LUND UNIVERSITY
Faculty of Science

LUND UNIVERSITY

BACHELOR THESIS
SPRING 2018

Smooth N-dimensional response landscape used to facilitate kinetic mechanism reduction

Author:
Anton MOBERG

Supervisors:
Dr. Elna HEIMDAL NILSSON
Christoffer PICHLER

June 6, 2018

Abstract

The threat of global warming means that combustion physics has become a hot research topic in order to minimize the negative environmental effects that otherwise arise from combustion. In order to effectively simulate effects of changes in the fuel composition, a so called reduced kinetic reaction mechanism is needed. The reduction process contains several steps where one of the more important ones to facilitate the reduction process is to evaluate the accuracy of the reduced process compared to the detailed version with a so called response score. An important step in mechanism reduction is to split reversible reactions into irreversible ones in order to remove unnecessary directions. In this paper a method was developed to calculate the reverse reaction rates using NASA formatted thermodynamic constants but ended up unsuccessfully doing so. Most likely reason is erroneous assumptions in the Arrhenius equations. The main finding of this paper is that that a smooth response landscape does facilitate the automated reduction process in a computer program. Another important finding is that Mean Absolute Error might be able to evaluate a reduced mechanism on its own in contrast to using Root Mean Square error and Cross-Correlation together, due to its similar behaviour. This has the potential to increase the efficiency of the reduction process.

Contents

1 Introduction **3**

2 Combustion simulations **4**
2.1 Detailed chemistry 4
2.2 Reduced kinetic mechanisms 5
2.3 Thermodynamic relations 5
2.3.1 Modified Arrhenius equation 5
2.3.2 Standard-state thermodynamic properties 6
2.3.3 Thermodynamic coefficients and polynomials 6

3 Evaluation of accuracy with response functions **7**
3.1 Computer program 7
3.2 Response functions 8
3.3 Weighted arithmetic mean value 9

4 Method **9**
4.1 Calculating reverse reaction rates 9
4.2 Determining response functions 9
4.3 Calculating response score 10
4.4 Smooth response landscape 11
4.5 Usage within the program 11

5 Results and conclusions **11**
5.1 Reverse reaction rates 11
5.2 Smooth response landscape 12
5.2.1 Response functions 12
5.2.2 Evaluation of reduced mechanism 15

6 Summary and outlook **18**

Appendix A File formats **21**
A.1 NASA thermodynamic format 21
A.2 ACO93 example mechanism - reactions 1-6 21

Appendix B Response landscape and reverse reaction rates **22**
B.1 MAE for low σ_{input} 22
B.2 k_r plots for reactions 3-6 in ACO93 22

1 Introduction

Combustion processes are a fundamental part of modern society, and can found in electricity generation, heating, jet engines, and Internal Combustion (IC) engines in cars to mention a few. Coal, natural gas and petroleum power plants stand for the majority of electricity produced[1], and they all use combustion to drive the generators. The same combustion process is used to drive the IC engines in road-based transportation and the jet-engines in aeroplanes. However, due to the the looming threat of global warming, combustion physics has become a hot research topic and the research is essential to decrease the negative environmental effects caused by combustion. Researchers are trying to find ways to increase the efficiency of the combustion engine, and thus decrease the harmful effects, as well as finding fuel compositions that lead to lower amount of harmful pollution.

Reaction kinetics is an interdisciplinary field between chemistry and physics and is used to describe the chemical mechanisms in a combustion process. A combustion mechanism is a representation of all the intermediate chemical reactions that occur in between the initial reactants and final products. A combustion process in real life consists of a very large and complex reaction mechanism in the order of thousands of reactions. They are also heavily coupled to physical interactions such as turbulence, diffusion and convection which adds to the complexity. This means that the detailed mechanism is too complex to be usable in Computational Fluid Dynamics(CFD) and hence a reduced mechanism is needed. Detailed chemistry, CFD and the reduction of a mechanism will be further discussed in **sec. 2**. A reduced mechanism is, as the name suggests, a simplified form of the full mechanism which is used to make high performance and computationally cheap simulations of combustion. A reduced kinetic mechanism only considers the most important intermediate reaction steps in order to minimize the needed computational power without compromising the accuracy of the model. The reduction process usually reduces the detailed mechanism from a couple of thousands reactions down to around fifty. The performance of such a reduced mechanism is evaluated with a so called response landscape, which is used to find an ideal reduced mechanism. Different response functions are needed in order to evaluate different types of features in the mechanism data. Two commonly appearing response functions when comparing data-sets in general are the Root Mean Square Error, and Cross-Correlation functions and they will therefore be applied in this project. The values of these response functions are then combined into a so called response score which when measured for many different reduced mechanisms produce the previously mentioned response landscape.

