Preprint Cambridge Centre for Computational Chemical Engineering ISSN 1473 – 4273

Automated calibration of a poly(oxymethylene) dimethyl ether oxidation mechanism using knowledge-graph technology

Jiaru Bai [1], Rory Geeson [1,2], Feroz Farazi [1], Sebastian Mosbach [1,3], Jethro Akroyd [1,3], Eric J. Bringley [1], Markus Kraft [1,3,4]

released: November 10, 2020

[1] Department of Chemical Engineering and Biotechnology
University of Cambridge
Philippa Fawcett Drive
Cambridge, CB3 0AS
United Kingdom

[2] Department of Computer Science and Technology
University of Cambridge
15 JJ Thomson Avenue
Cambridge, CB3 0FD
United Kingdom

[3] CARES
Cambridge Centre for Advanced Research and Education in Singapore
1 Create Way
CREATE Tower, #05-05
Singapore, 138602

[4] School of Chemical and Biomedical Engineering
Nanyang Technological University
62 Nanyang Drive
Singapore, 637459

Preprint No. 262

UNIVERSITY OF CAMBRIDGE

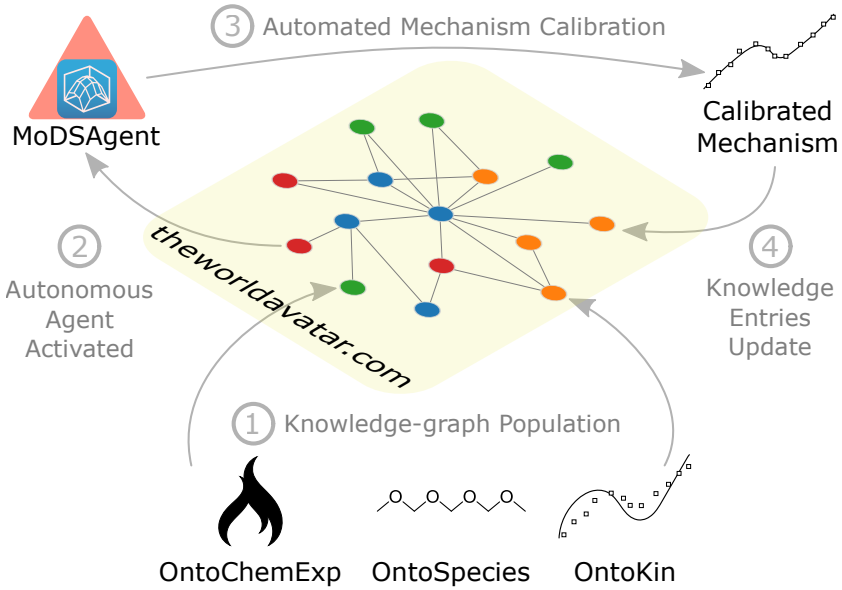
Keywords: Knowledge-graphs, automated mechanism calibration, combustion experiment ontology, chemical mechanisms, PODE

Edited by
Computational Modelling Group
Department of Chemical Engineering and Biotechnology
University of Cambridge
Philippa Fawcett Drive
Cambridge, CB3 0AS
United Kingdom

E-Mail: mk306@cam.ac.uk
World Wide Web: https://como.ceb.cam.ac.uk/

Abstract

In this paper, we develop a knowledge-graph based framework for the automated calibration of combustion reaction mechanisms and demonstrate its effectiveness on a case study of poly(oxymethylene) dimethyl ether ($PODE_n$, where $n = 3$) oxidation. We develop an ontological representation for combustion experiments, OntoChemExp, that allows for the semantic enrichment of experiments within the J-Park Simulator (JPS, theworldavatar.com), an existing cross-domain knowledge-graph. OntoChemExp is fully capable of supporting experimental results in the Process Informatics Model (PrIMe) database. Following this, a set of software agents are developed to perform experimental results retrieval, sensitivity analysis, and calibration tasks. The sensitivity analysis agent is used for both generic sensitivity analyses and reaction selection for subsequent calibration. The calibration process is performed as a sampling task followed by an optimisation task. The agents are designed for use with generic models but are demonstrated with ignition delay time and laminar flame speed simulations. We find that calibration times are reduced while accuracy is increased compared to manual calibration, achieving fittings 92% more accurate. Further, we demonstrate how this workflow is implemented as an extension of the JPS.



Highlights

- An ontology for representing certain types of combustion experiments is created.
- Autonomous agents for sensitivity analysis and mechanism calibration are developed.
- Interactions between the agents and the knowledge-graph are managed by a coordination agent.
- A $PODE_3$ mechanism is calibrated against ignition delay time and laminar flame speed.
- The proposed knowledge-graph approach outperforms manual calibration substantially.

Contents

1 Introduction

The contribution of human activity to climate change and the potential for ecological devastation this presents has become a widely accepted fact within the scientific community [27]. One of the key contributors to this effect is the release of greenhouse gases from combustion processes. Improvements in the design of energy conversion system has already resulted in significant reductions in their contribution to greenhouse gas emissions and presents one of the potential paths towards even lower emissions in the future. Another approach involves the use of alternative fuels, particularly synthetic fuels, offering potential for reductions to pollution and greenhouse gas emissions.

Modern workflows for the design and optimisation of combustion equipment and devices now routinely employ computational modelling techniques. These are most often used to screen designs and offer invaluable insights into processes occurring within [19]. Thus, to achieve the desired reduction in climate change potential from combustion equipment, provision of accurate combustion chemistry mechanisms is becoming essential.

In practice, the development of these combustion chemical mechanisms consists of two parts: mechanism generation and mechanism calibration. The first step constructs a tentative mechanism that maps the reaction pathways via elementary reaction generation and selection, and the second step adjusts the rate parameters, attempting to faithfully reproduce experimental observations.

Much of the construction of combustion mechanisms involves the selection and combination of reaction families, elementary reactions and sub-mechanisms from various existing mechanisms. This is facilitated by the CHEMKIN [26] mechanism format, acting as a *de facto* standard for mechanism sharing within the combustion community. Aspects not formally captured by this format include semantics and provenance. This allows errors to propagate through combustion models due to the inability to ensure the quality of individual reactions and the difficulty of tracking their origin.

Moreover, accuracy and consistency of combustion mechanisms is not guaranteed across applications, even with well-calibrated sub-mechanisms [17]. These problems are further exacerbated when the scale of the mechanisms is considered; potentially hundreds of species and thousands of reactions may be involved. This leads to the manual curation and provenance determination of all the components of these mechanisms being a near-impossible task for researchers. Even if attempts were to be made, these are likely to fall foul to human error and so a very real need for an automated approach to this mechanism development task is present.

Reaction Mechanism Generator (RMG) [18] is one of the available tools for the automation of the first stage of mechanism development. The technique is based on the use of a set of chemical rules to predict chemical pathways along with a database of chemical properties. Values of unknown chemical properties are estimated on-the-fly. One of the methods used for this purpose is that of Li et al. [31], employing a graph neural network (GNN) on molecular graphs to make formation enthalpy predictions. The GNN was trained on a dataset generated at B3LYP/6-31G(2df,p) level of density functional theory (DFT). The framework further incorporated quantum chemistry calculations for additional model training in case of uncertain predictions, improving accuracy and generalisability.

Progress has also been made in automating transition state theory calculations [2], an important step towards generating accurate reaction rates. However, generating a chemical mechanism with purely quantum calculated rate parameters remains infeasible, given that even the most detailed model would not include all possible pathways [27]. This necessitates the use of automated calibration processes for these coefficients to reproduce experimental results.

The mechanism development and curation process may be improved in two ways: 1) semantically focused and machine-interpretable formats for mechanism representation with clear provenance should be used and 2) automated updating and verification of existing mechanisms throughout their lifetimes should be implemented. Task 1 has received attention within the community, with various efforts to create standardised databases of combustion data with unique identifiers and easily processable formatting. One of the key efforts in this direction is that of the Process Informatics Model (PrIMe) [16] database, containing combustion data in a standardised eXtensible Markup Language (XML) format.

Although projects such as PrIMe have started the process, further strides towards a fully provenanced and machine-interpretable standard for the combustion community must continue. The relatively granular structure of databases and the lack of semantics prevents them from reaching the true potential of modern technologies within artificial intelligence (AI) and knowledge discovery. One of the potential technologies to aid with these processes is knowledge-graphs. A knowledge-graph is a dynamic knowledge eco-system interconnecting individual pieces of information as well as software. This is implemented using ontologies, commonly written in Web Ontology Language (OWL), to define the abstract concepts and relations that are shared within the knowledge-graph. Such a design offers both clear semantics to its entries and a highly linked structure, thus enabling item locating, easy provenance determination, and reasoning over entries with software agents.

In the context of combustion chemistry, we have developed OntoKin [12] as an ontology for representing chemical kinetic reaction models. We have also developed OntoCompChem [29], based on Chemical Markup Language (CML [39]), to store quantum chemistry calculations. We further introduced OntoSpecies [13] for unique chemical species identification, generating a more comprehensive description of combustion chemistry with the three ontologies seamlessly linked together to enable consistency checking across multiple mechanisms [14]. This framework was further enhanced by the development of a set of autonomous agents for quantum chemistry and enthalpy of formation calculations, employing error-cancelling balanced reactions for the enthalpy of formation calculations [37].

The purpose of this work is to propose a knowledge-graph based framework for automated combustion mechanism calibration. This forms a clear path for achieving the second of the highlighted tasks. We aim to achieve this by constructing an ontology to link combustion experiment measurements to chemical reaction mechanisms and developing a set of software agents that automatically perform mechanism calibration against ignition delay time and laminar flame speed experimental data. A demonstration of this is performed on a reduced poly(oxymethylene) dimethyl ether 3 ($PODE_n$, where n=3) mechanism [32]. This alternative fuel, with a molecular formula of $CH_3O(CH_2O)_3CH_3$, is deliberately chosen for its current interest as a fuel additive to help with the push for cleaner and more

efficient vehicles. The additive has been demonstrated to lower soot emissions and improve combustion efficiency in engines [33]. To match the current interest and to further $PODE_3$'s commercialisation, a reduced yet robust mechanism is required, making it an ideal candidate for a demonstration of our framework.

The presentation of this paper is structured as follows: section 2 situates this work by introducing the wider knowledge-graph based project; section 3 details the components that together form the framework; section 4 presents the demonstration case of $PODE_3$ with significant improvements in model performance; section 5 concludes the work.

2 The World Avatar

The 'World Avatar' (theworldavatar.com) begins the process of creating a fully interconnected virtual representation of the world through semantic web technologies. The term originated from the idea of the 'Digital Twin' in Industry 4.0, but extended the 'Digital Twin's' representation of factories to conceptualising and representing everything that physically exists. With this vision, the 'World Avatar' aims to standardise the language used across knowledge domains to enable cross-domain communications, offering extensive opportunities for solving more complex and interesting problems [28].

The J-Park Simulator (JPS) [10] is an instantiation of the 'World Avatar' at the intersection of chemical and electrical engineering. The initial scope of JPS is to create a digital replica of the eco-industrial park on Jurong Island, Singapore [38]. The effectiveness of JPS has been demonstrated through its ability to solve many energy related problems. JPS has been applied to the utilisation of waste energy [48], network optimisation of the eco-industrial park [49], and simulations of a carbon tax for scenario analysis in policy making [11].

