Step 3, train a machine learning model with XGBoost, using data from Step 1, in order to correct the forward reaction rate coefficient of the lumped reaction.

There are some important remarks to be made on the actual implementation. Firstly, groups of isomers are formed automatically: the chemical composition of each species present in the mechanism is assessed and those with identical composition (same number of atoms of each element) are identified as a separate group of isomers to be

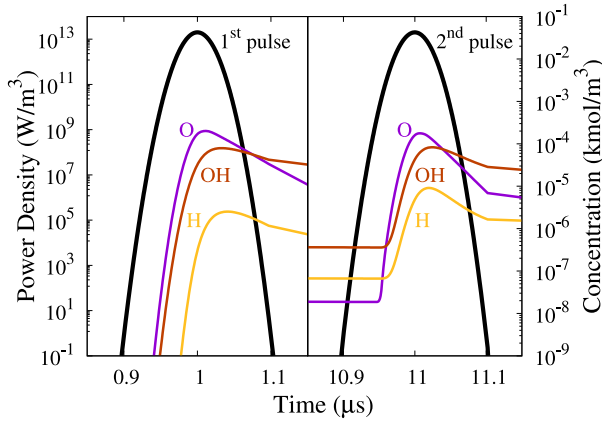


Fig. 4. Temporal evolution of power density and concentrations of O, OH, and H radicals, during the first two discharge pulses for a C_8H_{18} /air plasma mixture. Simulation with the detailed mechanism (S2805R18457) at $T_0 = 750$ K, $p_0 = 10$ atm, $\phi = 1.0$.

lumped. Secondly, the hyper-parameters of the machine learning models (e.g., number of estimators, learning rate, minimum and maximum number of nodes to be added) are tuned with a grid search approach in order to optimize learning.

Two different variations of the machine learning lumping are presented in this work. In the 1D approach, all machine learning regression models are trained using gas temperature as the only input feature. In the 2D approach, apart from gas temperature, a second input feature is also employed. This second input feature is one of the following: O, OH, or H. The procedure of selecting one of these three candidates as the second input feature is performed separately for every model. The final selection is decided for each model based on the feature importance, which is a metric that indicates how useful every feature is in the construction of the trees within the specific model. The candidate feature with the highest importance is selected as the second input feature for the model.

The O, OH, and H radicals are selected as candidate secondary input features, because their production strongly relies on the presence of plasma and their contribution to the ignition process has proved to be significant [51]. The strong interaction between these three radicals and plasma is evident in Fig. 4. During each discharge pulse, the sudden peak in the amount of deposited energy triggers similar peaks in the concentration of O, OH, and H. This implies that the said radicals are strongly related to the plasma physics of the mechanism under examination and could contribute to capturing the irregular evolution of the reaction rate coefficients of lumped reactions.

Training is not performed directly on the lumped reaction rate coefficients k_L , rather on correction factors ξ , which are defined as:

$$\xi = \frac{k_L}{k_{L,Arr}}, \quad (17)$$

where k_L target values are computed using data from simulations and $k_{L,Arr}$ represent approximate values, as they are computed from the fitted Arrhenius equation. Naturally, there is no need for the construction of a regression model, if $k_{L,Arr}$ approximates k_L accurately. For this reason, the fit between k_L and $k_{L,Arr}$ is assessed, using the coefficient of determination R^2 . If R^2 is less than a user-specified threshold (here set equal to 0.99), then k_L is deemed inappropriate to be approximated by the Arrhenius equation alone and the machine learning based regression model is trained for the specific lumped reaction.

The performance of the models obtained from the machine learning approach is evaluated against training and testing results. Here, a priori results are presented in Fig. 5 for the XGBoost single-input

predictive regression model developed for reaction (16). In order to avoid over-fitting, 80% of the available data are used for training and the remaining 20% are used for testing alone. The low normalized root mean squared error (NRMSE) in both training and testing, as well as the high score of the Pearson correlation coefficient between actual and predicted values, highlight the excellent performance of the XGBoost model. Another tool used here against over-fitting is the division of the training data into K folds. At every iteration during training, one of the folds is held out and a model is created using the remaining $(K-1)$ folds. The model is then fitted to the hold-out fold. The process is repeated for every fold and the mean performance after all iterations is reported.

The whole method described above is implemented in Python. The ignition simulations, are carried out using the in-house code PACMAN, which is written in Fortran. In order to be able to use machine learning models in Fortran, a lookup table is used. After training all models, the Python code outputs a file with the necessary information to build the table. At the beginning of a simulation, the file is read by PACMAN and a lookup table created. For the 1D approach, it contains a user-defined number of equally partitioned values of gas temperature and the corresponding correction factor predictions at every temperature point, from each one of the XGBoost models. For the 2D approach, equally partitioned values of the second input feature for every model are created as well. Predictions are made with every XGBoost model for all possible combinations between the partitioned values of the first and second input feature. In this way, a square matrix containing the predictions is created for each model, which allows for interpolation in two dimensions. This matrix is incorporated in the lookup table. It is important to note that in both cases the lookup table is generated using a new set of points, which have not been used by the models during training or testing. During simulation, at every time step and for every reaction that needs correction, the lookup table is accessed and the correction factor results from the interpolation between the existing values in the table, based on the current value(s) of the one (or two) input feature(s). This is repeated multiple times until CVODE converges to a solution for the specific time step. Finally, the forward reaction rate coefficient of the lumped reaction, as calculated by the Arrhenius equation, is multiplied by the correction factor.