In real life all chemical reactions are reversible, however for many reactions one of the directions dominates over the other, meaning that only one of them contribute in a meaningful way to the mechanism. This means that only one of the reactions are needed to accurately describe the process. On top of this, if one of the directions in an irreversible reaction is very fast, then the other direction is very slow. The slow direction takes a lot of time in the simulation without providing very much information, and can therefore, if removed, improve the efficiency of the simulation. For the reduced mechanism it is useful to split the reversible reactions into two separate irreversible ones and analyse them separately in order to reduce the amount of reactions in the final reduced mechanism as well as increasing the simulation speed[2]. Doing this by hand is a very labour intensive work which takes a lot of time due to the fact that the complex mechanisms contains many

thousands of intermediate reaction steps. Therefore, being able to let the computer do this automatically is of great importance in making the reduction process more effective.

This project is part the development of an automated program that reduces mechanisms, and the point of it is to show that for small changes of the parameters in an N-dimensional reaction mechanism will generate a smooth response landscape. This landscape is produced by a weighted sum of response functions which look at different features in the simulated data. This is done in order to facilitate the kinetic mechanism reduction process, since having a smooth response landscape means that one can compare two reduced mechanism to one another and get a concrete value of which one is more efficient. This will be used in order to help the selection of an ideal reduced kinetic mechanism. The reason a weighted sum is used when combining the response functions is to prioritize certain improvements. For example, an increase of a response function which previously gave a low value, factors in more in the total response landscape than the same increase of another response function which previously gave a high value since having the response functions giving opposite values, one good and one bad, is a clear indication that something is wrong. This in turn means that having response functions giving approximately the same value is an indication of a better mechanism and thus the response landscape has to mirror that. This project will also, as a secondary, discuss how to implement an automated method into the program that will automatically calculate the complementary reaction rates given some thermodynamic data for the species included in the reaction, i.e. reactants and products, and thus split a reversible reaction into two irreversible ones. The irreversible reactions will then be evaluated separately when considered as candidates for the reduced mechanism.

In order to evaluate the efficiency of the methods developed in the project, a combustion mechanism will be reduced using the newly implemented methods and then be compared to a previously reduced mechanism that does not utilize them.

2 Combustion simulations

In **Sec.1** it is touched upon that using the full mechanism for simulations of combustion is not possible because of the enormous amount of computational power it would require. This section will therefore in greater detail argue for why exactly a reduced mechanism is needed, and how the reduction process works.

2.1 Detailed chemistry

To improve the knowledge of combustion processes and deduce new findings, detailed kinetic mechanisms are studied. For simple fuels such as methane and hydrogen the detailed mechanism is small enough to successfully be applied in 0D and 1D simulations of combustion[2]. 0D and 1D simulations are the simplest form of combustion simulations and only cover the chemical processes in the combustion. They are therefore no longer accurate representation of the combustion in the case of real applications, such as in an IC engine or a jet-turbine. The combustion process become much more complex since it is affected by not only the chemistry but is also strongly coupled to turbulence, diffusion, and convection[3]. The processes of combustion in a complex system can be observed by doing quantitative experiments in real situations, however this is very expensive and

quite difficult. The alternative is to simulate these complex systems with something called Computational Fluid Dynamics (CFD)[4]. CFD is able to cover both the dynamic and chemical properties of the combustion by simulating the flow of the system, with the cost of requiring more computational power. Not even for methane and hydrogen are the detailed mechanisms small enough to be applicable in CFD, even less so for more complex fuels. To be able to simulate the combustion of any fuel with CFD or larger fuels with 0D and 1D, some type of reduced mechanism is needed to lower the computational cost.