The versatility of JPS is a result of two key components: modular and reusable ontologies, and inter-operable agents. As illustrated in **Fig. 1**, these components form the core of JPS – a distributed knowledge-graph. By design, JPS employs both in-house ontologies developed by domain experts (*e.g.*, OntoEIP [50], OntoCityGML [10], OntoPowSys [9], *etc.*) and existing ontologies developed by external researchers (*e.g.*, DBpedia [30], OntoCAPE [34], *etc.*). These ontologies are connected and merged within JPS to ensure the depth and breadth of the concepts and relations in the knowledge-graph. Data entries from different sources are described in the languages of these ontologies and stored in decentralised locations. The data are indexed with their own Internationalised Resource Identifiers (IRIs) that can be addressed without ambiguity.

Besides domain ontologies, JPS employs an agent ontology (OntoAgent [51]) to govern the concepts related to agents interacting within the knowledge-graph. Each agent is an individual building block, defined for specific tasks. As the agents share the same architecture and follow a similar design pattern, once activated, they are able to communicate with each other. The agents may further operate cooperatively in tasks ranging from manipulating the data within the knowledge-graph to coordinating between JPS and the outside world. To ensure secure agent operations, blockchain technology was implemented to support automated agent selection with a tamper-proof agent marketplace [52].

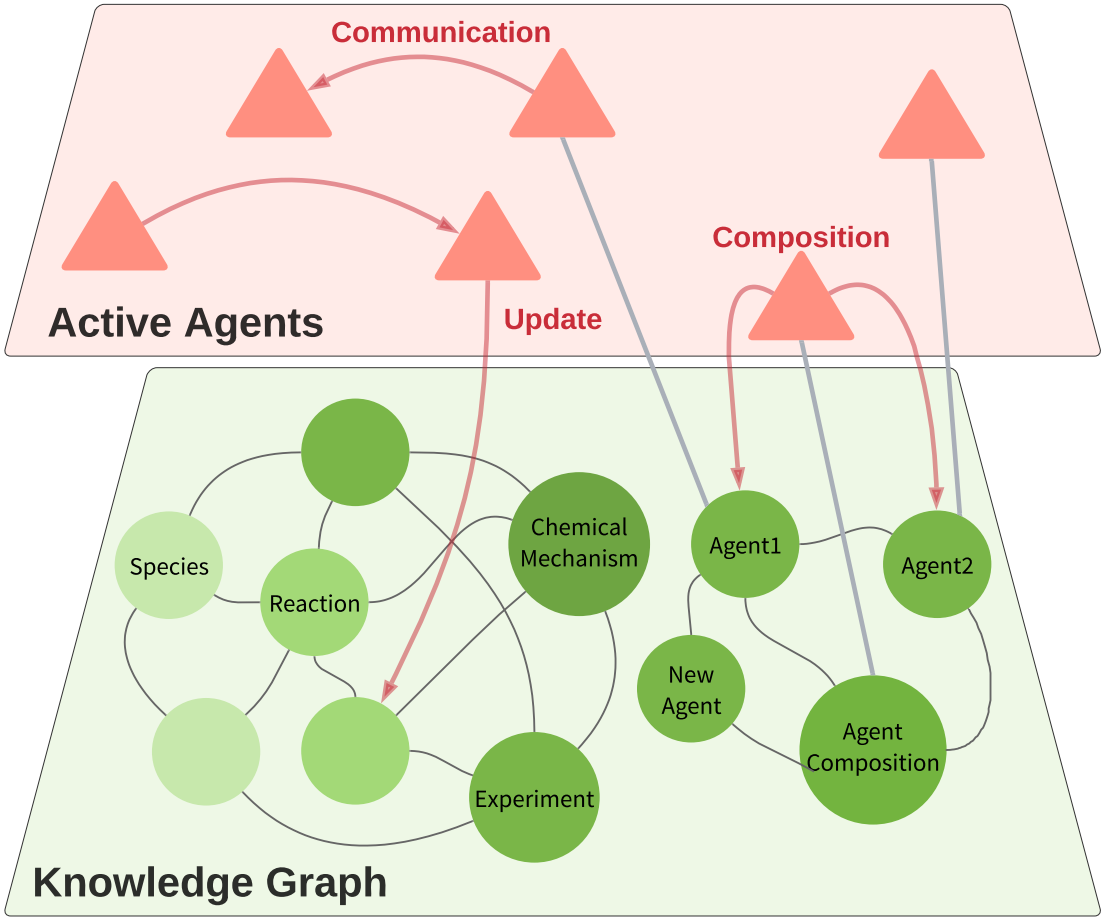


Figure 1: *Structure of the J-Park Simulator (JPS) as an implementation of a World Avatar. Inter-operable agents are part of the distributed knowledge-graph and operate on it once they are activated.*

Once granted access privileges, these tasks are envisaged to be completed without human interventions.

The current design and implementation of JPS has two significant advantages, namely reusability and extensibility. In that sense, JPS sets the standard and provides the basic toolbox, facilitating individual researchers and developers to build a customised knowledge-graph. As its ecosystem expands, the JPS has access to high volumes of real-time data from the real world, enabling simulation of dynamic physical processes in the cyber space. By connecting the real world and its virtual representation, JPS adds a new dimension of intelligent operation in the engineering sector.

The work described in this paper connects the measurements made during combustion experiments to relevant reaction mechanisms via the unique identification of chemical species. A distinguishing feature of the ontology and the agents we have developed is their seamless addition to an existing knowledge-graph in which all entities are relevantly linked. This presents a next step towards connecting macroscopic, observable data collected in real world experiments with quantum chemistry calculations. Relating experi-

mental data to reaction mechanisms, and subsequently to individual reactions, presents a valuable step towards enhancing mechanism development with accurate chemical kinetics. This step further pushes in the direction of combining data from a wide variety of domains, scales and sources to support cross-domain communication and scenario planning, the relevance of which is demonstrated in atmospheric pollutant dispersion simulation [10, 37].

3 Methodology

3.1 Mechanism calibration

3.1.1 PODE demonstration case and simulation procedure

This case serves as a demonstration for the developed framework. The particular case of $PODE_3$ has received recent interest and a range of previous efforts for modelling its combustion processes accurately when used as part of a fuel blend. Some of the prior works in this area are listed in **Table 1**, with many of the mechanisms intended for primary reference fuel (PRF) blends.

Table 1: *Summary of existing $PODE_n$ ($n = 2,3,4$) combustion mechanisms with their statistics counted in the OntoKin format. The reduced mechanism developed in Lin et al. [32], before optimisation, is chosen as the starting mechanism for the demonstration case of the knowledge-graph based automated mechanism calibration approach proposed in this work. It should be noted that the number of reactions of the OntoKin representation is different to that of CHEMKIN format.*

Mechanism	Type	No. Species	No. Reactions	Fuel	Fuel Carrier
Sun et al. [44]	Detailed	274	1674	$PODE_3$	No
He et al. [22]	Detailed	225	1178	$PODE_3$	No
He et al. [21]	Detailed	354	1392	$PODE_3$	Yes, PRF
Ren et al. [40]	Reduced	145	668	$PODE_3$	Yes, PRF
Lin et al. [32]	Reduced	61	215	$PODE_3$	Yes, PRF
Cai et al. [5]	Detailed	322	1611	$PODE_{2-4}$	No

The first mechanism for pure $PODE_3$ combustion under high-temperature conditions was developed by Sun et al. [44]. This was an example of a detailed combustion mechanism, whereby an attempt is made to model all elementary reactions believed to be present. In contrast, reduced combustion mechanisms are constructed to replicate results of key combustion metrics with a reduced set of reactions.

Following from the Sun et al. [44] mechanism, low- and intermediate-temperature conditions were covered by He et al. [22]. This model was further expanded by He et al. [21],

which is the first-ever mechanism to describe the combustion of a $PODE_3$/PRF blend. Given the complexity of engine simulations, the size of these mechanisms makes simulation largely intractable, requiring the development of a reduced mechanism. Two simplified mechanisms were proposed by Ren et al. [40] and Lin et al. [32]. Both employed the model of He et al. [22] as a basis, using different methodologies for selecting key species and reactions of $PODE_3$. Additional reactions were added by both for modelling the combustion of a PRF carrier fuel.

A notable alternative attempt was made in the work of Cai et al. [5]. In this, an automated mechanism development process is used to select the reactions for the detailed combustion mechanism of $PODE_n$ ($n = 2, 3, 4$). The work employed the class-based automatic reaction alternator (CLARA) and calibrated the selected reactions against experimental data for the ignition delay of $PODE_n$/air mixtures.

As a demonstration of the proposed knowledge-graph based approach, the starting mechanism of Lin et al. [32] is selected due to its relatively small size. This is the mechanism generated by selecting reactions from the wider He et al. [22] mechanism prior to any further calibration to experimental results. The focus of this paper remains the calibration of the $PODE_3$ combustion mechanism and so only $PODE_3$ combustion experiments are chosen for calibration.

The calibration was carried out against rapid compression machine (RCM) ignition delay time [22] and laminar flame speed [44] experiments. The ignition delay times of $PODE_3/O_2/N_2$ mixtures were measured at pressures of 10 bar and 15 bar, over a temperature range of 641 K to 865 K, with equivalence ratios of 0.5 (O_2:N_2 = 1 : 8), 1.0 (O_2:N_2 = 1 : 15), and 1.5 (O_2:N_2 = 1 : 20). The laminar flame speeds of $PODE_3$/air mixtures were measured at atmospheric pressure and an initial temperature of 408 K, with equivalence ratios ranging from 0.7 to 1.6. For the simulation stage, the ignition delay time is defined as the time interval between the starting point and the maximum rate of pressure rise due to the ignition. The laminar flame speeds were calculated with a mixture-averaged transport model. The simulations were performed using *k*inetics (version 2020.1.1) [7] for ignition delay times and Cantera (version 2.4.0) [20] for laminar flame speeds. For the laminar flame speed simulations, Soret effects were not considered and the solution gradient and curvature were both fixed at 0.05. The grid was set to be refined with a pruning coefficient of 0.01.

3.1.2 Sensitivity analysis

Sensitivity analysis acts as a screening process to identify reactions that have measurable effects on the model responses [15]. This is conducted by computing the normalised sensitivity coefficient of the chosen response with respect to the Arrhenius pre-exponential factors of the starting mechanism.

At the n^{th} point in the process condition space $\xi^{(n)}$, the normalised sensitivity coefficient of the i^{th} response $\eta_i(\xi^{(n)}, \theta)$ with respect to the j^{th} model parameter θ_j is defined as [45]:

$$\frac{\theta_j}{\eta_i(\xi^{(n)}, \theta)} \frac{\partial \eta_i(\xi^{(n)}, \theta)}{\partial \theta_j}. \tag{1}$$

Due to the complexity and stiffness of a typical combustion mechanism, a numerical solution is normally adopted [45]. The vector showing a small relative perturbation r of model parameters in the j^{th} positive direction can be denoted as:

$$\tilde{\theta}^j := (\theta_1, \ldots, \theta_{j-1}, (1+r) \times \theta_j, \theta_{j+1}, \ldots, \theta_n), \tag{2}$$

yielding a finite difference approximation of the normalised sensitivity coefficient as:

$$\frac{\theta_j}{\eta_i(\xi^{(n)}, \theta)} \frac{\eta_i(\xi^{(n)}, \tilde{\theta}^j) - \eta_i(\xi^{(n)}, \theta)}{(\tilde{\theta}^j - \theta)_j} = \frac{\eta_i(\xi^{(n)}, \tilde{\theta}^j) - \eta_i(\xi^{(n)}, \theta)}{r\eta_i(\xi^{(n)}, \theta)}. \tag{3}$$

Considering the sensitivities across the entire range of process condition space N, the overall sensitivity S_{ij} of the i^{th} response with respect to the j^{th} model parameter can be determined either by a maximum absolute value:

$$S_{ij} = \max_n \left\{ \left| \frac{\eta_i(\xi^{(n)}, \tilde{\theta}^j) - \eta_i(\xi^{(n)}, \theta)}{r\eta_i(\xi^{(n)}, \theta)} \right| \right\}, \tag{4}$$

or an averaged absolute value:

$$S_{ij} = \frac{1}{N} \sum_N \left[\left| \frac{\eta_i(\xi^{(n)}, \tilde{\theta}^j) - \eta_i(\xi^{(n)}, \theta)}{r\eta_i(\xi^{(n)}, \theta)} \right| \right]. \tag{5}$$

It should be noted that this analysis is *local* in the sense of model parameters while *global* in the process condition space, such that sensitivities at every collected point in the experiment are taken into account.