4. Results and discussion

The detailed isooctane/air plasma mechanism (S2805R18457) is used to assess the performance of P-DRGEP and the novel isomer lumping approach. A reduced mechanism is created with P-DRGEP first. Then, the lumping approach using XGBoost is applied to create an even more compact lumped mechanism. Results related to the performance of both reduction steps in zero-dimensional simulations with various initial conditions are presented and analyzed in this Section.

4.1. Reduction with P-DRGEP

Error tolerances are imposed for the set of desired targets. At every iteration of P-DRGEP, species and reactions are removed, based on the magnitude of their associated coefficients [34]. Once one of the error tolerances is exceeded, the process stops. For this application, a maximum error of 5% is set for τ_{ig} and 40% for energy transfer metrics, related to losses due to excitation, ionization and impact dissociation. The reactor is initialized at temperature $T = T_e = 750$ K, pressure $p_0 = 10$ atm, for the C_8H_{18} /air mixture at stoichiometric conditions.

With all these options set, a reduced mechanism of 415 species and 4716 reactions (S415R4716) is generated using P-DRGEP, forward and backward reactions counted separately. This is a reduction of around 85% in the total number of species and 75% in the total number of reactions. The excellent agreement between skeletal and detailed mechanisms is apparent from the data in Fig. 6. The error on τ_{ig} prediction is less than 2%.

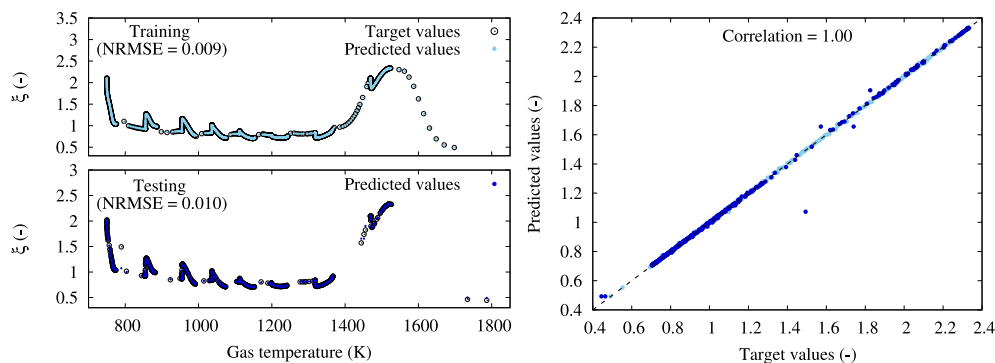


Fig. 5. XGBoost regression results for the prediction of correction factors for reaction (16), using a single-input model. Left: training (top) and testing (bottom). Right: correlation between predicted and actual values. Data originate from simulation with a reduced mechanism at $T_0 = 750$ K, $p_0 = 10$ atm, $\phi = 1.0$.

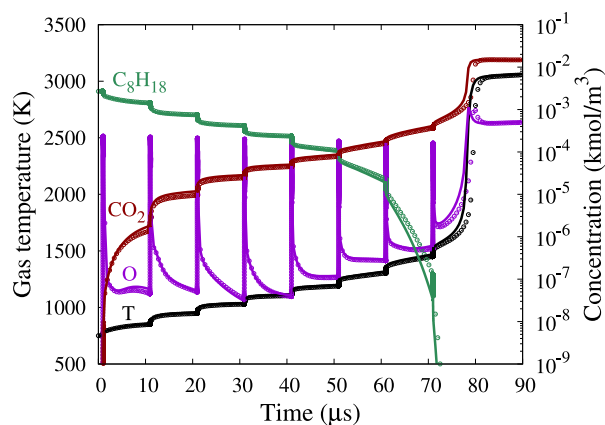


Fig. 6. Temporal evolution of gas temperature and various key combustion species for a C_8H_{18} /air plasma mixture. Comparison between simulations with detailed S2805R18457 (lines) and reduced S415R4716 (circles) mechanisms at $T_0 = 750$ K, $p_0 = 10$ atm, $\phi = 1.0$.

The ability of the skeletal mechanism to reproduce the behavior of the detailed one for a range of initial conditions is always one of the main goals during reduction. Here temperatures ranging from 750 K to 1200 K in lean and rich conditions, with equivalence ratios from 0.75 to 1.50, are explored. Results for the prediction of τ_{ig} are presented in Fig. 7(c), for every zero-dimensional reactor simulation executed with the reduced mechanism. The light blue area in the plots represents a $\pm 5\%$ deviation from reference values predicted by the detailed mechanism. This is the same tolerance used for the P-DRGEP reduction for the specific metric. In any case, the absolute error is less than 3%, when comparing predictions between detailed and reduced mechanisms.

4.2. Application of the novel isomer lumping strategy

With an accurate reduced mechanism available, the lumping procedure is conducted. The Python-based lumping software reads all species in S415R4716 and identifies 49 groups of unique isomers. The forward reaction rate coefficient of every newly formed lumped reaction is fit to Eq. (11), in order to obtain the new Arrhenius parameters. If the fit is not accurate, an XGBoost model is trained for the reaction, using output data from a simulation with the reduced mechanism S415R4716 at $T_0 = 750$ K, $p_0 = 10$ atm and $\phi = 1.0$. When the reaction to be lumped with XGBoost is reversible, it is split into two forward reactions first and two separate models are created, one for each reaction direction. Around 550 models are trained and their predictions of correction factors ξ are written into the file, which is then read by PACMAN to generate lookup tables at run-time.

