# Fast optimization of combustion mechanisms through model and experimental data reduction

László Horváth[1,2], Máté Papp[3], Tamás Turányi[3], Tibor Nagy[*1]

[1]*Institute of Materials and Environmental Chemistry, Research Centre of Natural Sciences, Budapest, Hungary*
[2]*Hevesy György Doctoral School of Chemistry, Eötvös Loránd University, Budapest, Hungary*
[3]*Institute of Chemistry, Eötvös Loránd University, Budapest, Hungary*

---

**Abstract**

Bioalcohols are promising alternatives to reduce dependence on fossil fuels. While bioethanol is widely used, its properties limit its compatibility with current petrol engines. Longer-chain alcohols, such as biobutanol and biopentanol, are more suitable and can be used in existing vehicles. Among several combustion mechanisms developed for the four butanol isomers, the mechanism of Sarathy et al. shows the best overall performance but still has significant room for improvement. This work focuses on improving its predictivity through kinetic parameter optimization to enhance its performance for a large set of combustion data. Parameter optimization can be very challenging in the case of large combustion models and/or large experimental data collections. Recently, Horváth et al. proposed the Mechanism Reduction Assisted Parameter Optimization and Model Development (RAPOD) method to overcome the challenge of optimizing the *n*-pentanol chemistry model in the large multifuel NUIGMech 1.1 combustion mechanism against a small data collection. According to RAPOD, a reduced mechanism is developed first that can very accurately reproduce the simulation results of the detailed model, and then its parameters are optimized and reintroduced into the full model. In the case of butanol, due to the huge size of the data collection, it is not feasible to perform the parameter optimization of the Sarathy model, even when using a reduced model. In this work, the RAPOD approach has been extended to reduce the large data collection as well. The optimized model shows similar improvements in accuracy for the entire data collection. We downloaded the experimental ignition delay time data (835 data points) collected for butanol isomers from the ReSpecTh information system. First, the Sarathy butanol mechanism was reduced, considering a small number of representative conditions covering the relevant conditions. Then, representative data points were selected from the data collection by dividing the condition and simulation error space into bins and taking single representative points from each with appropriate weights. The extended RAPOD method allowed for faster optimization of two orders of magnitude.

*Keywords: Combustion chemistry; mechanism optimization; biofuels; green chemistry*

---

*Corresponding author: nagy.tibor@ttk.hu

## 1. Introduction

Our modern society relies on the large-scale consumption of energy. Our household appliances, working tools, services, and methods of transportation are working using the power of electricity or direct burning of fuel. Today, energy production is still overwhelmingly dominated by the combustion of fossil fuels, whose resources, according to some more pessimistic outlooks, may be depleted within half a century at the current rate of utilization.

Popularly considered renewable alternatives to transportation fuels are bioalcohols. Bioethanol is already in use as a petrol additive in most countries (e.g. E10 petrol standard), and in certain countries (e.g. Brazil) as the main fuel component. However, the hygroscopic property of ethanol causes limited mixing with petroleum fuels and may also cause corrosion in the fuel system. For this reason, longer chain bioalcohols are considered as a viable alternative, although their production is not economical yet. Larger bioalcohols like n-butanol and n-pentanol are expected to be better ingredients, as due to their longer carbon chains, their physical and combustion properties are more similar to those of hydrocarbon fuel. Regarding the combustion of the long chain bioalcohols, biobutanol has already been researched for more than a decade, whereas on n-pentanol combustion still relatively little is known.

Kinetic models are developed for understanding the combustion of fuels and to promote the development of new combustion technologies based on them. These kinetic models contain not only the reaction steps but also their rate parametrization, which, in theory, defines their pressure and temperature dependence. Furthermore, thermodynamic and transport data are provided for each chemical species with the models; thus, it allows simulations of systems with complex geometries. Nowadays, computational fluid dynamics (CFD) simulations using chemical kinetic models are central tools in the development of modern combustion devices.