As chemical mechanisms can either be assembled from reaction classes or individual elementary reactions, it is natural to either optimise reactions on a class basis or just based on individual reaction's contributions. Cai and Pitsch [4] demonstrated a comparable performance between both bases, claiming that a significant distinction would only appear when reactions in the same class show low sensitivity individually but high sensitivity collectively.

In the case of a combustion mechanism constructed for a group of similar species, optimisation based on reaction rules often provides a consistent improvement of model performance. This was found to be the case with the mechanism developed in Cai et al. [5], describing PODE_{2-4} combustion. The comparable performance is seen as justification for implementing only one of the bases at present. The selected basis is that of individual reaction contributions, chosen because many of the envisaged use cases will involve only one or few key staring species. Further option for calibration on a class basis will be implemented in future work.

3.1.3 Global search and local optimisation

In order to find an optimal balance between the two considered responses, a weighted least-squares objective function was implemented for the experiment responses:

$$\begin{aligned}\Phi(\theta) = &\sum_{n=1}^{N}\left[[\eta_{\text{ign}}(\xi^{(n)},\theta)]' - [\eta_{\text{ign}}^{\text{exp}}(\xi^{(n)})]'\right]^2 \\ &+ \alpha \times \sum_{m=1}^{M}\left[[\eta_{\text{fls}}(\xi^{(m)},\theta)]' - [\eta_{\text{fls}}^{\text{exp}}(\xi^{(m)})]'\right]^2,\end{aligned} \tag{6}$$

where α refers to the weighting of laminar flame speed in the calibration process.

Following selection of the target reactions through sensitivity analysis, an optimisation routine is followed to calibrate the mechanism with the objective function defined above. The process initially employs low-discrepancy quasi-random global sampling through a Sobol sequence generator [43]. This provides initial points for a Hooke-Jeeves optimisation algorithm [23], selected for its gradient free nature to better handle the stiff system.

In each evaluation, experiment and model responses are scaled with respect to the upper η_{ub} and lower η_{lb} bounds of the experimental observations, as defined by the experimental unceratinty. For ignition delay times, a $\pm 20\%$ uncertainty was assigned to the measurements. This was selected to align with common practices within the community [3, 5, 24].

As ignition delay times can vary by orders of magnitude, a logarithmic scaling was applied to balance the contribution of each data point towards to the objective function:

$$\eta' = \frac{\log(\frac{\eta^2}{\eta_{\text{ub}}\eta_{\text{lb}}})}{\log(\frac{\eta_{\text{ub}}}{\eta_{\text{lb}}})}. \tag{7}$$

For laminar flame speed data, the error used was that reported by the source [44]. A linear scaling was applied:

$$\eta' = \frac{2\eta}{\eta_{\text{ub}} - \eta_{\text{lb}}}. \tag{8}$$

Uncertainty bounds may be obtained from uncertainty factors (UF) for Arrhenius rate equation parameters [41], derived from the uncertainties in quantum chemistry calculations. There are also alternative optimisation principles for the reactions involved in combustion chemistry that optimise both activation energies and pre-exponential rate parameters in a coupled manner [1] as well as techniques that include the temperature dependence exponent.

At present, the optimisation of just pre-exponential factors has been performed. This was chosen to simplify the process for a proof-of-concept framework demonstration and as the main focus of this work is to demonstrate a knowledge-graph based approach. There is existing precedent for works adjusting the pre-exponential factors alone [32], with large hypercubes of potential values. This approach is further justified when reduced mechanisms are optimised as some reaction pathways and species are already not present within the mechanism. It should be noted that other optimisation techniques can be easily made available in future work.

3.2 Ontological representation

The OntoChemExp ontology is developed to describe combustion *experiments*, detailing both the overall experimental setup and the individual, independent variable values for each data point. The overall experimental description of the ontology incorporates the *apparatus* used and the various *common properties* shared amongst datapoints. Independents are used to form *data groups* that share the same set of independent variables, with individual data points forming subsets of these data groups.

Take the *experiment* conducted by Sun et al. [44] as an example. A spark-ignited spherical flame was formed in the combustion vessel (*apparatus*). A mixture of $PODE_3$/air at atmospheric pressure and an initial temperature of 408 K was used in the experiment for all data points (*common properties*). The laminar flame speed measurements (*data groups*) were recorded at equivalence ratios ϕ ranging from 0.7 to 1.6.

The current structure of OntoChemExp is developed following discussions with domain experts and takes inspiration from existing databases, including the experimental data stored in the PrIMe database. The complete ontology contains 24 concepts and 46 relations. OntoChemExp is published at: `http://www.theworldavatar.com/ontology/ontochemexp/OntoChemExp.owl`.

Figure 2 illustrates the core concepts and relations defined in OntoChemExp. The conceptual structure is divided into four modules following a heuristic approach:

- **Experiment**: An *experiment* refers to the process of observing and measuring characteristics of interest from an energy release chemical process of a mixture of fuel and air. Dependent upon the original source, a set of metadata may be employed to provide more details and more precise identification of an experiment. This metadata includes *copyright*, *bibliography link* and *additional data item*.
- **Setup**: The setup outlines the global conditions of an experiment, including the *apparatus* in which the experiment was conducted, and the shared process conditions, forming *common properties*. The concepts defined in this section are normally left unchanged throughout an experiment.
- **Results**: Experimental *results* are grouped within *data groups* that share the same set of independent variables. Within the data groups, individual *data points* describe each experimental measurement, including independent and dependent variable values that are detailed within *X*.
- **Specification**: The specification is a shared data structure, supporting both the Setup and Results sections with an abstract concept *property*. Property is used to group a wide range of properties. The most straightforward usage is detailing the size of equipment with *value*. Property is also used to provide information about chemical *components* with the species described by a *species link* and an *amount*, *e.g.*, initial composition of fuel/air mixture. A further use of property is describing *derived properties*, that include *features* such as *indicators* and *observables*.

Figure 3 depicts how combustion experiment measurements and chemical mechanisms may be connected. The task of linking species with reactions has already been achieved

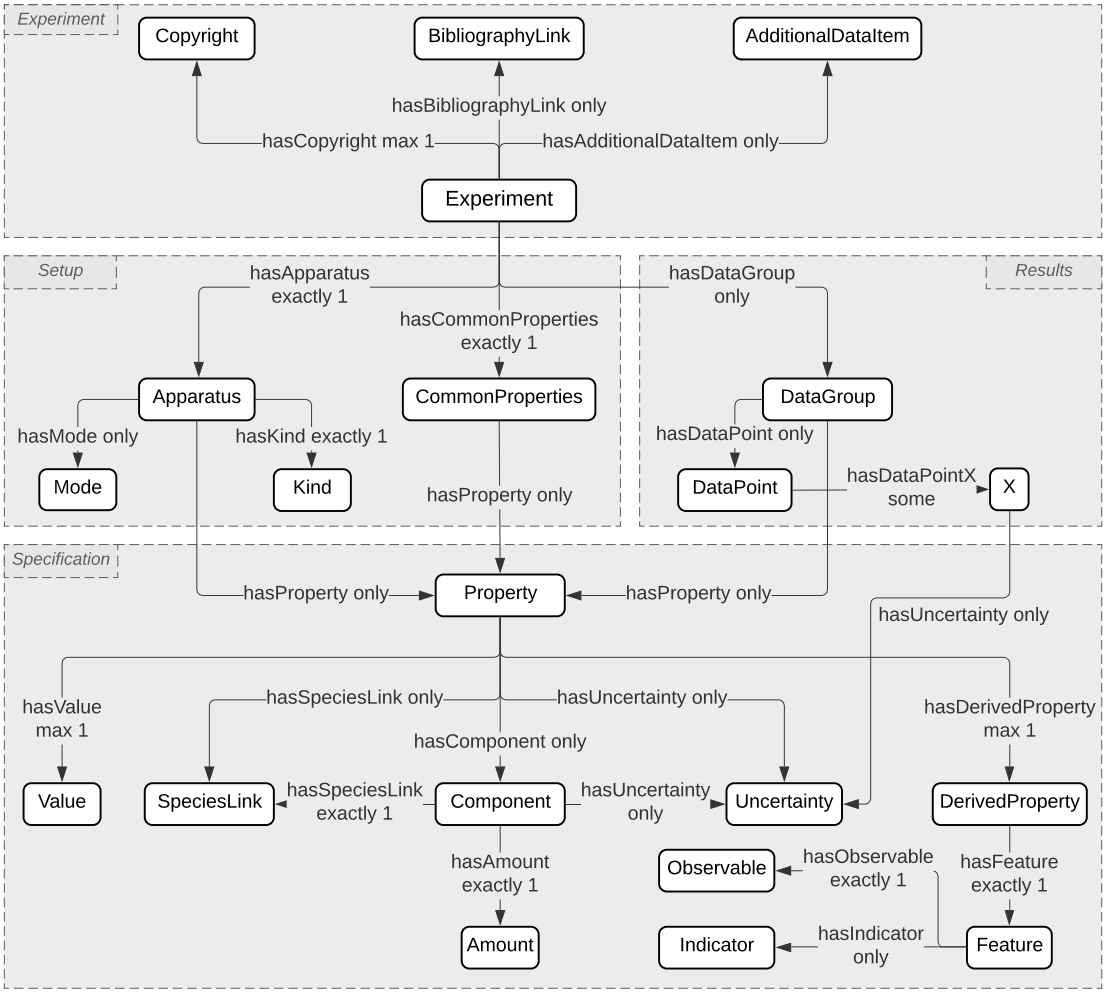


Figure 2: *Core concepts and relations of OntoChemExp ontology. This ontology is constructed to represent measurements from combustion experiments. The complete ontology consists of 24 concepts and 46 relations.*

in previous work [14], linking OntoKin with OntoSpecies. This allows the linking of OntoChemExp to both species and reactions via provision of unique species identifiers within OntoSpecies.

Two connections are made between OntoChemExp and the prior ontologies: (1) data property `hasPreferredKey`, equivalent to `skos:altLabel`, that refers to the common name of a species within the community and (2) data property `hasUniqueSpeciesIRI` that directly links to the exact OntoSpecies instance. This design can help resolve inconsistencies between data from different sources, through the unique identification of chemical species in OntoSpecies. The importance of the use of this approach is shown by the $PODE_3$ demonstration case whereby poly(oxymethylene) dimethyl ether 3 is denoted differently as $PODE_3$ by He et al. [22], $POMDME_3$ by Sun et al. [44], DMM_3 by Lin et al. [32], and OME_3 by Cai et al. [5].