The final lumped mechanism consists of 300 species and 3827 reactions (S300R3827), always counting forward and backward reactions separately. When compared to the reduced mechanism after the P-DRGEP step, this translates to an additional 28% reduction in the number of species and 19% reduction in the number of reactions.

4.2.1. Single-input machine learning models (1D approach)

All models are trained using gas temperature as the only input feature. The performance of the 1D lumped mechanism is evaluated by simulating a plasma-assisted ignition of the C_8H_{18} /air mixture at the same initial conditions as those used to train the predictive regression machine learning models. The temporal evolution of temperature and concentration of a number of species are compared to those from a simulation with detailed mechanism in Fig. 8 and a good agreement is observed. For the same conditions, a similar comparison is carried out for the concentration of two lumped isomer groups, namely $C_8H_{17}(L)$ and $C_8H_{17}O_2(L)$, presented in Fig. 9. There is an almost perfect agreement between the results obtained with the lumped model and the target values. The latter are computed via the additivity rule, as expressed in Eq. (7), using simulation data from the detailed model. This indicates that the mass balance is respected in the simulations with the lumped mechanism.

It is interesting to note that the high accuracy of the mechanism produced by P-DRGEP, which is evident in Figs. 6 and 7(c), allows the training of all models with data from the reduced mechanism (S415R4716), instead of from the detailed one (S2805R18457). This is computationally convenient when training models with XGBoost for a broad set of initial conditions. Another important advantage is that there is no need to gather training data from multiple simulations with different initial conditions. As it can be seen by the gas temperature rise in Fig. 6, a single simulation provides enough data, to train the models over a wide range of temperatures.

Focusing again on the performance of the lumped mechanism, we assess its accuracy when used in predictions of plasma-assisted ignition of mixtures at initial thermochemical states different from those used during training. For this purpose, multiple zero-dimensional simulations were executed over the same range examined with the reduced mechanism. For each simulation, the time to ignition τ_{ig} is plotted in Fig. 10(c) together with the corresponding value obtained with the detailed mechanism and a $\pm 5\%$ error band.

Even though the lumping and training of models were based on data from simulation with specific initial conditions, the resulting lumped mechanism is accurate over a much broader range of temperatures and equivalence ratios: $750 \leq T_0 \leq 1200$ K and $0.75 \leq \phi \leq 1.50$. More specifically, for lean conditions, the average absolute error in the prediction of time to ignition is around 1%, with a maximum value of 2.3%. Similar results are observed for stoichiometric conditions, with an average error of 1.5%, and a maximum error of 3.3%. For simulations with rich mixtures, a maximum error of 6% is observed for

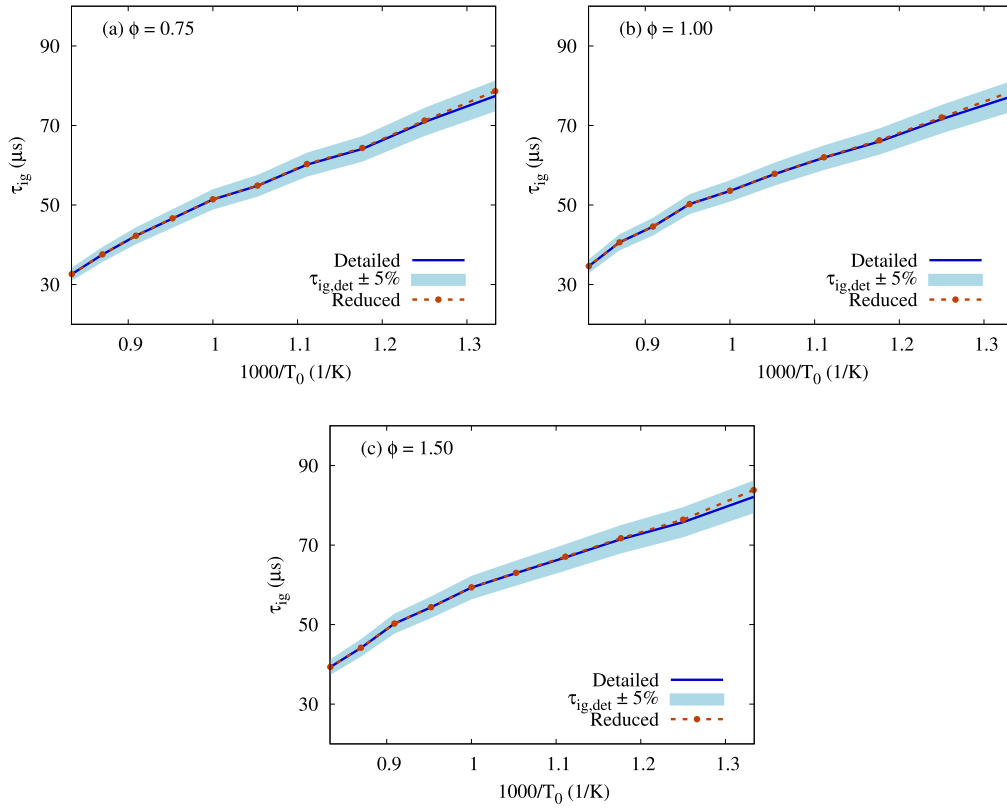


Fig. 7. Comparison between the time to ignition prediction made by simulations with detailed S2805R18457 (lines) and reduced S415R4716 mechanism (dots) at $p_0 = 10$ atm: (a) $\phi = 0.75$, (b) $\phi = 1.00$ and (c) $\phi = 1.50$, at various initial temperatures.