2.2 Reduced kinetic mechanisms

In order to be usable for simulation in CFD the detailed kinetic mechanism needs to be reduced from several thousands reactions to under a hundred, while still maintaining the ability to predict the combustion as well as the detailed mechanism does.

The basics of the mechanism reduction process that is used in this project was outlined in **sec.1** where two methods are mentioned. The first one mentioned is called skeletal reduction, where unimportant reactions and species are removed with the help of something called sensitivity analysis[2]. Sensitivity analysis is done by slightly changing parameters in the detailed mechanism and then looking at how the final result is affected[3]. The second one is time-scale analysis which in this project is used when evaluating a reversible reaction. Due to the fact that one of the directions may have a vastly shorter time-scale than the other, time is lost without gaining a lot of information from evaluating the direction with longer time-scales. There are more methods for reduction of mechanisms, however, they are not as relevant for this project as the ones mentioned above. A paper published by Lu and Law[2] describe skeletal reduction, time-scale analysis and other methods in great detail.

2.3 Thermodynamic relations

To describe the chemistry of a combustion process, thermodynamic relations are needed. For sensitivity and time-scale analysis the most important thermodynamic relations are the specific molar entropy and enthalpy, specific heat capacity, and the modified Arrhenius equation.

2.3.1 Modified Arrhenius equation

The Modified Arrhenius equation is as the name suggests a modified version of the standard Arrhenius equation where the difference is where the so called pre-exponential factor, A, has a temperature dependence added, T^n. T is the temperature and n is a temperature constant which varies for different reactions. The modified Arrhenius equation is expressed as

$$k = AT^n \exp\left(\frac{-E_a}{RT}\right) \tag{1}$$

where k is the rate constant, E_a is the activation energy of the reaction, and R is the Universal Gas constant [8][3].

The forward and reverse rate constants are related to each other in terms of the reaction

equilibrium constant, K_{eq}^{c}

$$K_{eq}^{c} = \frac{k_f}{k_r} \tag{2}$$

where then k_f and k_r are the reaction rate constants for the forward and reverse cases, respectively [8][3].

The forward reaction rate constant is calculated from experimental data using **eq.1**. In order to find the reverse reaction constant, the equilibrium constant K_{eq^c} needs to be calculated. K_{eq} is expressed in terms of concentration units so to calculate the reverse reaction rate constant from thermodynamic properties, the equilibrium constant needs to be expressed in terms of the pressure. This means that we get the following expression

$$K_{eq}^{c} = K_{eq}^{p} \left(\frac{P}{RT} \right)^{\sum_{k=1}^{K} v_k} \tag{3}$$

where K_{eq}^{p} is the equilibrium Constants deduced from thermodynamic properties and v_k are the so called Stoichiometric coefficients.[6]

K_{eq}^{p} is expressed by simply combining the expressions for Gibbs free energy and Gibbs free energy for isotherm reactions [9] which give

$$K_{eq}^{p} = \exp\left(-\frac{\Delta H}{RT} + \frac{\Delta S}{T} \right). \tag{4}$$

2.3.2 Standard-state thermodynamic properties

The standard-state specific molar entropy and enthalpy is given by an integral expression of the molar heat capacity. The standard-state specific molar enthalpy is given by [3]

$$H^0 = \int_0^{T_k} C_p^0 dT + H^0(0) \tag{5}$$

and the standard-state specific molar entropy is given by

$$S^0 = \int_{298}^{T_k} \frac{C_p^0}{T} dT + S_k^0(0) \tag{6}$$

where C_p^0 is the specific molar heat capacity, T is the temperature and $S, H_k^0(0)$ refers to the standard-state molar entropy/enthalpy evaluated at the reference temperature [6].

2.3.3 Thermodynamic coefficients and polynomials

In order to empirically determine the standard-state molar entropy and enthalpy it is required to express the molar heat capacity as polynomial using thermodynamic coefficients. These coefficients are given in data files following the NASA thermodynamic format which contain 15 coefficients. The first 7 are the coefficients $a_{1,\ldots,7}$ for the upper temperature interval with $T \geq 1000$K. The following 7 ones are the coefficients $a_{1,\ldots,7}$ for the lower temperature interval with $T \leq 1000$K [5]. The last coefficient is not always needed but it represents the value of the standard-stade specific molar enthalpy at zero Kelvin, $H^0(298.15)$ [7].