These ambiguities may be handled by human operators but present a significant challenge

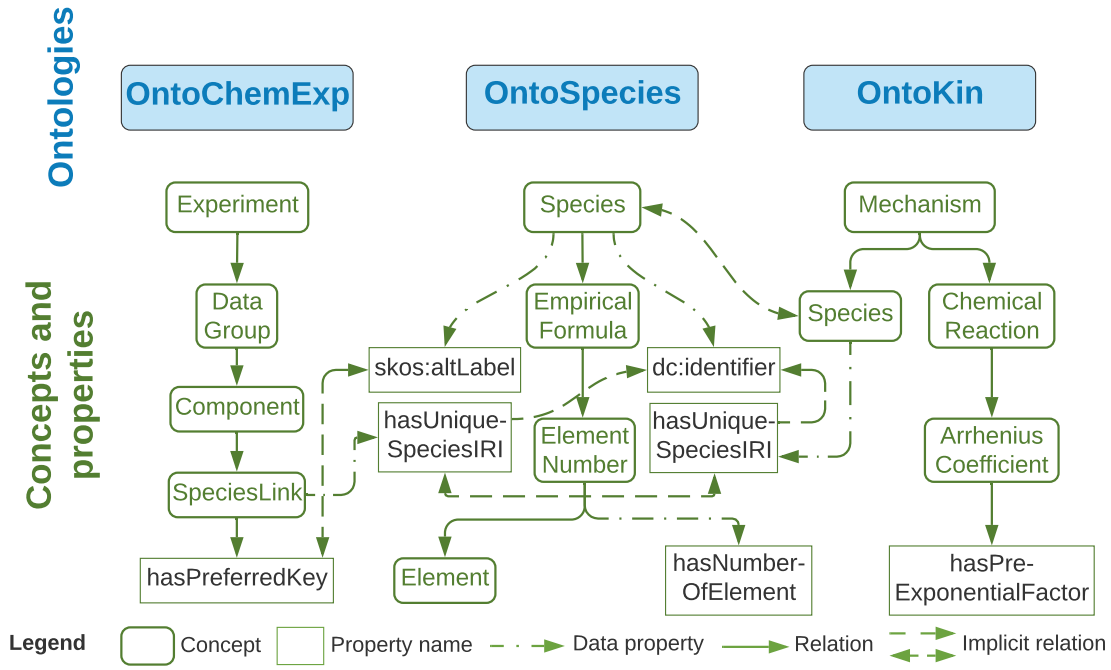


Figure 3: *Selected concepts, properties, and relations that demonstrate the links between OntoChemExp, OntoSpecies and OntoKin ontologies. The main purpose of these links is to enable unique identification of species.*

to machine interpretability. The ontological approach adopted allows for the same level of understanding by both humans and machines, allowing for greater interoperability between agents, more comprehensive querying, and many opportunities for semantically driven tasks.

Populating the knowledge-graph is managed by a tailor-made toolset developed in this work for generating OntoChemExp-conformant OWL files. The toolset is based on that developed by Farazi et al. [12] for converting chemical mechanisms to the format of OntoKin. Experimental data related to $PODE_3$ were manually constructed in the OWL format of OntoChemExp and then uploaded to the knowledge-graph. The knowledge-graph was subsequently deployed on an RDF4J (`https://rdf4j.org/`) triple store, queriable by the SPARQL Protocol and RDF Query Language (SPARQL).

3.3 Agent integration

The framework detailed and developed in this work is intended to act as an agent within the JPS ecosystem. It is structured as an instantiation of the agent template proposed by Mosbach et al. [37]. A Unified Modeling Language (UML) activity diagram of the agent is provided in **Fig. 4**, with the agent template surrounding the Model Development Suite (MoDS) [8] software package. MoDS is an integration of multiple tools developed for various generic model development tasks, such as parameter estimation [25], surrogate model creation [42, 47], and experimental design [35].

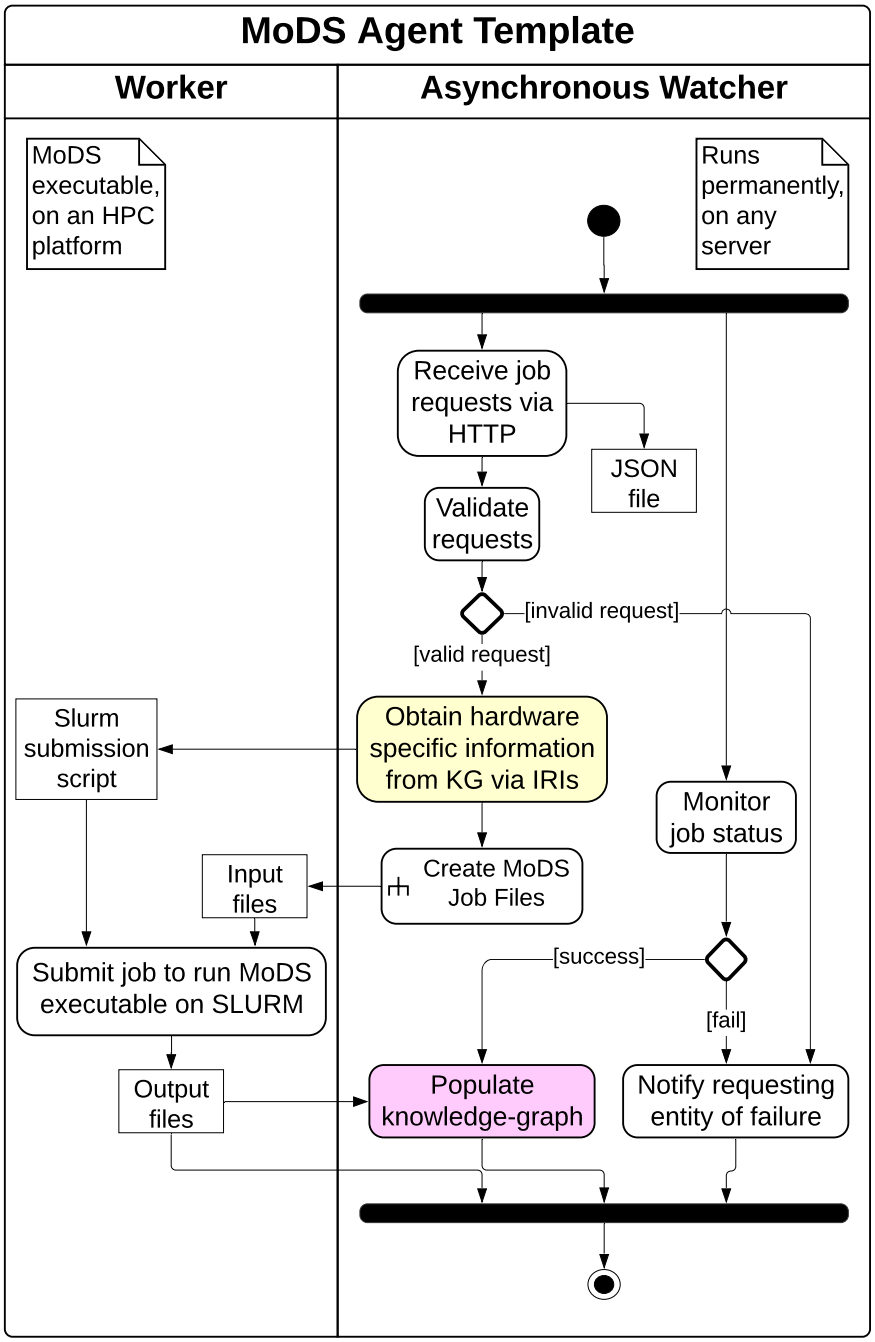


Figure 4: *UML activity diagram of a template agent that enables MoDS jobs to be executed asynchronously on an HPC platform upon HTTP requests. The same design is followed by all MoDS wrapper agents, distinguished by different activate node for jobs files creation. The yellow shaded action indicates the data retrieving operation of agents over knowledge-graph, whereas magenta refers to the knowledge-graph populating operation.*

Detailed documentation of the agent template is provided by Mosbach et al. [37] so only the changes from the template design will be detailed. One such change is the addition of a validation step for received job requests. This is added in order to improve the robustness of the agent. An additional change has been made to merge the process of querying executable-specific information from the knowledge-graph with the process of creating job files. This was implemented to accommodate different types of jobs being requested due to the integration of MoDS with multiple tools and its capability as a generic model development tool [25, 35, 36]. This results in an agent capable of automatically generating specific job files corresponding to supplied job requests.

The MoDS agent is designed to accept a target mechanism and experimental results at a range of process conditions and to perform parameter estimation for the target mechanism. To achieve this, the agent performs simulations with the experimental conditions and adjusts parameters within the target mechanism to replicate the experimental responses. The responses cover different simulation tasks which are performed in two different software packages, necessitating the generation of individual executable models. The software packages were *k*inetics [7] for ignition delay time simulation and Cantera [20] for laminar flame speed simulation.

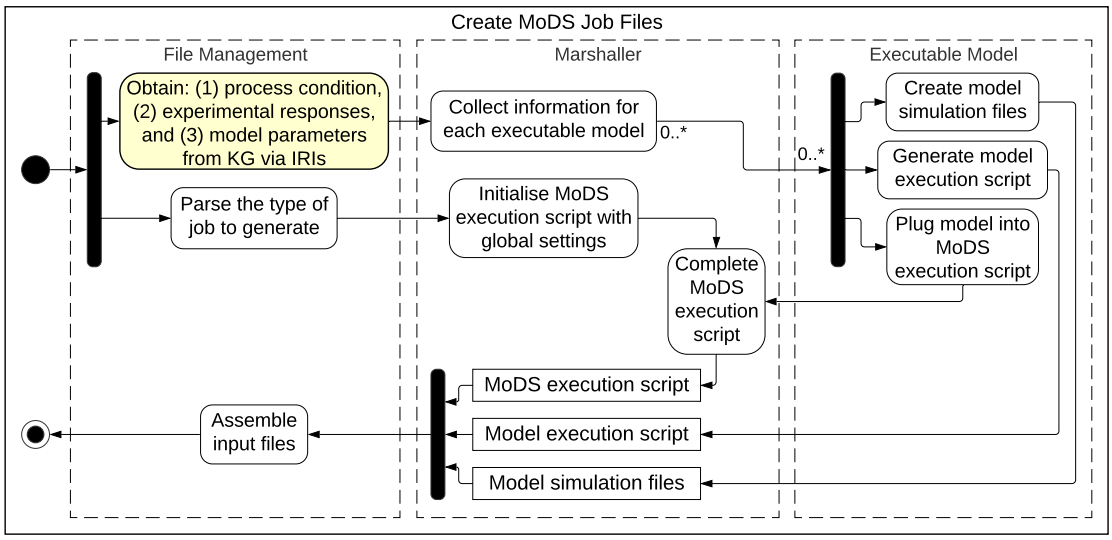


Figure 5: *Workflow of the process of creating requested job files. The whole process corresponds to the activity node in Fig. 4.*

Figure 5 illustrates the automated process of generating MoDS job files. This process is structured in three layers: the file management centre communicates the input and output, a marshaller collects all information required by the MoDS executable, and finally the layer that manages the individual executable models. In the file management centre, the process starts by querying the knowledge-graph for information related to each experiment response. This information is then passed to the marshaller to allocate executable models for simulating each response. The simulation files and execution script required for the selected models are then generated in the executable model layer and sent back to the marshaller.

At the same time, the type of job requested is parsed in the file management centre and passed to the marshaller to initialise a MoDS execution script with predefined global settings. This script is connected to the selected models generated by the executable model layer. All generated files are then assembled in the file management centre and transferred to an HPC platform.

The first two stages in the automated mechanism calibration are performed by two MoDS template agents: MoDSSensAnaAgent for sensitivity analysis and MoDSMechCalibAgent for mechanism calibration. Three parameters are currently available for the sensitivity analysis: magnitude of the relative perturbation, the type of overall sensitivity (maximum absolute value or average absolute value), and the number of reactions to be optimised.