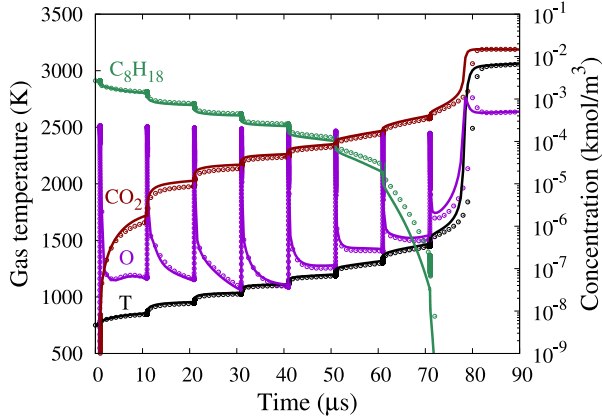


Fig. 8. Temporal evolution of gas temperature and various species for a C_8H_{18} /air plasma mixture. Comparison between simulations with detailed S2805R18457 (lines) and 1D lumped S300R3827 (circles) mechanisms at $T_0 = 750$ K, $p_0 = 10$ atm, $\phi = 1.0$.

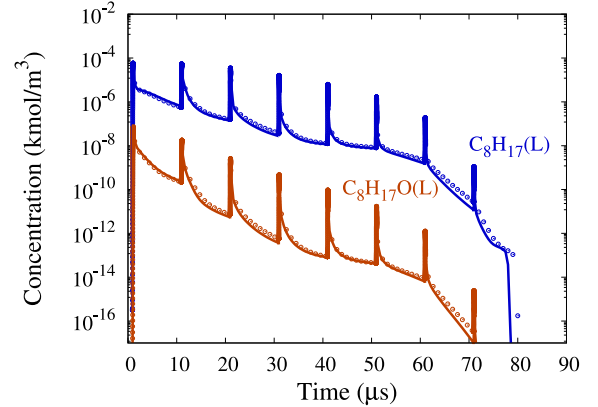


Fig. 9. Temporal evolution of two lumped groups for a C_8H_{18} /air plasma mixture. Comparison between simulations with detailed S2805R18457 (lines) and lumped S300R3827 (circles) mechanisms at $T_0 = 750$ K, $p_0 = 10$ atm, $\phi = 1.0$.

low temperatures. The average absolute error remains low and close to 2.2%.

The alleviation of computational demand is also significant: 87% decrease in central processing unit (CPU) time is observed for a simulation with the 1D lumped mechanism at $T_0 = 750$ K, $P_0 = 10$ atm, and $\phi = 1.00$, when compared to the respective simulation using the detailed mechanism.

Apart from the initial thermochemical conditions, different pulse settings were also tested for the lumped mechanism. More specifically, the pulse frequency f was varied over the range 50–400 kHz and single pulse power density P over the range 10–75 MW/cm³. Thus, different values for the energy deposition rate W were explored, where $W = Ef$,

and E is the energy density per pulse, which is a function of P . For each value of W , a simulation was performed at $p_0 = 10$ atm, $T_0 = 750$ K, and $\phi = 1.00$. τ_{ig} for every configuration is shown in Fig. 11, compared to predictions made by corresponding simulations with the reduced mechanism (S415R4716), with a $\pm 10\%$ error band. Larger errors are observed for lower values of W . In order to make the simulations with the lumped mechanism more accurate for low values of W , data from simulations with lower W should be included in the training set as well.

4.2.2. Two-input machine learning models (2D approach)

In this case, the same models developed for the lumped mechanism S300R3827 are trained with two input features, instead of one: the first