The assembled models are then validated against data from indirect experimental measurements, whose results can be simulated using only detailed mechanisms; that is, they contain indirect information on the rate of elementary reactions. Such experimental measurements are, for example, ignition delay times (IDT) measured in rapid compression machines (RCM) and shock tubes (ST), concentration data from jet-stirred reactor (JSR) measurements, and laminar burning velocities (LBV) from various devices, etc.

The size of detailed combustion mechanisms, depending on the fuel molecule, can range from tens of species and reactions (hydrogen, methane) to hundreds (natural gas, gasoline surrogates) and thousands (e.g. diesel surrogates, biodiesel). The accuracy of these models can be significantly improved by systematic parameter tuning; however, their sheer size makes their application unfeasible not only in CFD simulations but also for parameter tuning against a large amount of experimental data.

In our previous work, we presented a help of mechanism reduction to accelerate parameter optimization for large, multifuel mechanisms [1]. In that case, the simulation (and optimization) speed of NUIGMech 1.1 mechanism's [2] n-pentanol chemistry could be increased 40-fold. However, in the case of other biofuels, such as butanol, a significant amount of research data is available, making the optimization computationally expensive, while selecting only parts of them may decrease robustness (i.e. overall accuracy) of the optimized model. As these large experimental data collections contain substantial amounts of data measured at the same or similar conditions, the information they carry is often redundant. This redundancy allows the creation of a reduced data collection, which contains representative data points spanning the entire condition space of the data collection, with assigned weights to account for the redundant measurements and to mimic the error function over the whole data collection.

## 2. The investigated combustion mechanism

The Sarathy 2014 combustion mechanism [3] is designed to describe the combustion of a wide array of fuels: hydrogen, syngas, $C_1$-$C_5$ hydrocarbons, aldehydes and $C_1$-$C_4$ alcohols.

The original mechanism contains a large number of species (687) and reactions (3435), which makes its optimization time and power-consuming. This necessitated its reduction to accelerate its optimization using the RAPOD concept to a significantly smaller size: 285 species and 1535 reactions, which allowed a 10-fold increase in simulation speed with very similar predictivity.

## 3. Experimental data collection

The experimental data used as optimization targets includes many ignition delay time measurements from various research groups, compiled by Bolla et al. [4]. These measurements cover a wide range of experimental conditions and contain experiments with all four butanol isomers, albeit in different proportions.

All experimental conditions and measured data we collected were stored in ReSpecTh Kinetics Data (RKD) XML data files [5], which allow their automated use for setting up simulations in the Optima++ parameter optimization environment [6].

The total error of each experimental data series was estimated based on the error value given by the experimentalists and the inherent statistical noise[7] of the data series using the following

formula, where the reported experimental error was taken as $2\sigma_{\mathrm{exp}}$ unless it was stated differently in the original publication:

$$\sigma_{\mathrm{exp,tot}}^2 = \sqrt{\sigma_{\mathrm{exp}}^2 + \sigma_{\mathrm{stat}}^2} \qquad (1)$$

## 4. Reduction of experimental data (RAPOD)

To reduce the computational time and resources of the optimization work, we propose the following framework of data reduction-assisted optimization:

1. Selecting a representative set of experimental data, which covers all regions of the condition space of experiments.
2. Optimization of the important parameters of the mechanism against the representative set of data.
3. Validating the optimized model against the total data collection.

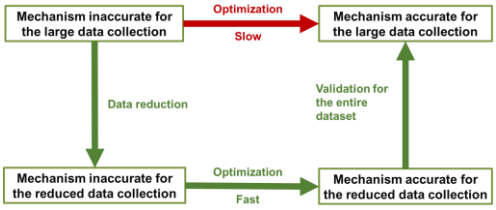


**Figure 1.** Reduction-Assisted Parameter Optimization and Model Development (RAPOD)

This method can be (as in this case) combined with the previous model reduction method. [1]

## 5. Data reduction procedure

The following protocol is proposed to reduce the data collection to a representative subset.