For the mechanism calibration, the parameters are made available in two folds: the global settings for the algorithms used in the MoDS job and the calibration objective parameters. For the algorithms, the total number of Sobol points to be generated and the termination tolerance of the Hooke-Jeeves algorithm can be specified by the user. For the calibration objective parameters, two parameters are available: the weighting in the objective function and the response scaling type (logarithmic or linear). Provision is also made for users to supply their own list of reactions to guarantee their inclusion in the calibration process.

A coordination agent manages the interactions between the MoDSSensAnaAgent and MoDSMechCalibAgent with the knowledge-graph. These three agents form the overall automated mechanism calibration agent, AutoMechCalibAgent.

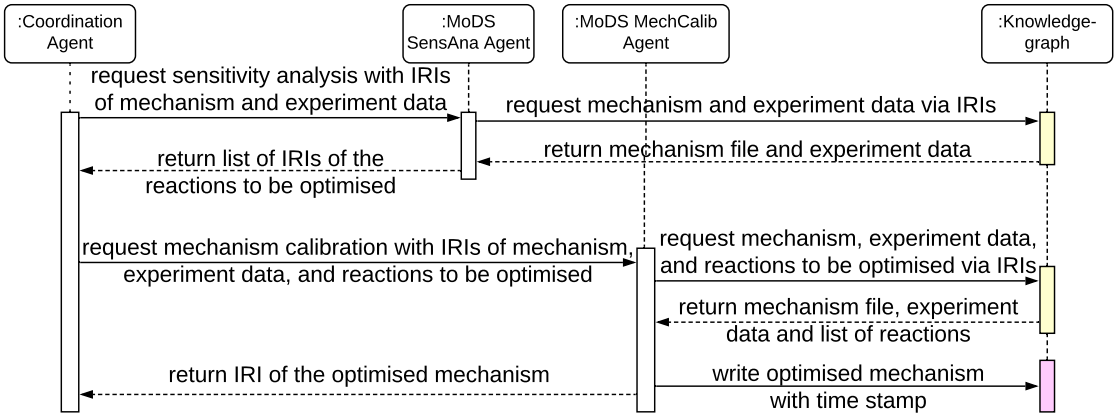


Figure 6: *UML sequence diagram of the automated mechanism calibration process that captures the interaction between the different agents and the knowledge-graph. Actions where the agent retrieves data from the knowledge-graph are shaded in yellow and those where the agent populates the knowledge-graph in magenta.*

Figure 6 illustrates the UML sequence diagram of the AutoMechCalibAgent as a five-step process. Initially, the coordination agent requests a sensitivity analysis via IRIs to identify the active parameters for calibration. Secondly, the MoDSSensAnaAgent communicates with the knowledge-graph via IRIs to obtain the chemical model to be calibrated and the process conditions over which the sensitivity analysis is to be conducted. MoDSSensAnaAgent returns the list of IRIs of the identified reactions to the coordination agent. Thirdly, the coordination agent requests a mechanism calibration job for the reactions

identified. Fourthly, the MoDSMechCalibAgent carries out the calibration. Global search and local optimisation are used with experimental data retrieved from the knowledge-graph. As the final step, the MoDSMechCalibAgent populates the knowledge-graph with the calibrated mechanism and returns its IRI to the coordination agent.

The process relied upon linked ontologies. These provided the connection between combustion experiments (OntoChemExp) and kinetic mechanisms (OntoKin) via unique identification of chemical species (OntoSpecies).

4 Results and Discussions

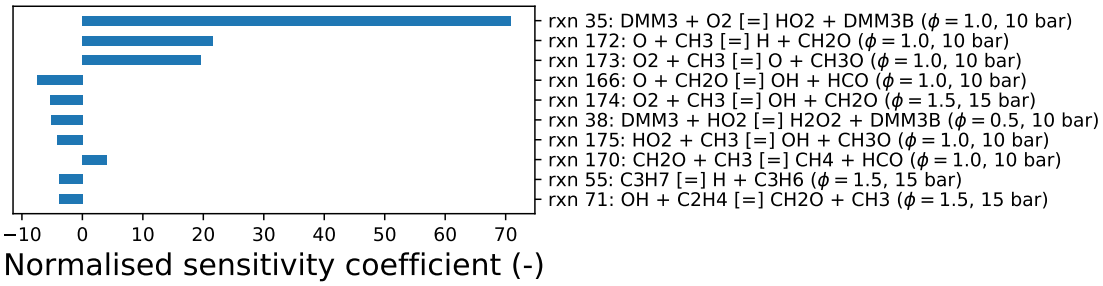
4.1 Sensitivity study

Prior to the automated calibration task, a sensitivity study was performed to assess the effect of the relative perturbation size used during the reaction selection sensitivity analysis in the automated calibration process. This sensitivity study was performed using the MoDSSensAnaAgent developed. The study involved 21 relative perturbation sizes for the finite difference approximations (r in **Equ. 3**) of the derivatives required for the sensitivity coefficients (1×10^{-n}, 2×10^{-n}, and 5×10^{-n}; $n=1,\ldots,7$) and assess the sensitivities for both ignition delay times and laminar flame speeds to the Arrhenius pre-exponential factor for all 215 reactions. For every reaction, a normalised sensitivity coefficient was computed for all 73 experimental conditions used for the calibration process. The maximum absolute value form of the sensitivity coefficient was used and the reactions ranked based on this value to determine their relative importance. We present the 10 reactions with the greatest absolute sensitivity coefficients.

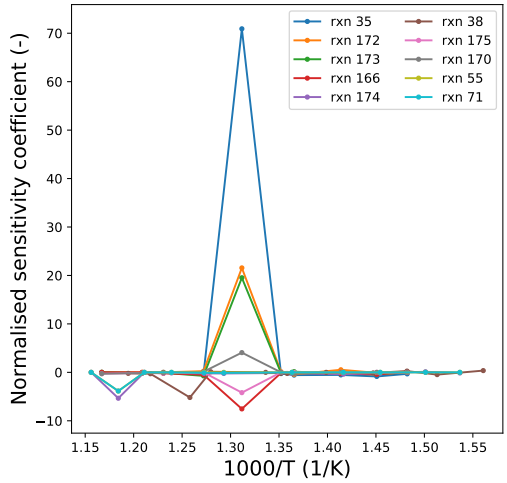
Figure 7 presents the ignition delay time results of the sensitivity study. The values of the sensitivity coefficients at their maximum absolute value are shown in Fig. 7(a) along with the conditions for each case. The effect of temperature on the sensitivities at fixed equivalence ratio and pressure is shown in Fig. 7(b). As demonstrated by the variability in normalised sensitivity coefficients to changing conditions, it is necessary to assess reaction importance through a *global* perspective. Figure 7(c) shows the stability of the sensitivity coefficients over the range of relative perturbation sizes investigated. The peaks and troughs in Fig. 7(b) correspond to the sensitivity coefficients used for ranking the reactions, matching the values at the dotted line in Fig. 7(c) and those presented in Fig. 7(a).

This study identified two reactions from the $PODE_3$ sub-mechanism, reactions 35 and 38. Reaction 35 belongs to the first O_2 addition reaction class. This class is suggested by Ren et al. [40] as a good choice for calibration as ignition delay times are usually sensitive to this class of reactions at low initial temperatures. Reaction 38 involves H-abstraction by HO_2 radicals, shown to increase the fuel reactivity thus important to $PODE_3$ combustion [5].

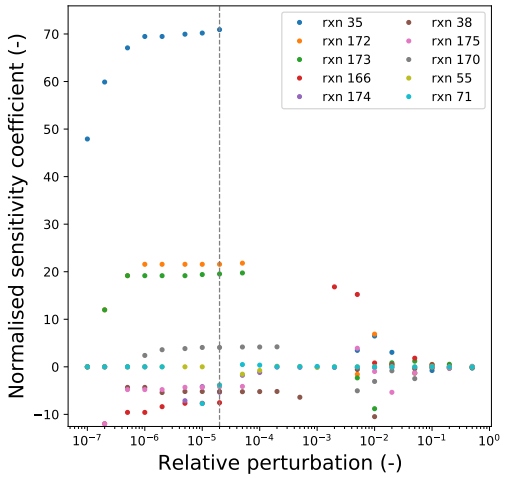
Figure 8 presents the sensitivity analysis results for laminar flame speeds. The structure remains similar to that of Fig. 7 but with Fig. 8(b) showing the effect of the equivalence ratio rather than temperature.



(a) Selected 10 most sensitive reactions using a relative perturbation of 2×10^{-5}.



(b) Sensitivities as function of temperature.



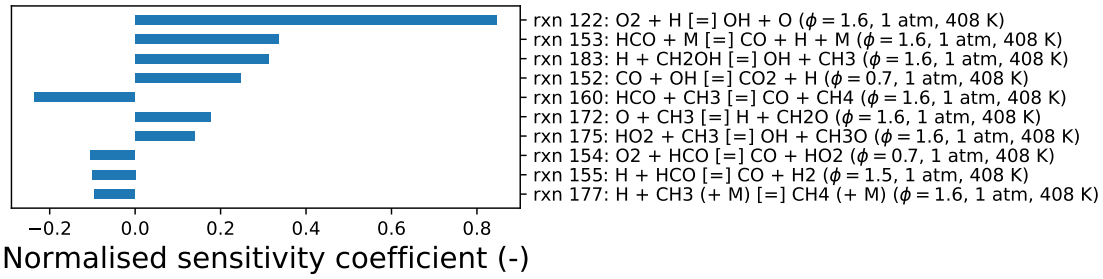
(c) Sensitivities as function of relative perturbation.

Figure 7: *Sensitivity analysis of the ignition delay times with respect to Arrhenius pre-exponential factors in the starting mechanism. The list of reactions is selected based on the maximum value of its sensitivities across all considered points in experimental condition space with a relative perturbation of* 2×10^{-5}.

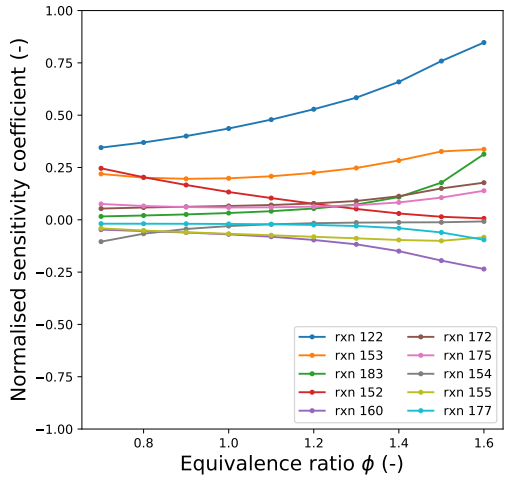
Reaction 122 was found to have the greatest influence on the laminar flame speed. This is a chain-branching step $O_2 + H \longleftrightarrow OH + O$ and is expected to have a significant effect on the laminar flame speed [46, chap. 8]. Other reactions identified mostly involve small species and radicals which governs a large part of the heat release (*e.g.* $CO + OH \longleftrightarrow CO_2 + H$).

The reactions identified for both laminar flame speed and ignition delay times fall in line with those selected while Lin et al. [32] constructing the reduced mechanism. The mechanism was developed using a decoupling methodology, separating the mechanism detail into three levels: detailed for H_2/CO/C_1, reduced for $C_2 - C_3$, and skeletal for $C_4 - C_N$. As appeared in the list of reactions found to be sensitive for laminar flame speed, most of them are in the detailed H_2/CO/C_1 sub-mechanism, representing high temperature combustion process. While for ignition delay times, most sensitive reaction is from a skeletal structure for $C_4 - C_N$, representing low temperature combustion process.