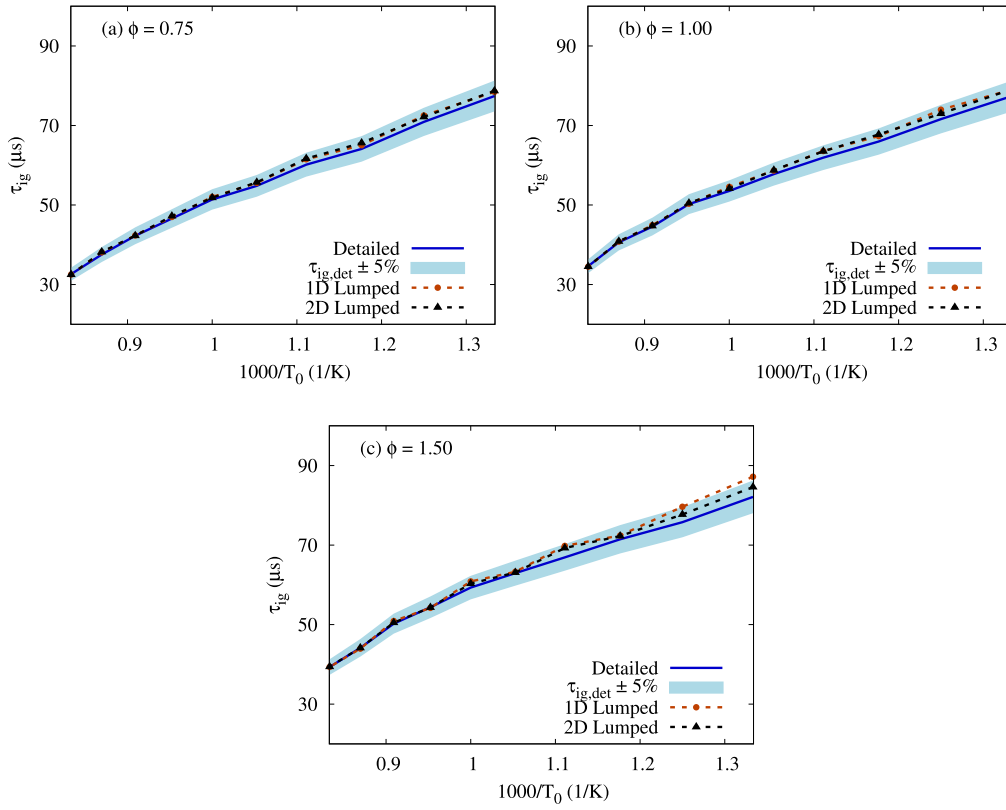


Fig. 10. Comparison between the time to ignition with detailed S2805R18457 (lines), 1D lumped S300R3827 (dots), and 2D lumped S300R3827 (triangles) mechanisms at $p_0 = 10$ atm: (a) $\phi = 0.75$, (b) $\phi = 1.00$ and (c) $\phi = 1.50$, at various initial temperatures.

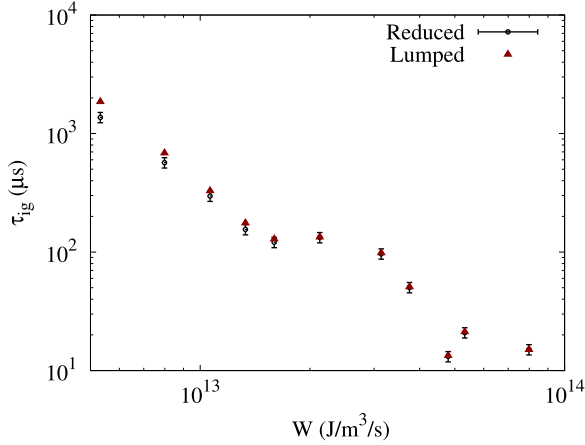


Fig. 11. Comparison between time to ignition predictions made by simulations with reduced S415R4716 (dots with $\pm 10\%$ error bars) and 1D lumped S300R3827 mechanism (triangles), for different values of energy deposition rate W .

one is the gas temperature and the second one is selected for every model, based on the feature importance ranking of the three candidate species (O, OH, and H), as described above. It is interesting to note that almost half of the models use O as their second input feature. Another 35% use OH and the rest H.

The promising potential of this approach is already evident by assessing some preliminary results. Indicatively, NRMSE, both in training and testing of the models, is decreased by half compared to the single-input approach.

These a priori good results are confirmed by the numerical simulations performed using the two-input machine learning models, as it can be observed in Fig. 10(c). The average absolute error on τ_{ig} with

the 2D lumped mechanism is just 1.4% in the same range of initial conditions examined for the 1D approach: $p_0 = 10$ atm, $750 \leq T_0 \leq 1200$ K, and $0.75 \leq \phi \leq 1.50$. The 2D machine learning lumping approach offers better modeling capabilities compared to a classic least square approach. The superiority of the 2D lumped mechanism is evident from the improved accuracy in the set of initial conditions where the 1D version produced the worst results, namely the low temperature region in rich conditions. Fuel concentration and gas temperature evolution in time are presented for the two different isomer lumping approaches in Fig. 12 and Fig. 13, for simulations at initial conditions $\phi = 1.50$, $p_0 = 10$ atm, $T_0 = 750$ K, and $\phi = 1.50$, $p_0 = 10$ atm, $T_0 = 800$ K respectively. Results from simulations with the detailed mechanism are also shown. The 2D lumped mechanism approximates more accurately the fuel breakdown and gas temperature rise in both cases, thus achieving low errors or τ_{ig} : 2.8% for $T_0 = 750$ K and 2.5% for $T_0 = 800$ K. The respective errors with the 1D lumped mechanism are 6% and 5%.

As far as the computational time is concerned, a simulation at $T_0 = 750$ K, $P_0 = 10$ atm, and $\phi = 1.00$ using the 2D lumped mechanism is 78% faster in terms of CPU time, compared to the detailed mechanism.

As for the training phase, a single simulation is sufficient to produce the needed data for both the 1D and 2D approaches. Adding one more input feature to a model, while keeping the same output, does not necessarily require a larger training dataset, since the complexity of the model does not increase significantly [52]. Thus, an adequate time step size during the simulation, produces enough training data for both models.