1. In the first step, the parameters used for the categorization and segmentation of the data are defined and classified according to various aspects:

**a. experimental parameters:** the experiment method and the measured quantity, the fuel composition (mole fraction of components or index of isomers), temperature (*T*), pressure (*p*), dilution (*dil*) and equivalence ratio ($\varphi$).

**b. uncertainty normalized signed simulation error** (*D*): repeated measurements by different laboratories can have different predictions for a measured quantity, thus their initial mechanism can be significantly different. The outcome of the optimization largely depends on which of them is selected into the reduced data collection.

**c. continuous parameters (i.e. real numbers):** *p*, *T*, *dil*, $\varphi$, *D*. Each of these continuously changing values with many distinct values in the data collection is assigned to a low number of bins. Before binning, these parameter values can be transformed to reciprocal (e.g. 1/*T*) or logarithmic scales (e.g. $\log_{10}p$, $\log_{10}\varphi$). The resolution of this segmentation can be adjusted for each parameter separately.

**d. discrete parameters:** pure fuels, isomers, experiment type, measured quantity, **and any condition parameter which takes on only few values** in the data collection, thus can be indexed with small integer numbers. Here, every distinct value is considered and not represented by other values.

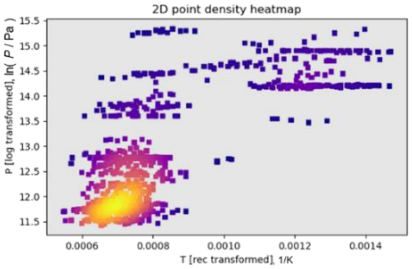


**Figure 2.** Point density of data on a *T-P* map

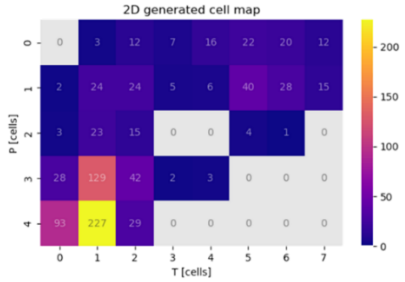


**Figure 3.** An example of generated cells on a *T-P* map

2. From each cell defined by the discrete and continuous parameters, one representative point is selected. Multiple possible selection methods were tested, including selecting the point with the highest simulation error (*D*) and with the median error value. Empty cells are ignored in the following.

3. Various combination of a number of bins is considered (e.g. (3,3,3), (4,3,3), (3,4,3), etc for ($p$,$T$, $\varphi$)), and the representative experiments are picked from them. The coverage of the full data collection by the reduced set of points is characterized by the convex hull ratio of the reduced data collection to that of the original one:

$$Q_{\mathrm{C.H.}} = \frac{V_{red}}{V_{full}} \qquad (2)$$

4. The accepted representative data collection is the smallest set where the convex hull ratio is higher than a pre-defined limit (e.g. 90%)

The algorithm has been implemented as Python code, and will be available on the respecth.hu webpage upon journal publication of the method.

## 6. Selection of active parameters

Detailed combustion mechanisms usually contain large number of uncertain parameters. Frenklach et al. [8] suggested that only those parameters ('the active parameters') that have a high influence on the simulated value of the experimental data need to be fitted. We carried out a brute-force local sensitivity analysis in the Optima++ [6] framework, using the OpenSMOKE++ [8,9] simulation code. Ranking the global maxima of the reaction sensitivities.