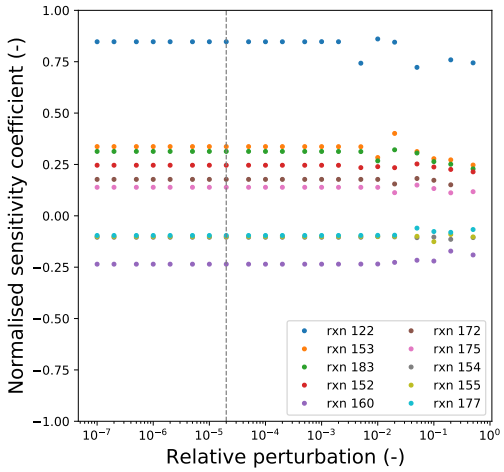
Following this analysis, a relative perturbation size of 2×10^{-5} was chosen for the cali-



(a) Selected 10 most sensitive reactions using a relative perturbation of 2×10^{-5}.



(b) Sensitivities as function of equivalence ratio.



(c) Sensitivities as function of relative perturbation.

Figure 8: *Sensitivity analysis of the laminar flame speed with respect to Arrhenius pre-exponential factors in the starting mechanism. The list of reactions is selected based on the maximum value of its sensitivities across all considered points in experimental condition space with a relative perturbation of* 2×10^{-5}.

bration process. This value is selected to act as a trade-off between numerical errors and the onset of non-linearities.

4.2 Mechanism calibration

In parameter estimation tasks, deciding the number of parameters to include within the estimation task requires consideration of the trade-offs between precision and tractability. For this calibration case, the reactions to optimise have been set as the top 10 reactions identified for ignition delay times and laminar flame speeds by the MoDSMechCalibAgent. This resulted in a total of 18 reactions to optimise due to reactions appearing in both sensitivity analyses.

Another trade-off requiring consideration is that of the weighting between ignition delay times and laminar flame speeds. In many cases, differing quantities of experimental data are available and there may exist differences in users' preference in weightings. The

weighting of the two is handled by the value of α in the objective function of the calibration process (*i.e.* **Eqn. 6**). In this case, there are 63 ignition delay time experimental data points and 10 laminar flame speed. Correcting this imbalance in number of experimental results forms a natural starting point for selecting a value for α, and so values of α in the range of 6.3 to 1 were investigated. This range is intended to cover values that offer a good balance between the two responses and prevent the domination of ignition delay times for the calibration.

The calibration routine seeks to optimise the values of the pre-exponential factors for the target reactions. The range selected for the task was 10^{-2} to 10^{2} times the original value. The process began with Sobol sampling within the selected range of values. 10^{4} logarithmic-evenly distributed points were used to determine three starting points for the optimisation routine that displayed the lowest values of the objective function.

Following on from the sampling stage, a Hooke-Jeeves optimisation routine is performed. The routine was performed with 400 iterations and a termination stepsize of 0.001, with an initial stepsize of 0.2 and stepsize reduction factor of 0.5. The results of the sampling and optimisation stages are presented in **Table 2**.

Table 2: *Objective function of global search and local optimisation results of the starting mechanism. The best three Sobol points from global search were chosen for further local optimisation with Hooke-Jeeves (H-J) algorithm. Value in boldface indicates the best performing mechanism for each response ratio and was chosen as the starting mechanism for the next iteration of calibration.*

Ratio α	Best Sobol	H-J	2^{nd} Sobol	H-J	3^{rd} Sobol	H-J
6.3	4446	**659**	5153	1375	5356	1263
4.98	4170	**657**	4697	1378	4788	895
3.65	3892	**710**	4216	712	4238	1398
2.33	3617	896	3648	**714**	3783	1373
1	3075	654	3165	**653**	3324	1370

Prior to the sampling stage, the scaled sum-of-squares-errors value was found to be 14554 for ignition delay times and 774 for laminar flame speeds, resulting in objective function values of 19430 and 15328 for α values of 6.3 and 1, respectively. This indicates the value of performing the initial Sobol sampling stage, with a significant improvement in the objective function being achieved prior to any optimisation. This is particularly valuable as the Hooke-Jeeves algorithm performs local search, significantly benefiting from a good initial point.

The optimisation stage is further seen to be providing an improvement in the objective function, significantly reducing its value from the sampling stage. The results of the optimisation stage are comparable with those of Lin et al. [32], which achieved an objective function value of 445 with an α value of 1.

Although variation is seen in the objective function values after the sampling stage in response to changing α values, this same change is not observed after the optimisation

stage. This is a result of the contributions to the objective function from the laminar flame speeds becoming very small after optimisation. The laminar flame speed is largely governed by small molecule oxidation which remains a detailed sub-mechanism within the decoupling methodology adopted in the Lin et al. [32] mechanism. This suggests a better fit would be expected for both the initial mechanism and calibrated mechanism for laminar flame speeds than ignition delay times.

During the initial sensitivity analysis, the mechanism is unable to accurately reproduce the combustion characteristics. This suggests that reaction selection at this stage may be premature and may not select all reactions that are of the most importance local to the optimum fitting. Additional reactions may also only become important after the rates of the initially identified reactions are closer to their optimum values. For these reasons, a further iteration of the calibration algorithm is necessary, which is consistent with the recommendation by Frenklach [15].

4.3 Second iteration

A second calibration was performed on the best performing mechanism for each value of α. The ontological structure of the framework aided in this process, allowing for the task to be completed by the passing of the IRIs for the experiments and calibrated mechanisms from the last iteration to the AutoMechCalibAgent agent. The agent again performed a sensitivity analysis to identify key reactions and a mechanism calibration to further optimise the mechanism. The settings for the sensitivity analysis as well as mechanism calibration were left unchanged. The results after both the sampling and calibration stages are summarised in **Table 3**.

Table 3: *Objective function values after sampling and optimisation for the best performing mechanisms selected from the first iteration of mechanism calibration for each α value. All mechanisms showed significant improvement in this iteration of calibration, with the best performing mechanism highlighted in green.*

Ratio α	Best Sobol	H-J	2^{nd} Sobol	H-J	3^{rd} Sobol	H-J
6.3	605	278	627	296	651	**260**
4.98	375	89	505	**79**	507	187
3.65	459	**243**	500	259	552	253
2.33	571	**38**	658	133	702	127
1	355	151	376	**133**	377	156

After the calibration stage, the best performing mechanism was found with an α value of 2.33. The objective value is found to show a 92% decrease compared to that for the mechanism of Lin et al. [32] when the same α values are used in both.

The performance of the mechanism of Lin et al. [32] (manual calibration) and the mechanism of this work (automated calibration) is compared in **Fig. 9** and **Fig. 10**. The automatically calibrated mechanism shows a good fit to the experimental data in all cases.

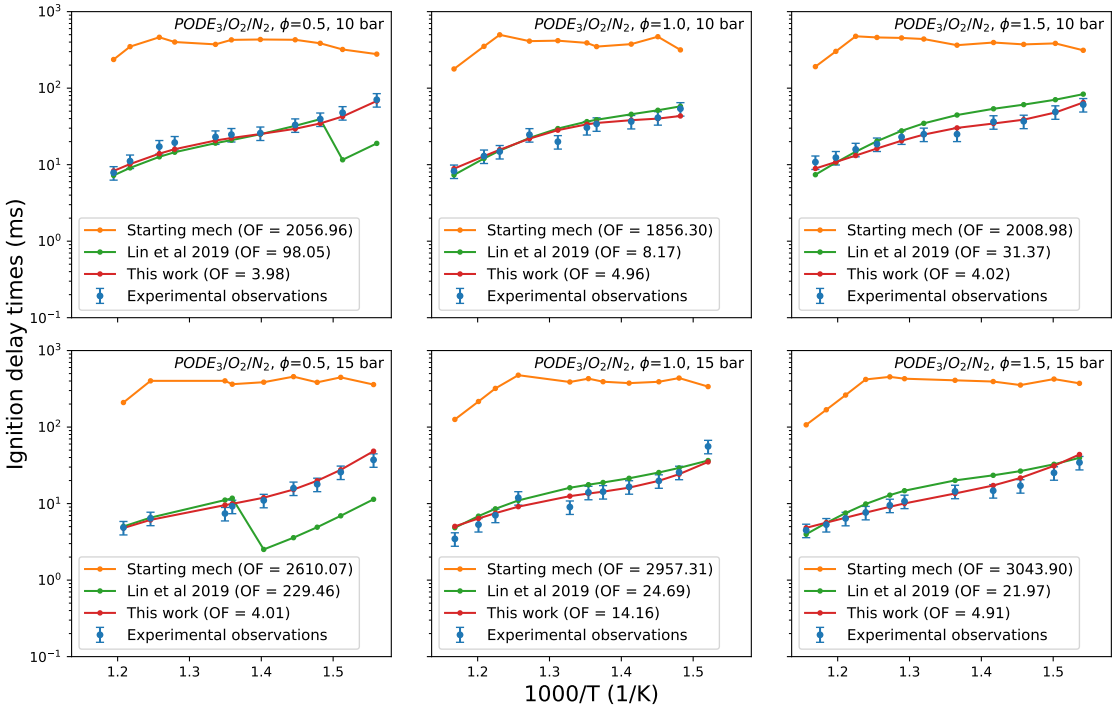


Figure 9: *Comparison of the mechanisms from [32] and the AutoMechCalibAgent agent (this work) at simulating ignition delay times (maximum rate of pressure increase ignition criterion) of $PODE_3/O_2/N_2$ mixtures at three equivalence ratios [22]. The model performance is displayed as the ignition delay time contribution to the objective function. As per the experimental results, the oxidizer used in this study has different compositions: (1) $\phi = 0.5$, O_2:$N_2 = 1:8$; (2) $\phi = 1.0$, O_2:$N_2 = 1:15$; (3) $\phi = 1.5$, O_2:$N_2 = 1:20$.*

The mechanism of Lin et al. [32] is found to perform worse than was presented in the original paper but this is a result of the change in ignition criteria between this work and that used in the original paper. In the original paper by Lin et al. [32], a temperature rise criterion was used for assessing ignition delay times while in this work a rate of pressure increase criterion is used, bringing the ignition criterion in line with what was used for the experimental results.

The negative temperature coefficient (NTC) behaviour of the fuel is captured in both the mechanism of this work and that of Lin et al. [32]. In Lin et al. [32], it is claimed that capturing the NTC region is achieved through optimisation of the isomerisation reaction $DMM_3BO_2 \longleftrightarrow DMM_3OOH_{35}$. In this work, this reaction was not identified as important and so was not calibrated, instead remaining at the same value as used in He et al. [22] where no apparent NTC region was captured.

It is believed the capturing of the NTC behaviour in this work is a result of the sensitivity analysis identifying reactions of importance in the intermediate-temperature regime (around 770 K), corresponding to the NTC region. This effect may be seen in Fig. 7(b) where the majority of the sensitivities show a peak in the intermediate-temperature region.