5. Conclusions

This work focuses on the reduction of a detailed iso-octane/air plasma-assisted combustion mechanism, consisting of 2805 species and 18457 reactions, by combining two novel techniques. Using P-DRGEP, a reduced mechanism of 415 species and 4716 reactions is produced. The

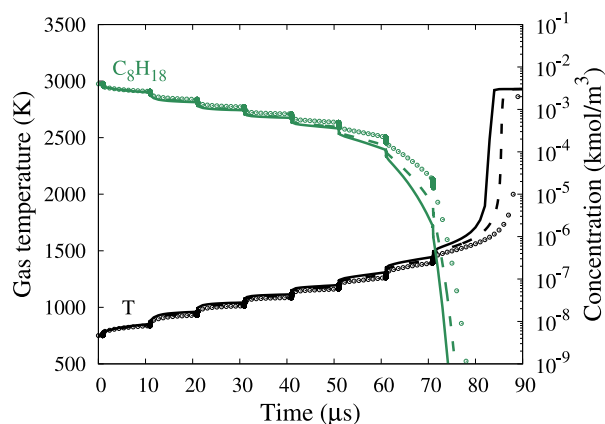


Fig. 12. Temporal evolution of gas temperature and fuel for a C_8H_{18} /air mixture. Comparison between simulations with detailed S2805R18457 (lines), 1D lumped mechanism S300R3827 (circles), and 2D lumped mechanism S300R3827 (dashed lines) at $T_0 = 750$ K, $p_0 = 10$ atm, $\phi = 1.50$.

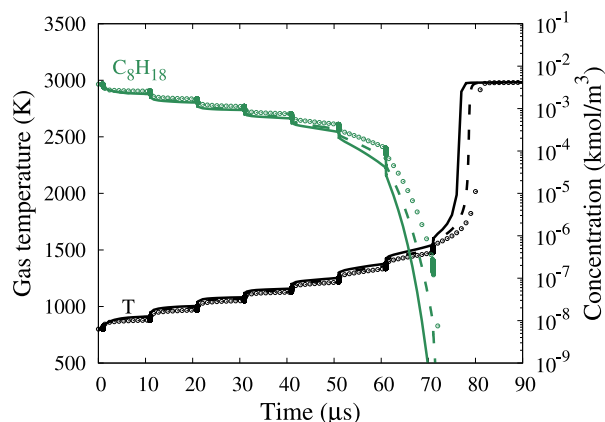


Fig. 13. Temporal evolution of gas temperature and fuel for a C_8H_{18} /air mixture. Comparison between simulations with detailed S2805R18457 (lines), 1D lumped mechanism S300R3827 (circles), and 2D lumped mechanism S300R3827 (dashed lines) at $T_0 = 800$ K, $p_0 = 10$ atm, $\phi = 1.50$.

mechanism has excellent predictive capabilities over a wide range of initial conditions: temperatures from 750 K to 1200 K and equivalence ratios from 0.75 to 1.50. Thus, the ability of the reduction framework to handle kinetic mechanisms of such a large size and complexity is demonstrated for the first time. In addition, a plasma-specific isomer lumping approach is proposed and its viability and accuracy demonstrated. With the gradient boosting machine learning method and data from a zero-dimensional reactor simulation that employ the reduced mechanism, predictive regression models are developed, which are found to describe accurately lumped reaction rate coefficients. Using this novel methodology, a lumped mechanism is developed, which contains 300 species and 3827 reactions. Two variations of this method are presented: one with models that use just gas temperature as input feature and another with two-input models, which apart from gas temperature also employ one of the radicals O, H, or OH as a second input feature. In the former approach and over the broad range of initial conditions used to test the reduced mechanism with zero-dimensional simulations, absolute errors fall within 6% on time to ignition, when compared to the detailed mechanism. The latter approach produces even lower errors, which do not exceed 3%.

CRediT authorship contribution statement

Georgios Rekkas-Ventiris: Main researcher, Wrote paper. **Alfredo Duarte Gomez:** Feedback on simulations. **Nicholas Deak:** Feedback

on simulations. **Nicholas Kincaid:** Feedback on reduction strategies. **Perrine Pepiot:** Feedback on lumping. **Fabrizio Bisetti:** Designed applications. **Aur lie Bellemans:** Designed research, Supervision.

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

N. Kincaid and P. Pepiot were supported by National Science Foundation (NSF, USA) Grant DGE-2139899, A. Duarte Gomez and F. Bisetti were sponsored in part by NSF, USA Grant No. 1903775 and U.S. Department of Energy (DOE, USA) Grant DE-EE0008874.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.combustflame.2023.113252>.