## 7. Mechanism optimization procedure

For the parameter optimization we applied the Optima++ code [10], which implements the optimization methodology developed by Turányi et al. [11] and uses the numerical optimization algorithm FOCTOPUS developed by T. Nagy [11,12]. In this work we used the OpenSMOKE++ [9,13] code for the combustion simulations. The Optima++ code minimizes the following error function:

$$E(\mathbf{P}) = \frac{1}{N}\sum_{f=1}^{N_f}\sum_{s=1}^{N_{fs}}\frac{w_{fsd}}{N_{fsd}}\sum_{d=1}^{N_{fsd}}\left(\frac{Y_{fsd}^{\mathrm{sim}}(\mathbf{P}) - Y_{fsd}^{\mathrm{exp}}}{\sigma_{fsd}^{\mathrm{exp,tot}}}\right)^2, \qquad (3)$$

where $N$ is the total number of data series, $N_f$ is the number of data files, $N_{fs}$ is the number of data series in the $f$-th file and $N_{fsd}$ is the number of data in its $s$-th data series. In this data series, $Y_{fsd}^{\mathrm{exp}}$ is the optionally transformed $d$-th data and $Y_{fsd}^{\mathrm{sim}}$ is its simulated value, which depends on parameter vector **P** that contains transformed Arrhenius parameters ($\ln A$, $n$, $E/R$). $R$ is the universal gas constant. The $w_{fsd}$ factors are the weights, which can help to develop optimized mechanism with balanced performance in reproducing all type of experimental data in the case of an unbalanced data collection.

The experimental uncertainty of a data point is also taken into account via the $\sigma_{fsd}^{\mathrm{exp,tot}}$ standard deviation of its determination. Thus, square root of $E$(**P**) measures the root mean squared deviation of the simulation results from the experimental data relative to the standard deviation of the experimental data. Square root of $E$(**P**) is around 1 for a perfect model, and below 3 for a model which is accurate on average within three experimental standard deviation.

The selected Arrhenius parameters were optimized in such a way that the rate coefficients calculated with the optimized rate parameters always remained within their prior uncertainty interval [$k_{\min}(T)$; $k_{\max}(T)$] in the temperature interval of 600 K to 3000 K. The uncertainty parameter is defined as the radius of a symmetric uncertainty range around the nominal $k^0(T)$ value on $\log_{10}$ scale:

$$f_{\mathrm{prior}}(T) = \log_{10}\frac{k_{\max}(T)}{k^0(T)} = \log_{10}\frac{k^0(T)}{k_{\min}(T)}. \qquad (4)$$

As prior information was not known for the uncertainty range of the influential reactions, one order of magnitude of uncertainty was assumed for the rate coefficients in the 600-3000K temperature range, which corresponded to $f_{\mathrm{prior}} = 1$ value. The optimization was continued for 500 iterations, with 3 random parameter sets tested in each iteration. During the optimization, the parameter sets improving for the reduced dataset were also validated for the full one. The mechanism improved for the reduced set was retained for the next iteration, even if the validation error increased.

## 8. Results

The following combinations of convex hull accuracy and selection rules were tested:

**Table 1.** Table of reduced data collections

| Codeword | $Q_{\mathrm{C.H.}}$ | Selection rule |
|---|---|---|
| CH90max | 90% | max($E$) |
| CH80max | 80% | max($E$) |
| CH70max | 70% | max($E$) |
| CH60max | 60% | max($E$) |
| CH90med | 90% | median($E$) |
| CH80med | 80% | median($E$) |
| CH70med | 70% | median($E$) |
| CH60med | 60% | median($E$) |

Testing these settings shows the profound effect of both the convex hull accuracy and selection rule in the overall improvement and correlation between the optimization and validation error function curves: The results show that a median selection rule provided greater retention of correlation, even with smaller convex hull accuracy criteria, while selecting a higher convex hull accuracy decreased the shift between the representative and validation dataset. CH90med, in particular, showed remarkable agreement between the two, staying accurate throughout the optimization. However, the overall outcome of full set accuracy was very similar in lower accuracy median selections, such as CH70med.