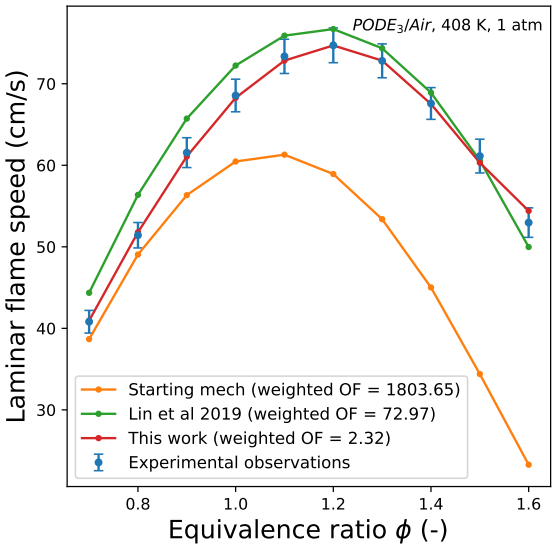


Figure 10: *Comparison between the model from [32] and the AutoMechCalibAgent agent (this work) on simulating laminar flame speed of $PODE_3$/Air mixtures at atmospheric pressure and an initial temperature of 408 K [44]. Model performance is displayed as the value of the laminar flame speed contribution to the objective function with an α value of 2.33.*

Table 4 summarises the changes made to the Arrhenius pre-exponential factors during the calibration process. The range of adjustment for the rate parameters during the calibration process was 10^{-4} to 10^{4}. Similar orders of adjustment were employed by both Lin et al. [32] and Chang et al. [6] while calibrating mechanisms constructed with decoupling methodologies. It is further noted that even more complete $PODE_3$ mechanisms, such as that of Ren et al. [40], modify the pre-exponential factors by an order of magnitude during calibration. This is to balance necessary levels of adjustment against unnecessarily large search spaces.

Two additional H-abstraction reactions from the $PODE_3$ sub-mechanism and a total of eight reactions were identified by the second calibration iteration that were not identified by the first. The second iteration fully captures the governing reactions of the low temperature combustion process, as found to be important to modelling ignition delay times [32]. Thus, the substantial improvement of model performance found in the second iteration is not surprising. The calibrated mechanism is available in CHEMKIN format (doi:10.17863/CAM.59826).

5 Conclusions

In this work, a knowledge-graph based automatic mechanism calibration framework is developed. This acts as an extension to the world avatar (theworldavatar.com), a dynamic knowledge-graph eco-system. All components developed in this work are standardised and modularised to allow them to easily integrate with the wider knowledge-graph infras-

Table 4: *Summary of the calibrated Arrhenius pre-exponential factors. Omitted values imply a reaction rate is unchanged. The unit of the pre-exponential factors is $m^3mol^{-1}s^{-1}$ or s^{-1} for two and one reactant respectively. The indexes of reactions follow the labels generated while converting mechanism from CHEMKIN to OntoKin format. Reactions identified as sensitive for different response are denoted as † for ignition delay time and ‡ for laminar flame speed. Note that $PODE_3$ is denoted as DMM_3.*

Reaction	Equation	Original A factor	1^{st} iteration	2^{nd} iteration
35	$DMM_3 + O_2 \longleftrightarrow HO_2 + DMM_3B$	6.66×10^{6}	6.66×10^{8}†	1.37×10^{10}†
36	$DMM_3 + OH \longleftrightarrow H_2O + DMM_3B$	3.79×10^{-2}	-	4.22×10^{-4}†
37	$DMM_3 + H \longrightarrow H_2 + DMM_3B$	7.40×10^{6}	-	5.05×10^{5}†
38	$DMM_3 + HO_2 \longleftrightarrow H_2O_2 + DMM_3B$	4.00×10^{7}	2.25×10^{9}†	1.24×10^{9}†
55	$C_3H_7 \longleftrightarrow H + C_3H_6$	1.25×10^{14}	1.08×10^{13}†	-
71	$OH + C_2H_4 \longleftrightarrow CH_2O + CH_3$	1.00×10^{8}	3.23×10^{6}†	-
122	$O_2 + H \longleftrightarrow OH + O$	1.04×10^{8}	1.04×10^{10}‡	1.72×10^{11}‡
124	$H_2 + O \longleftrightarrow OH + H$	8.79×10^{8}	-	1.35×10^{8}‡
126	$OH \longleftrightarrow H_2O + O$	3.34×10^{-2}	-	4.19×10^{-3}‡
152	$CO + OH \longleftrightarrow CO_2 + H$	2.23×10^{-1}	6.22×10^{-1}‡	1.16×10^{1}‡
153	$HCO + M \longleftrightarrow CO + H + M$	5.75×10^{5}	2.13×10^{5}‡	8.30×10^{5}‡
154	$O_2 + HCO \longleftrightarrow CO + HO_2$	7.58×10^{6}	2.02×10^{6}‡	-
155	$H + HCO \longleftrightarrow CO + H_2$	7.23×10^{7}	4.63×10^{9}‡	1.04×10^{8}‡
160	$HCO + CH_3 \longleftrightarrow CO + CH_4$	1.20×10^{8}	9.40×10^{9}‡	3.80×10^{10} †‡
166	$O + CH_2O \longleftrightarrow OH + HCO$	1.81×10^{7}	1.21×10^{7}†	-
167	$OH + CH_2O \longleftrightarrow H_2O + HCO$	3.43×10^{3}	-	6.79×10^{3}†
170	$CH_2O + CH_3 \longleftrightarrow CH_4 + HCO$	3.64×10^{-12}	5.78×10^{-11}†	5.38×10^{-9}†
172	$O + CH_3 \longleftrightarrow H + CH_2O$	8.43×10^{7}	8.43×10^{5} †‡	-
173	$O_2 + CH_3 \longleftrightarrow O + CH_3O$	1.99×10^{12}	1.99×10^{14}†	7.25×10^{14}†
174	$O_2 + CH_3 \longleftrightarrow OH + CH_2O$	3.74×10^{5}	6.56×10^{6}†	9.33×10^{4}†
175	$HO_2 + CH_3 \longleftrightarrow OH + CH_3O$	1.00×10^{6}	1.00×10^{8} †‡	-
177	$H + CH_3 (+ M) \longleftrightarrow CH_4 (+ M)$	1.27×10^{10}	1.03×10^{9}‡	-
180	$OH + CH_4 \longleftrightarrow H_2O + CH_3$	5.72×10^{0}	-	5.03×10^{2}‡
183	$H + CH_2OH \longleftrightarrow OH + CH_3$	9.64×10^{7}	3.30×10^{9}‡	3.78×10^{9}‡
186	$O_2 + CH_2OH \longleftrightarrow HO_2 + CH_2O$	2.41×10^{8}	-	2.41×10^{10}‡
193	$CH_3O + M \longleftrightarrow H + CH_2O + M$	8.30×10^{11}	-	2.26×10^{12}†

tructure.

For the development process, an ontology, OntoChemExp, was created. OntoChemExp provides an ontological description of combustion experiments and allows for linking these to existing ontologies for reaction mechanisms and chemical species, semantically

enriching the description of experiments and drawing links between mechanisms for combustion processes and their experimental validation.

Another contribution of this work is a set of agents for coupled sensitivity analysis and mechanism calibration. These are based on a standardised JPS agent template and are designed to employ generic model development software.

As a demonstration of these technologies, a case study was used of a reduced $PODE_3$ combustion mechanism. It was found that two iterations of the coupled agent process were required to sufficiently optimise this mechanism due to its initially poor fitting. The initial iteration brought the calibration objective to a value of a similar order to that of the manual calibration of Lin et al. [32], while the second iteration reduced its value to 8% of the manual calibration value. This represents the development of a significantly improved mechanism in a shorter time-span than manual calibration. Should multiple iterations be performed, users need simply provide the IRIs of experimental data and the mechanism from the previous iteration.

Research data

Research data supporting this publication is available in the University of Cambridge data repository (doi:10.17863/CAM.59826).

Acknowledgements

This research was supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme. The authors are grateful to the UK Engineering and Physical Sciences Research Council (EPSRC, grant number: EP/R029369/1) and ARCHER for financial and computational support as a part of their funding to the UK Consortium on Turbulent Reacting Flows (www.ukctrf.com). The authors would like to express their gratitude to Qinjie Lin and Prof Wenming Yang from the Green and Sustainable Transportation and Power Generation Laboratory of National University of Singapore, for their correspondence and sharing of the PODE3 experiment file. J. Bai acknowledges financial support provided by CSC Cambridge International Scholarship from Cambridge Trust and China Scholarship Council. R. Geeson acknowledges financial support EPSRC (grant EP/S024220/1) for SynTech Centre for Doctoral Training, University of Cambridge. E. J. Bringley was funded by a Gates Cambridge Scholarship (OPP1144). M. Kraft gratefully acknowledges the support of the Alexander von Humboldt foundation.

References

[1] P. Azadi, G. Brownbridge, I. Kemp, S. Mosbach, J. S. Dennis, and M. Kraft. Microkinetic modeling of the Fischer-Tropsch synthesis over cobalt catalysts. *ChemCatChem*, 7(1):137–143, 2015. doi:10.1002/cctc.201402662.

[2] P. L. Bhoorasingh, B. L. Slakman, F. Seyedzadeh Khanshan, J. Y. Cain, and R. H. West. Automated transition state theory calculations for high-throughput kinetics. *The Journal of Physical Chemistry A*, 121(37):6896–6904, 2017. doi:10.1021/acs.jpca.7b07361.

[3] S. M. Burke, U. Burke, R. Mc Donagh, O. Mathieu, I. Osorio, C. Keesee, A. Morones, E. L. Petersen, W. Wang, T. A. DeVerter, M. A. Oehlschlaeger, B. Rhodes, R. K. Hanson, D. F. Davidson, B. W. Weber, C.-J. Sung, J. Santner, Y. Ju, F. M. Haas, F. L. Dryer, E. N. Volkov, E. J. K. Nilsson, A. A. Konnov, M. Alrefae, F. Khaled, A. Farooq, P. Dirrenberger, P.-A. Glaude, F. Battin-Leclerc, and H. J. Currana. An experimental and modeling study of propene oxidation. Part 2: Ignition delay time and flame speed measurements. *Combustion and Flame*, 162(2):296–314, 2015. doi:10.1016/j.combustflame.2014.07.032.

[4] L. Cai and H. Pitsch. Mechanism optimization based on reaction rate rules. *Combustion and Flame*, 161(2):405–415, 2014. doi:10.1016/j.combustflame.2013.08.024.

[5] L. Cai, S. Jacobs, R. Langer, F. vom Lehn, K. A. Heufer, and H. Pitsch. Auto-ignition of oxymethylene ethers (OMEn, n= 2–4) as promising synthetic e-fuels from renewable electricity: Shock tube experiments and automatic mechanism generation. *Fuel*, 264:116711, 2020. doi:10.1016/j.fuel.2019.116711.

[6] Y. Chang, M. Jia, Y. Liu, Y. Li, and M. Xie. Development of a new skeletal mechanism for n-decane oxidation under engine-relevant conditions based on a decoupling methodology. *Combustion and Flame*, 160(8):1315–1332, 2013. doi:10.1016/j.combustflame.2013.02.017.

[7] CMCL Innovations. *k*inetics & SRM engine suite (version 2020.1.1), 2020. URL `https://cmclinnovations.com/solutions/products/kinetics/`.

[8] CMCL Innovations. MoDS: Model Development Suite (version 2020.2.2), 2020. URL `https://cmclinnovations.com/solutions/products/mods/`.

[9] A. Devanand, G. Karmakar, N. Krdzavac, R. Rigo-Mariani, Y. S. E. Foo, I. A. Karimi, and M. Kraft. OntoPowSys: A power system ontology for cross domain interactions in an eco industrial park. *Energy and AI*, 1:100008, 2020. doi:10.1016/j.egyai.2020.100008.

[10] A. Eibeck, M. Q. Lim, and M. Kraft. J-Park Simulator: An ontology-based platform for cross-domain scenarios in process industry. *Computers & Chemical Engineering*, 131:106586, 2019. doi:10.1016/j.compchemeng.2019.106586.

[11] A. Eibeck, A. Chadzynski, M. Q. Lim, K. Aditya, L. Ong, A. Devanand, G. Karmakar, S. Mosbach, R. Lau, I. A. Karimi, E. Y. S. Foo, and M. Kraft. A parallel world framework for scenario analysis in knowledge graphs. *Data-Centric Engineering*, 1:e6, 2020. doi:10.1017/dce.2020.6.