References

- [1] W. Sun, S.H. Won, Y. Ju, In situ plasma activated low temperature chemistry and the S-curve transition in DME/Oxygen/Helium mixture, *Combust. Flame* 161 (8) (2014) 2054–2063.
- [2] F. Wang, C. Jiang, A. Kuthi, M. Gundersen, J. Sinibaldi, C. Brophy, L. Lee, Transient plasma ignition of hydrocarbon-air mixtures in pulse detonation engines, in: 42nd AIAA Aerospace Sciences Meeting and Exhibit, 2004, paper AIAA-2004-0834.
- [3] T. Ombrello, S.H. Won, Y. Ju, S. Williams, Flame propagation enhancement by plasma excitation of oxygen. Part I: Effects of O₃, *Combust. Flame* 157 (10) (2010) 1906–1915.
- [4] T. Ombrello, S.H. Won, Y. Ju, S. Williams, Flame propagation enhancement by plasma excitation of oxygen. Part II: Effects of O₂ (a1Δg), *Combust. Flame* 157 (10) (2010) 1916–1928.
- [5] T. Ombrello, X. Qin, Y. Ju, A. Gutsol, A. Fridman, C. Carter, Combustion enhancement via stabilized piecewise nonequilibrium gliding arc plasma discharge, *AIAA J.* 44 (1) (2006) 142–150.
- [6] A. Starikovskiy, N. Aleksandrov, Plasma-assisted ignition and combustion, *Prog. Energ. Combust. Sci.* 39 (1) (2013) 61–110.
- [7] Y. Ju, W. Sun, Plasma assisted combustion: Dynamics and chemistry, *Prog. Energ. Combust. Sci.* 48 (2015) 21–83.
- [8] S. Starikovskaia, D.A. Lacoste, G. Colonna, Non-equilibrium plasma for ignition and combustion enhancement, *Eur. Phys. J. D* 75 (8) (2021) 1–25.
- [9] N. Popov, Kinetics of plasma-assisted combustion: effect of non-equilibrium excitation on the ignition and oxidation of combustible mixtures, *Plasma Sources Sci. Technol.* 25 (4) (2016) 043002.
- [10] N. Tsolas, R.A. Yetter, Kinetics of plasma assisted pyrolysis and oxidation of ethylene. Part 1: Plasma flow reactor experiments, *Combust. Flame* 176 (2017) 534–546.
- [11] N. Tsolas, R.A. Yetter, I.V. Adamovich, Kinetics of plasma assisted pyrolysis and oxidation of ethylene. Part 2: Kinetic modeling studies, *Combust. Flame* 176 (2017) 462–478.
- [12] S.B. Leonov, D.A. Yarrantsev, A.P. Napartovich, I.V. Kochetov, Plasma-assisted combustion of gaseous fuel in supersonic duct, *IEEE Trans. Plasma Sci.* 34 (6) (2006) 2514–2525.
- [13] I. Kosarev, N. Aleksandrov, S. Kindysheva, S. Starikovskaia, A.Y. Starikovskii, Kinetics of ignition of saturated hydrocarbons by nonequilibrium plasma: CH₄-containing mixtures, *Combust. Flame* 154 (3) (2008) 569–586.
- [14] W. Sun, Y. Ju, A multi-timescale and correlated dynamic adaptive chemistry and transport (CO-DAC) method for computationally efficient modeling of jet fuel combustion with detailed chemistry and transport, *Combust. Flame* 184 (2017) 297–311.
- [15] J. Griffiths, Reduced kinetic models and their application to practical combustion systems, *Prog. Energ. Combust. Sci.* 21 (1) (1995) 25–107.
- [16] P. Pepiot, L. Cai, H. Pitsch, Model reduction and lumping procedures, in: T. Faravelli, F. Manenti, E. Ranzi (Eds.), *Computer Aided Chemical Engineering*, Elsevier, 2019, pp. 799–827.
- [17] A. Massias, D. Diamantis, E. Mastorakos, D. Goussis, Global reduced mechanisms for methane and hydrogen combustion with nitric oxide formation constructed with CSP data, *Combust. Theor. Model.* 3 (2) (1999) 233.