In terms of these thermodynamic coefficients the heat capacity is expressed as

$$\frac{C_p^0}{R} = a_1 + a_2 T + a_3 T^2 + a_4 T^3 + a_5 T^4 \tag{7}$$

which means that the standard-state molar enthalpy, in terms of thermodynamic coefficients, is expressed as

$$\frac{H^0}{RT} = a_1 + \frac{a_2}{2} T + \frac{a_3}{3} T^2 + \frac{a_4}{4} T^3 + \frac{a_5}{5} T^4 + \frac{a_6}{T} \tag{8}$$

and the standard-state molar entropy as

$$\frac{S^0}{R} = a_1 \ln T + a_2 T + \frac{a_3}{2} T^2 + \frac{a_4}{3} T^3 + \frac{a_5}{4} T^4 + a_7 \tag{9}$$

where, as mentioned above, $a_{1,\ldots,7}$ are the NASA polynomial thermodynamic coefficients, T is the temperature and R is the molar gas constant[6][7].

3 Evaluation of accuracy with response functions

3.1 Computer program

The computer program, developed by PhD student Christoffer Pichler at the department of Combustion Physics at Lund University, serves as a basis for this project. The program is made to automate the reduction process of kinetic mechanisms, and without human interaction find the optimal mechanism to use for combustion simulations.

The program is split into three different main sections which perform different tasks, see **fig. 1** for a schematic. These sections are: mechanism reduction, simulation, and comparison. The mechanism reduction starts with the initial species that are are present before combustion process starts and ends with the final products of the reaction. It then writes out all the reactions that are present in the full kinetic mechanism like a tree of different reactions and species. Then by using an ant-colony optimization (ACO) algorithm different probability values are assigned to the different reactions[10]. Based on the probabilities a reduced mechanism is selected and sent to the simulation part of the program which obtains simulated results based on it. These results are then sent to the comparison part of the program where it evaluates and compares the simulated data to actual experimental data in order to determine the performance of the reduced mechanism.

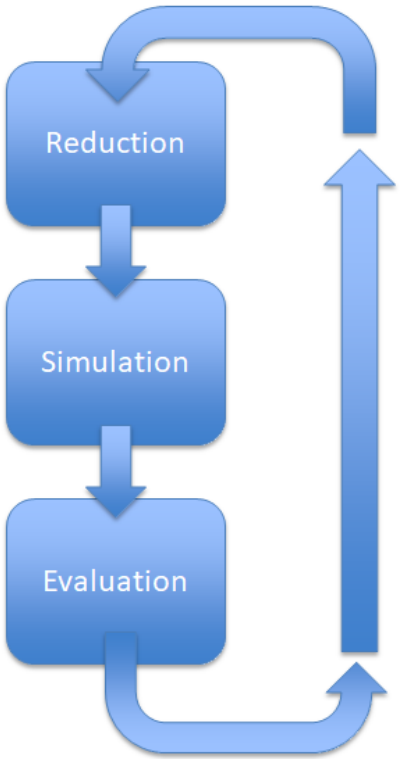

Figure 1: *Figure showing a general schematic of how the reduction program by Christoffer Pichler works.*

Depending on the response score of the reduced mechanism the probabilities in the mechanism tree are tweaked and the process is repeated. After many iterations the program will converge onto one mechanism which is the mechanism that simulates the combustion process the best under certain conditions such as the size (number of reaction steps) or simulation speed.

The response function currently used in the evaluation part of this program is defined as follows

$$RS(a,b) = \max\left(-1, 1 - \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{a_i}{b_i} - 1\right)^2}\right) \tag{10}$$

where a_i and b_i are the i:th value in the two datasets, a and b being compared to each other.

3.2 Response functions

In order to measure the accuracy of simulated data compared to the target data, some kind of normalized response functions are needed. Below this section some of the most widely used response functions are described.

Root Mean Squared Error (RMSE) is widely used when it comes to measuring the performance of models. RMSE provides a measurement of standard deviation between the predicted model and the observed data. The RMSE is a scale dependent measurement which means that it is only applicable if the data that is to be compared is on the same scale. RMSE is as the name suggests the root of the Mean Square Error (MSE)

$$RMSE = \sqrt{MSE} \tag{11}$$

where the MSE is described as the arithmetic mean of errors

$$MSE = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \tag{12}$$

where n is the total amount of data points and e_i is the error of data point with index i [11]. The RMSE is appropriate to use for error measurement where the error distribution is expected to follow a Gaussian and the errors are unbiased. If this is not the case then Mean Absolute Error (MAE) is a better choice to measure the performance [12].