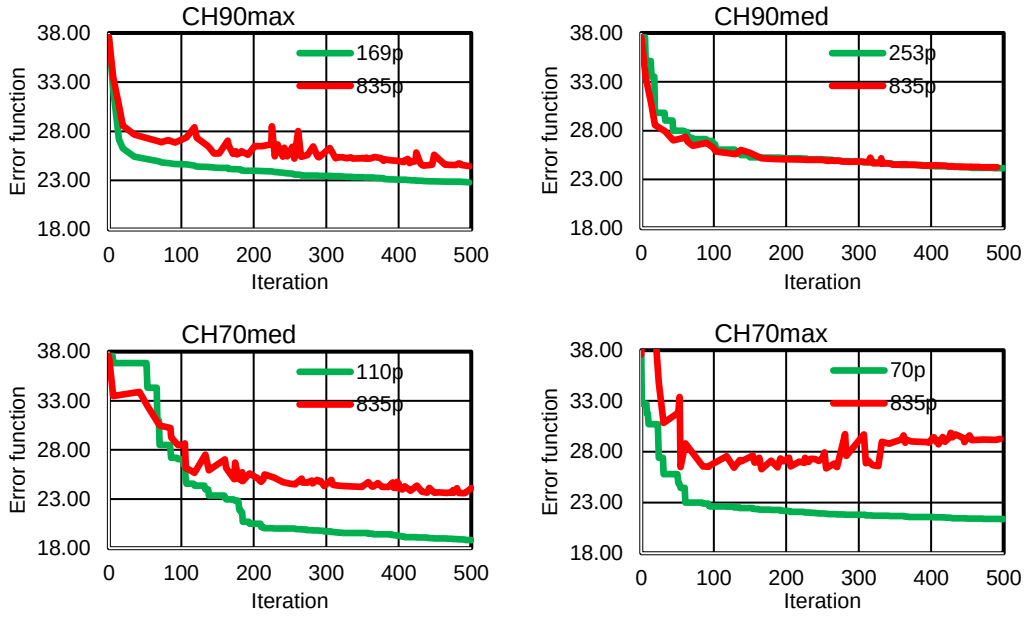


**Figure 4.** Comparison of various reduction settings regarding validation consistency

If we compare these optimization protocols with the the direct optimization of the full data collection, including the time necessary to reach the best results, we get the following results:

With an initial *E* of 37.6, each reduced data collection shows significant improvement in 500 iterations, with CH70med (median selection, with 70% convex hull accuracy) showing similar error reduction as the original set, with a significantly lower (44 hours compared to 365) optimization time. If we consider the tenfold increase in simulation time due to mechanism reduction, the process was roughly 80 times faster than direct optimization of Sarathy 2014.

**Table 2.** Comparison of reduced datasets with a direct optimization

| Codeword | Lowest *E* in 500 iterations | Time (500 iterations) (h) |
|---|---|---|
| Original | 23.7 | 365.4 |
| CH90max | 24.4 | 62.3 |
| CH80max | 24.5 | 44.8 |
| CH70max | 26.3 | 38.0 |
| CH60max | 25.0 | 27.6 |
| CH90med | 24.2 | 95.0 |
| CH80med | 24.9 | 52.5 |
| CH70med | 23.6 | 44.0 |
| CH60med | 26.3 | 37.4 |