[12] F. Farazi, J. Akroyd, S. Mosbach, P. Buerger, D. Nurkowski, M. Salamanca, and M. Kraft. OntoKin: An ontology for chemical kinetic reaction mechanisms. *Journal of Chemical Information and Modeling*, 60(1):108–120, 2020. doi:10.1021/acs.jcim.9b00960.

[13] F. Farazi, N. B. Krdzavac, J. Akroyd, S. Mosbach, A. Menon, D. Nurkowski, and M. Kraft. Linking reaction mechanisms and quantum chemistry: An ontological approach. *Computers & Chemical Engineering*, 137:106813, 2020. doi:10.1016/j.compchemeng.2020.106813.

[14] F. Farazi, M. Salamanca, S. Mosbach, J. Akroyd, A. Eibeck, L. K. Aditya, A. Chadzynski, K. Pan, X. Zhou, S. Zhang, et al. Knowledge graph approach to combustion chemistry and interoperability. *ACS omega*, 5(29):18342–18348, 2020. doi:10.1021/acsomega.0c02055.

[15] M. Frenklach. Modeling. In W. C. Gardiner, editor, *Combustion Chemistry*, chapter 7, pages 423–453. Springer Verlag, New York, 1984.

[16] M. Frenklach. Transforming data into knowledge — Process Informatics for combustion chemistry. *Proceedings of the Combustion Institute*, 31(1):125–140, 2007. doi:10.1016/j.proci.2006.08.121.

[17] M. Frenklach, H. Wang, and M. J. Rabinowitz. Optimization and analysis of large chemical kinetic mechanisms using the solution mapping method — combustion of methane. *Progress in Energy and Combustion Science*, 18(1):47–73, 1992. doi:10.1016/0360-1285(92)90032-V.

[18] C. W. Gao, J. W. Allen, W. H. Green, and R. H. West. Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms. *Computer Physics Communications*, 203:212–225, 2016. doi:10.1016/j.cpc.2016.02.013.

[19] A. Giusti and E. Mastorakos. Turbulent combustion modelling and experiments: Recent trends and developments. *Flow, Turbulence and Combustion*, 103(4):847–869, 2019. doi:10.1007/s10494-019-00072-6.

[20] D. G. Goodwin, R. L. Speth, H. K. Moffat, and B. W. Weber. Cantera: An object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes. `https://www.cantera.org`, 2018. Version 2.4.0.

[21] T. He, H. Liu, Y. Wang, B. Wang, H. Liu, and Z. Wang. Development of surrogate model for oxygenated wide-distillation fuel with polyoxymethylene dimethyl ether. *SAE International Journal of Fuels and Lubricants*, 10(3):803–814, 2017. doi:10.4271/2017-01-2336.

[22] T. He, Z. Wang, X. You, H. Liu, Y. Wang, X. Li, and X. He. A chemical kinetic mechanism for the low-and intermediate-temperature combustion of polyoxymethylene dimethyl ether 3 (PODE3). *Fuel*, 212:223–235, 2018. doi:10.1016/j.fuel.2017.09.080.

[23] R. Hooke and T. A. Jeeves. "Direct Search" solution of numerical and statistical problems. *Journal of the ACM (JACM)*, 8(2):212–229, 1961. doi:10.1145/321062.321069.

[24] S. Jacobs, M. Döntgen, A. B. Alquaity, W. A. Kopp, L. C. Kröger, U. Burke, H. Pitsch, K. Leonhard, H. J. Curran, and K. A. Heufer. Detailed kinetic modeling of dimethoxymethane. Part II: Experimental and theoretical study of the kinetics and reaction mechanism. *Combustion and Flame*, 205:522–533, 2019. doi:10.1016/j.combustflame.2018.12.026.

[25] C. A. Kastner, A. Braumann, P. L. Man, S. Mosbach, G. P. Brownbridge, J. Akroyd, M. Kraft, and C. Himawan. Bayesian parameter estimation for a jet-milling model using Metropolis-Hastings and Wang-Landau sampling. *Chemical Engineering Science*, 89:244–257, 2013. doi:10.1016/j.ces.2012.11.027.

[26] R. J. Kee, F. M. Rupley, E. Meeks, and J. A. Miller. CHEMKIN-III: A FORTRAN chemical kinetics package for the analysis of gas-phase chemical and plasma kinetics. Technical report, Sandia National Labs., Livermore, CA (United States), 1996.

[27] K. Kohse-Höinghaus, M. Reimann, and J. Guzy. Clean combustion: Chemistry and diagnostics for a systems approach in transportation and energy conversion. *Progress in Energy and Combustion Science*, 65(1), 2018. doi:10.1016/j.pecs.2017.10.001.

[28] M. Kraft and S. Mosbach. The future of computational modelling in reaction engineering. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1924):3633–3644, 2010. doi:10.1098/rsta.2010.0124.

[29] N. Krdzavac, S. Mosbach, D. Nurkowski, P. Buerger, J. Akroyd, J. Martin, A. Menon, and M. Kraft. An ontology and semantic web service for quantum chemistry calculations. *Journal of Chemical Information and Modeling*, 59(7):3154–3165, 2019. doi:10.1021/acs.jcim.9b00227.

[30] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, and C. Bizer. DBpedia – A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, 6(2):167–195, 2015. doi:10.3233/SW-140134.

[31] Y.-P. Li, K. Han, C. A. Grambow, and W. H. Green. Self-evolving machine: A continuously improving model for molecular thermochemistry. *The Journal of Physical Chemistry A*, 123(10):2142–2152, 2019. doi:10.1021/acs.jpca.8b10789.

[32] Q. Lin, K. L. Tay, D. Zhou, and W. Yang. Development of a compact and robust polyoxymethylene dimethyl ether 3 reaction mechanism for internal combustion engines. *Energy Conversion and Management*, 185:35–43, 2019. doi:10.1016/j.enconman.2019.02.007.

[33] H. Liu, Z. Wang, J. Wang, and X. He. Improvement of emission characteristics and thermal efficiency in diesel engines by fueling gasoline/diesel/PODEn blends. *Energy*, 97:105–112, 2016. doi:10.1016/j.energy.2015.12.110.

[34] J. Morbach, A. Yang, and W. Marquardt. OntoCAPE — A large-scale ontology for chemical process engineering. *Engineering Applications of Artificial Intelligence*, 20(2):147–161, 2007. doi:10.1016/j.engappai.2006.06.010.

[35] S. Mosbach, A. Braumann, P. L. Man, C. A. Kastner, G. P. Brownbridge, and M. Kraft. Iterative improvement of Bayesian parameter estimates for an engine model by means of experimental design. *Combustion and Flame*, 159(3):1303–1313, 2012. doi:10.1016/j.combustflame.2011.10.019.

[36] S. Mosbach, J. H. Hong, G. P. Brownbridge, M. Kraft, S. Gudiyella, and K. Brezinsky. Bayesian error propagation for a kinetic model of n-propylbenzene oxidation in a shock tube. *International Journal of Chemical Kinetics*, 46(7):389–404, 2014. doi:10.1002/kin.20855.

[37] S. Mosbach, A. Menon, F. Farazi, N. Krdzavac, X. Zhou, J. Akroyd, and M. Kraft. A multi-scale cross-domain thermochemical knowledge-graph. 2020. Submitted for publication. Preprint available from: https://como.ceb.cam.ac.uk/preprints/257/. Accessed: 13th Oct 2020.

[38] M. Pan, J. Sikorski, C. A. Kastner, J. Akroyd, S. Mosbach, R. Lau, and M. Kraft. Applying industry 4.0 to the jurong island eco-industrial park. *Energy Procedia*, 75:1536–1541, 2015. doi:10.1016/j.egypro.2015.07.313.

[39] W. Phadungsukanan, M. Kraft, J. A. Townsend, and P. Murray-Rust. The semantics of Chemical Markup Language (CML) for computational chemistry: CompChem. *Journal of Cheminformatics*, 4(1):15, 2012. doi:10.1186/1758-2946-4-15.

[40] S. Ren, Z. Wang, B. Li, H. Liu, and J. Wang. Development of a reduced polyoxymethylene dimethyl ethers (PODEn) mechanism for engine applications. *Fuel*, 238:208–224, 2019. doi:10.1016/j.fuel.2018.10.111.

[41] D. A. Sheen and H. Wang. The method of uncertainty quantification and minimization using polynomial chaos expansions. *Combustion and Flame*, 158(12):2358–2374, 2011. doi:10.1016/j.combustflame.2011.05.010.

[42] J. J. Sikorski, G. Brownbridge, S. S. Garud, S. Mosbach, I. A. Karimi, and M. Kraft. Parameterisation of a biodiesel plant process flow sheet model. *Computers & Chemical Engineering*, 95:108–122, 2016. doi:10.1016/j.compchemeng.2016.06.019.

[43] I. Sobol. On the systematic search in a hypercube. *SIAM Journal on Numerical Analysis*, 16(5):790–793, 1979. doi:10.1137/070709359.

[44] W. Sun, G. Wang, S. Li, R. Zhang, B. Yang, J. Yang, Y. Li, C. K. Westbrook, and C. K. Law. Speciation and the laminar burning velocities of poly(oxymethylene) dimethyl ether 3 (POMDME3) flames: An experimental and modeling study. *Proceedings of the Combustion Institute*, 36(1):1269–1278, 2017. doi:10.1016/j.proci.2016.05.058.

[45] T. Turányi and H. Rabitz. Local methods. In A. Saltelli, K. Chan, and E. M. Scott, editors, *Sensitivity Analysis*, Wiley Series in Probability and Statistics, pages 81–99. John Wiley & Sons, New York, 2000.

[46] J. Warnatz, U. Maas, and R. W. Dibble. *Combustion: Physical and Chemical Fundamentals, Modeling and Simulation, Experiments, Pollutant Formation*. Springer, Berlin, 2006.

[47] C. Yu, M. Seslija, G. Brownbridge, S. Mosbach, M. Kraft, M. Parsi, M. Davis, V. Page, and A. Bhave. Deep kernel learning approach to engine emissions modeling. *Data-Centric Engineering*, 1:e4, 2020. doi:10.1017/dce.2020.4.

[48] C. Zhang, A. Romagnoli, L. Zhou, and M. Kraft. Knowledge management of eco-industrial park for efficient energy utilization through ontology-based approach. *Applied Energy*, 204:1412–1421, 2017. doi:10.1016/j.apenergy.2017.03.130.

[49] L. Zhou, M. Pan, J. J. Sikorski, S. Garud, L. K. Aditya, M. J. Kleinelanghorst, I. A. Karimi, and M. Kraft. Towards an ontological infrastructure for chemical process simulation and optimization in the context of eco-industrial parks. *Applied Energy*, 204:1284–1298, 2017. doi:10.1016/j.apenergy.2017.05.002.

[50] L. Zhou, C. Zhang, I. A. Karimi, and M. Kraft. An ontology framework towards decentralized information management for eco-industrial parks. *Computers & Chemical Engineering*, 118:49–63, 2018. doi:10.1016/j.compchemeng.2018.07.010.

[51] X. Zhou, A. Eibeck, M. Q. Lim, N. B. Krdzavac, and M. Kraft. An agent composition framework for the J-Park Simulator - A knowledge graph for the process industry. *Computers & Chemical Engineering*, 130:106577, 2019. doi:10.1016/j.compchemeng.2019.106577.

[52] X. Zhou, M. Q. Lim, and M. Kraft. A Smart Contract-based agent marketplace for the J-Park Simulator - a knowledge graph for the process industry. *Computers & Chemical Engineering*, 139:106896, 2020. doi:10.1016/j.compchemeng.2020.106896.