MAE is less sensitive to outliers than RMSE and is therefore a better indicator of general features of the data however, just as RMSE the MAE is scale dependent [11]. MAE is expressed as

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |e_i| \tag{13}$$

where just as in the RMSE case n is the total amount of data points and e_i is the error of data point with index i [12].

Cross-correlation (CC) is a way to compare the similarity in shape for two functions. It measures the similarity of two signals as a function of time lag. Time-lag can be considered the displacement between the mean-values if Gaussian functions are considered

to be the signals. CC is used widely in signal analysis as a way to detect know features within a large signal, as well as time-series analysis. CC is defined as

$$(f \star g)(\tau) \equiv \int_{-\infty}^{\infty} f^*(t)\, g(t+\tau)\, dt \tag{14}$$

where f and g are two continuous functions and τ is the lag[13].

3.3 Weighted arithmetic mean value

The Weighted arithmetic mean values is a way to get certain data-points to contribute either more or less to the final mean value by multiplying weights to them. The mathematical formulation of the weighted arithmetic mean value is

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

where x_i is a set of mean values and w_i are the weights which sum to one, $\sum_{i=1}^{n} w_i = 1$ [14]. This weighted mean value is needed in order to accurately represent the performance of a reduced mechanism. This is due to the fact that bad response values have a bigger negative impact on the performance than the good response values have positive impact. This simply means that it's more appropriate to have all response functions give decent values rather than one giving a very high and the other one a very low value.

4 Method

4.1 Calculating reverse reaction rates

The method to calcualte the reverse reaction rates from the forward reaction rate will be described in this section. The general equations used are described in **Sec.2.3**, but this section will describe how the algorithm in the program uses those equations.

The algorithm starts by reading a NASA standard thermodynamic data file, for example format see **Appendix A.1**, containing thermodynamic coefficients with which it calculates the specific molar Entropy and Enthalpy, **Eq.8-9**. Using these expressions the equilibrium constant is caluclated with **Eq.3-4**. The algorithm then reads in the activation energy,temperature dependence and the pre-exponentialfactor for a irreversible reaction and calculates the forward reaction rate with the Arrhenius equation, **Eq.1**. Plugging these two results into **Eq.2**, and the reverse rate constants can be determined.

4.2 Determining response functions

In **Sec.3.2**, three types of functions are described, where CC is used to compare similarities in shapes between two functions or data-sets, and RMSE and MAE is used to compare the values at certain points. CC and RMSE are the two most commonly used ones for measuring the accuracy of data, however, MAE is often argued to be a better indicator than RMSE as described in the theory section. Determination of which response function/s to use is done by simply comparing the response given by them when fed know data. This is done by generating two Gaussian normal distributions keeping one at a

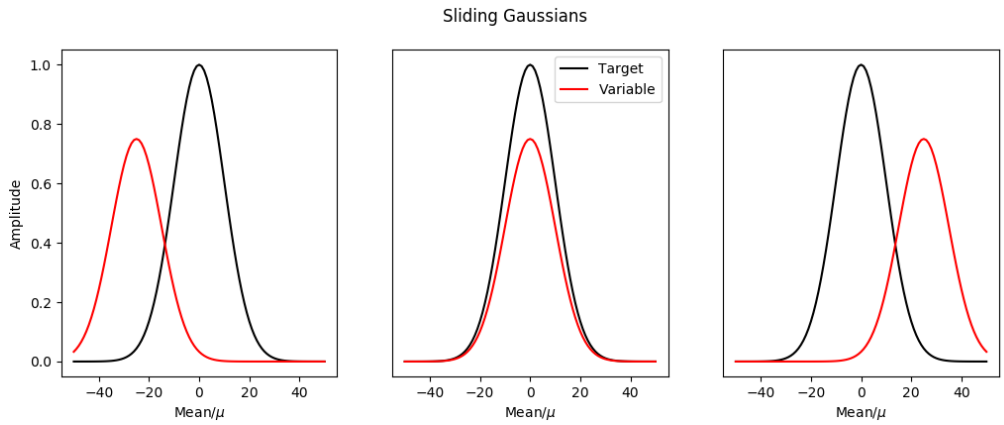

Figure 2: *Schematic of sliding two Gaussian functions ($\sigma = 10$), over each other by changing the mean for one (red) while keeping everything else fixed.*