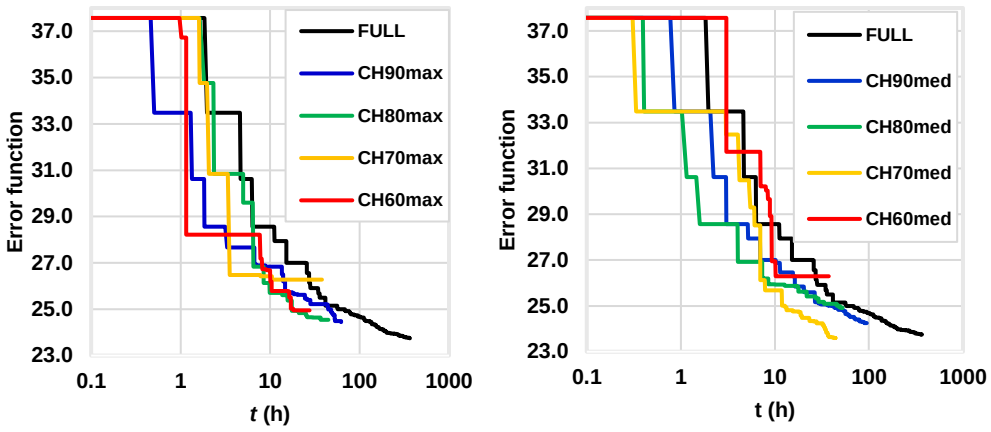


**Figure 5.** Diagrams of optimizations on reduced data collections

The optimized results show a significant improvement in the simulation accuracy, reducing the overall error function by 37.19%. This is comparable to the improvement on the original data, which was 36.80%.

There was a general improvement for both HPST and RCM simulations, with a larger improvement in RCM. For the butanol isomers, primer and seconder butanol showed moderate improvement, while tertiary butanol experiments' error improved significantly. In the case of isobutanol, there was no significant change in simulation error.

If we break down the error statistics for the two optimizations, the distribution is remarkably similar:

**Table 3.** Comparison of full- and reduced dataset optimization results

| | $\sqrt{E}$ | | |
|---|---|---|---|
| Data | Orig. | Opt.full | Opt- red. |
| HPST | 6.35 | 5.69 | 5.45 |
| RCM | 6.45 | 3.79 | 3.69 |
| I-BuOH | 4.42 | 4.26 | 4.45 |
| N-BuOH | 6.98 | 6.21 | 5.66 |
| S-BuOH | 5.86 | 5.08 | 5.44 |
| T-BuOH | 6.27 | 3.51 | 3.75 |

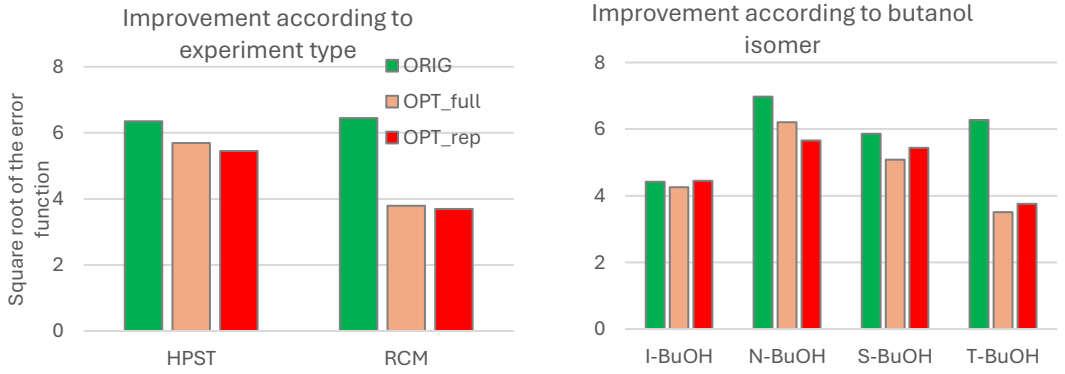


**Figure 6.** Comparison of full- and reduced dataset optimization results

## 9. Conclusions

The goal of this research is to develop a method to generate small, representative sets from large amounts of experimental data. This, combined with mechanism reduction, can increase the optimization efficiency of large multifuel mechanisms for fuels where a large amount of experimental data is available.

By mapping the experimental data points on a multidimensional grid defined by experimental parameters, we could generate a representative data collection, which allowed us to optimize the Sarathy 2014 mechanism's butanol ignition delay time simulation accuracy with very similar results to the full dataset control optimization, with 8 times less simulation time, with an 80 times speed increase combined with mechanism reduction.