- [18] M. Valorani, F. Creta, F. Donato, H. Najm, D. Goussis, A CSP-based skeletal mechanism generation procedure: Auto-ignition and premixed laminar flames in n-heptane/air mixtures, in: *European Conference on Computational Fluid Dynamics*, 2006.
- [19] P.P. Ciottoli, R.M. Galassi, P.E. Lapenna, G. Leccese, D. Bianchi, F. Nasuti, F. Creta, M. Valorani, CSP-based chemical kinetics mechanisms simplification strategy for non-premixed combustion: An application to hybrid rocket propulsion, *Combust. Flame* 186 (2017) 83–93.
- [20] T. Løvs, D. Nilsson, F. Mauss, Automatic reduction procedure for chemical mechanisms applied to premixed methane/air flames, *Proc. Combust. Inst.* 28 (2) (2000) 1809–1815.
- [21] Q.-D. Wang, Y.-M. Fang, F. Wang, X.-Y. Li, Systematic analysis and reduction of combustion mechanisms for ignition of multi-component kerosene surrogate, *Proc. Combust. Inst.* 34 (1) (2013) 187–195.
- [22] J.C. Sutherland, A. Parente, Combustion modeling using principal component analysis, *Proc. Combust. Inst.* 32 (1) (2009) 1563–1570.
- [23] B.J. Isaac, A. Coussement, O. Gicquel, P.J. Smith, A. Parente, Reduced-order PCA models for chemical reacting flows, *Combust. Flame* 161 (11) (2014) 2785–2800.
- [24] A. Stagni, A. Frassoldati, A. Cuoci, T. Faravelli, E. Ranzi, Skeletal mechanism reduction through species-targeted sensitivity analysis, *Combust. Flame* 163 (2016) 382–393.
- [25] R. Li, A.A. Konnov, G. He, F. Qin, D. Zhang, Chemical mechanism development and reduction for combustion of NH₃/H₂/CH₄ mixtures, *Fuel* 257 (2019) 116059.
- [26] T. Lu, C.K. Law, A directed relation graph method for mechanism reduction, *Proc. Combust. Inst.* 30 (1) (2005) 1333–1341.
- [27] P. Pepiot-Desjardins, H. Pitsch, An efficient error-propagation-based reduction method for large chemical kinetic mechanisms, *Combust. Flame* 154 (1–2) (2008) 67–81.
- [28] T. Lu, C.K. Law, Strategies for mechanism reduction for large hydrocarbons: n-heptane, *Combust. Flame* 154 (1–2) (2008) 153–163.
- [29] A. Stagni, A. Cuoci, A. Frassoldati, T. Faravelli, E. Ranzi, Lumping and reduction of detailed kinetic schemes: an effective coupling, *Ind. Eng. Chem. Res.* 53 (22) (2014) 9004–9016.
- [30] S.S. Ahmed, F. Mauß, G. Moréac, T. Zeuch, A comprehensive and compact n-heptane oxidation model derived using chemical lumping, *Phys. Chem. Chem. Phys.* 9 (9) (2007) 1107–1126.
- [31] K. Peerenboom, A. Parente, T. Kozák, A. Bogaerts, G. Degrez, Dimension reduction of non-equilibrium plasma kinetic models using principal component analysis, *Plasma Sources Sci. Technol.* 24 (2) (2015) 025004.
- [32] A. Bellemans, T. Magin, A. Coussement, A. Parente, Reduced-order kinetic plasma models using principal component analysis: Model formulation and manifold sensitivity, *Phys. Rev. Fluids* 2 (7) (2017) 073201.
- [33] A. Bellemans, N. Deak, F. Bisetti, Development of skeletal kinetics mechanisms for plasma-assisted combustion via principal component analysis, *Plasma Sources Sci. Technol.* 29 (2) (2020) 025020.
- [34] A. Bellemans, N. Kincaid, N. Deak, P. Pepiot, F. Bisetti, P-DRGEP: a novel methodology for the reduction of kinetics mechanisms for plasma-assisted combustion applications, *Proc. Combust. Inst.* 38 (4) (2021) 6631–6639.
- [35] D. Davidson, M. Oehlschlaeger, J. Herbon, R. Hanson, Shock tube measurements of iso-octane ignition times and OH concentration time histories, *Proc. Combust. Inst.* 29 (1) (2002) 1295–1301.
- [36] N. Atef, G. Kukkadapu, S.Y. Mohamed, M. Al Rashidi, C. Banyon, M. Mehl, K.A. Heufer, E.F. Nasir, A. Alfazazi, A.K. Das, et al., A comprehensive iso-octane combustion model with improved thermochemistry and chemical kinetics, *Combust. Flame* 178 (2017) 111–134.
- [37] Z.S. Eckert, Energy Transfer in Non-Equilibrium Reacting Gas Flows: Applications in Plasma Assisted Combustion and Chemical Gas Lasers (Ph.D. thesis), The Ohio State University, Columbus, OH, USA, 2018.
- [38] N. Deak, A. Bellemans, F. Bisetti, Plasma-assisted ignition of methane/air and ethylene/air mixtures: Efficiency at low and high pressures, *Proc. Combust. Inst.* 38 (4) (2021) 6551–6558.
- [39] Y. Bechane, B. Fiorina, Numerical investigations of turbulent premixed flame ignition by a series of Nanosecond Repetitively Pulsed discharges, *Proc. Combust. Inst.* 38 (4) (2021) 6575–6582.
- [40] S.D. Cohen, A.C. Hindmarsh, P.F. Dubois, CVODE, a stiff/nonstiff ODE solver in C, *Comput. Phys.* 10 (2) (1996) 138–143.
- [41] R.J. Kee, F.M. Rupley, J.A. Miller, Chemkin-II: A Fortran Chemical Kinetics Package for the Analysis of Gas-Phase Chemical Kinetics, Tech. rep., Sandia National Laboratories, Livermore, CA, USA, 1989.
- [42] A. Sharma, V. Subramaniam, E. Solmaz, L.L. Raja, Fully coupled modeling of nanosecond pulsed plasma assisted combustion ignition, *J. Phys. D Appl. Phys.* 52 (9) (2018) 095204.
- [43] E. Ranzi, M. Dente, A. Goldaniga, G. Bozzano, T. Faravelli, Lumping procedures in detailed kinetic modeling of gasification, pyrolysis, partial oxidation and combustion of hydrocarbon mixtures, *Prog. Energy Combust. Sci.* 27 (1) (2001) 99–139.
- [44] H. Huang, M. Fairweather, J. Griffiths, A. Tomlin, R. Brad, A systematic lumping approach for the reduction of comprehensive kinetic models, *Proc. Combust. Inst.* 30 (1) (2005) 1309–1316.
- [45] P. Pepiot-Desjardins, H. Pitsch, An automatic chemical lumping method for the reduction of large chemical kinetic mechanisms, *Combust. Theor. Model.* 12 (6) (2008) 1089–1108.
- [46] M. Kearns, Thoughts on Hypothesis Boosting, Tech. rep., 1988.
- [47] L. Mason, J. Baxter, P. Bartlett, M. Frean, Boosting algorithms as gradient descent, *Adv. New. In.* 12 (1999) 512–518.
- [48] J.H. Friedman, Greedy function approximation: a gradient boosting machine, *Ann. Statist.* (2001) 1189–1232.
- [49] L. Breiman, Prediction games and arcing algorithms, *Neural Comput.* 11 (7) (1999) 1493–1517.
- [50] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, et al., Xgboost: Extreme Gradient Boosting, 2015, R package version 0.4-2.
- [51] Y. Tang, Q. Yao, J. Zhuo, S. Li, Plasma-assisted pyrolysis and ignition of pre-vaporized n-heptane, iso-octane and n-decane, *Fuel* 289 (2021) 119899.
- [52] V. Lakshmanan, S. Robinson, M. Munn, Machine Learning Design Patterns, O'Reilly Media, 2020.