fixed mean value while sliding the other one over it, see **fig.2**, and measuring the value of the response functions. This is repeated several times where the standard deviation for one of the Gaussian is varied. The data generated is then evaluated on the basis of identifying which response functions are sensitive to what features, and from that argue for which of them are applicable for the task of measuring the accuracy of a reduced kinetic mechanism.

4.3 Calculating response score

This method uses CC to compare the shape, and a normalized RMSE to compare the values. The functions take the data from the reduced mechanism and compares it to the target data given by the complex mechanism and output a value between 1 and -1, where 1 indicates 100% correlation and/or match of values for the respective functions and -1 indicates the opposite. The scores from the individual response functions are then added together by a weighted arithmetic mean value, where the weights are determined by a logistics function. This weighted average of the response functions is considered to be the accuracy for the reduced mechanism and is called the response score.

The response score method developed in this project is

$$RS(a,b) = w_1 \cdot CC + w_2 \cdot NRMSE \tag{15}$$

where w_i are weights, CC is the Cross-correlation and NRMSE is the normalized Root Mean Square Error between the output (a) and target (b) values. The weights are determined by:

$$w_1 = \begin{cases} 1 - \frac{1}{1+\exp(k \cdot NRMSE)}, & \text{for } CC < NRMSE \\ \frac{1}{1+\exp(k \cdot CC)}, & \text{for } CC > NRMSE \end{cases} \tag{16}$$

and $w_2 = 1 - w_1$, where k determines how sensitive the weights are. For positive k-values the algorithm will put more weight towards the lowest of CC and RMS values, for negative k-values the opposite is true. CC is calculated as

$$CC(a,b) = \begin{cases} 1, & \sigma_a = \sigma_b = 0 \\ \frac{\frac{1}{n}\sum_{i=1}^{n}(a_i \cdot b_i)}{\sigma_a \sigma_b}, & \sigma_a, \sigma_b \neq 0 \\ 0, & \sigma_a \sigma_b = 0 \text{ and } \sigma_a \neq \sigma_b \end{cases} \tag{17}$$

where σ_i is the standard deviation and n is the total number of data points. RMSE is calculated as

$$NRMSE(a,b) = \max\left(-1, 1 - \frac{RMSE}{denom}\right) \tag{18}$$

where $a_{max,min}$ and $b_{max,min}$ are the maximum and minimum values of the output and target data-sets respectively, $denom = \min\left(mean(a), mean(b)\right)$ and RMSE is defined by **eq.11**.

4.4 Smooth response landscape

In order to being able to use response score to help facilitate in the reduction of a mechanism it is of extreme importance that the response landscape is smooth, which means that the response score should change in controlled manner if some parameter is changed slightly. Showing that the response landscape is smooth is done by having two functions, one represents the target and the other one is varying. The two functions gets evaluated with the method described in the section above and the varying function is given a response score. By changing some of the parameters in the varying function, for example the standard deviation or mean value for a Gaussian, it will receive a new response score. This is repeated until sufficient data is collected in order to evaluate the behaviour of the response landscape.

4.5 Usage within the program

This method of determining the response score of a mechanism is part of the evaluation module of the program outlined in **Sec.3.1**. The weights used by the ACO-algorithm is determined in part from the response score and is modified accordingly for each iteration in order to facilitate the reduction and finding an accurate reduced mechanism. Simply put the program takes a complex kinetic mechanism and uses an ACO-algorithm in order to pick out a reduced mechanism. This reduced mechanism is used generate data that is fed into the method described in the previous section above and gets a numeric response score back, which represents the accuracy of the mechanism. The weights in the ACO-algorithm are modified in accordance to this value and another iteration is started. The reduced mechanism using this new method will be compared to another reduced mechanism using the old response score method, see **Eq. 10**, to determine the efficiency of the new method. The comparison is done by comparing the mechanism reduced with the old and new response method to AramcoMech 2.0 detailed mechanism with an arbitrary response function giving a value between -1 and 1. This is done in order to get a value that is comparable for the two methods. This is done since the response scores given by the old and new response methods are not directly comparable.