While the method is capable of producing excellent reduced data collections, the proper cut-off criteria of convex hull accuracy is yet to be defined. Alternatively, an iterative reduction, where points can be added and removed based on the validation results, could further improve the speed increase. Additionally, the discretization of parameters could be improved by discretizing parameters with frequently used values (like $\varphi$) into groups according to these frequent values.

## Acknowledgements

T. Nagy and T. Turányi thank the support of the National Research, Development, and Innovation Fund (NKFIH) FK134332 and K147024 grants, respectively. L. Horváth thanks the support of the COST Action CYPHER CA22151 and the European Cooperation in Science and Technology (COST) association.

## References

[1] T. Horváth, L., Dong S., Saggese, C., Papp, M., Curran, H. J., Pitz, W. J., Turányi, T., Nagy, Mechanism reduction-assisted kinetic parameter optimization for the n-pentanol chemistry of the NUIGMech multifuel combustion mechanism, Proc. Eur. Combust. Meet. Rouen, Fr. (2023).

[2] S. Martinez, M. Baigmohammadi, V. Patel, S. Panigrahy, A.B. Sahu, S.S. Nagaraja, A. Ramalingam, A.A.E.S. Mohamed, K.P. Somers, K.A. Heufer, A. Pekalski, H.J. Curran, An experimental and kinetic modeling study of the ignition delay characteristics of binary blends of ethane/propane and ethylene/propane in multiple shock tubes and rapid compression machines over a wide range of temperature, pressure, equivalence ratio, , Combust. Flame 228 (2021).

[3] S.M. Sarathy, P. Oßwald, N. Hansen, K. Kohse-Höinghaus, Alcohol combustion chemistry, Prog. Energy Combust. Sci. 44 (2014) 40–102.

[4] M. Bolla, M. Papp, C. Olm, H. Böttler, T. Nagy, I.G. Zsély, T. Turányi, Comparison and Analysis of Butanol Combustion Mechanisms, Energy and Fuels 36 (2022) 11154–11176.

[5] T. Varga, C. Olm, M. Papp, Á. Busai, I.G. Zsély, ReSpecTh Kinetics Data Format Specification v2.3, (2020).

[6] M. Papp, T. Varga, Á. Busai, I.G. Zsély, T. Nagy, T. Turányi, Optima++ package v2.5: A general C++ framework for performing combustion simulations and mechanism optimization, (2024).

[7] T. Nagy, T. Turányi, Minimal Spline Fit: a model-free method for determining statistical noise of experimental data series, in: Proc. 10th Eur. Combust. Meet., 2021: p. 336.

[8] M. Frenklach, H. Wang, M.J. Rabinowitz, Optimization and analysis of large chemical kinetic mechanisms using the solution mapping method - combustion of methane, Prog. Energy Combust. Sci. 18 (1992) 47–73.

[9] A. Cuoci, A. Frassoldati, T. Faravelli, E. Ranzi, OpenSMOKE++ 0.12.0 (2020).

[10] M. Papp, T. Varga, Á. Busai, I.G. Zsély, T. Nagy, T. Turányi, Optima++ v2.1.0: A general C++ framework for performing combustion simulations and mechanism optimization (2021.05.03), (2021).

[11] T. Turányi, T. Nagy, I.G. Zsély, M. Cserháti, T. Varga, B.T. Szabó, I. Sedyó, P.T. Kiss, A. Zempléni, H.J. Curran, Determination of rate parameters based on both direct and indirect measurements, Int. J. Chem. Kinet. 44 (2012) 284–302.

[12] S.K. Goitom, M. Papp, M. Kovács, T. Nagy, I.G. Zsély, T. Turányi, L. Pál, Efficient numerical methods for the optimisation of large kinetic reaction mechanisms, Combust. Theory Model. 26 (2022) 1071–1097.

[13] A. Cuoci, A. Frassoldati, T. Faravelli, E. Ranzi, OpenSMOKE++: An object-oriented framework for the numerical modeling of reactive systems with detailed kinetic mechanisms, Comput. Phys. Commun. 192 (2015) 237–264.