5 Results and conclusions

5.1 Reverse reaction rates

By using the method described in **Sec.4.1** the reverse reaction rate in a reversible reaction is calculated. The calculated reverse reaction rate for two reversible reactions in an example mechanism is presented with the actual reverse (obtained by MODMECH) and forward reaction rates in **Fig.3**. The rates are plotted against the temperature as the

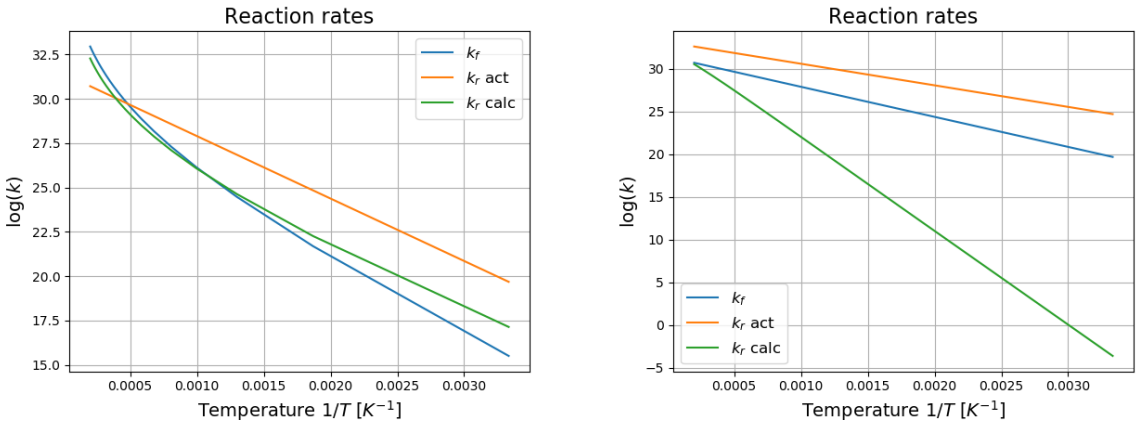

Figure 3: *Shows the forward, actual reverse, and calculated reverse reaction rates, k_f, k_r act, k_r calc, as a function of the inverse temperature, $1/T$. the left figure shows the first reversible reaction and the right figure shows the second for an example ACO92 mechanism.*

modified Arrhenius equation has a temperature dependence, see **Eq.1**. The calculated and actual reverse reaction rates does not match, as can be seen in the figure. For more figures of more reactions in the example ACO92 mechanism see **App.B.2**, and for the data used in calculating the reverse reaction rates in the presented reactions see **App.A.2**. The mismatch of reverse reaction rates between the calculated values and the values obtained by MODMECH indicates that something with the method used in this project is wrong. The method developed can therefore not be used within the program to facilitate the reduction process.

The method strictly use well established models developed by NASA, as is backed up in **Sec.2.3**, for determining K_{eq}^{c} in **Eq.3** which is used to determine k_r from **Eq.2**. This indicates that either MODMECH pre-process the thermodynamic data to handle some unknown condition or that one or more of the equations **Eq.1-9** are wrongly implemented in the program and/or wrongly written in the theory used for the method. The former has been checked multiple times, however it is still a possibility. The latter is based on left graph in **Fig.3**, where the calculated reverse reaction rate is curved while the actual one is linear. This indicates a difference in an exponent which means that there could be a potential error of assumption that the temperature constant n in **Eq.1** is the same for both forward and reverse case of a reversible reaction. However, there is probably some erroneous assumptions in the other equations as well due to the fact that for some cases the calculated and actual are both linear but still giving different values. For further research upon the matter these possible leads is where one should concentrate their efforts.

5.2 Smooth response landscape

5.2.1 Response functions

In **Sec.3.2** three response functions were mentioned, CC, RMSE and MAE that have different behaviours, where CC is comparing the shape of two functions or data-sets while RMSE and MAE are comparing the values. In order to determine which response functions are appropriate for usage when calculating the response score, the method described

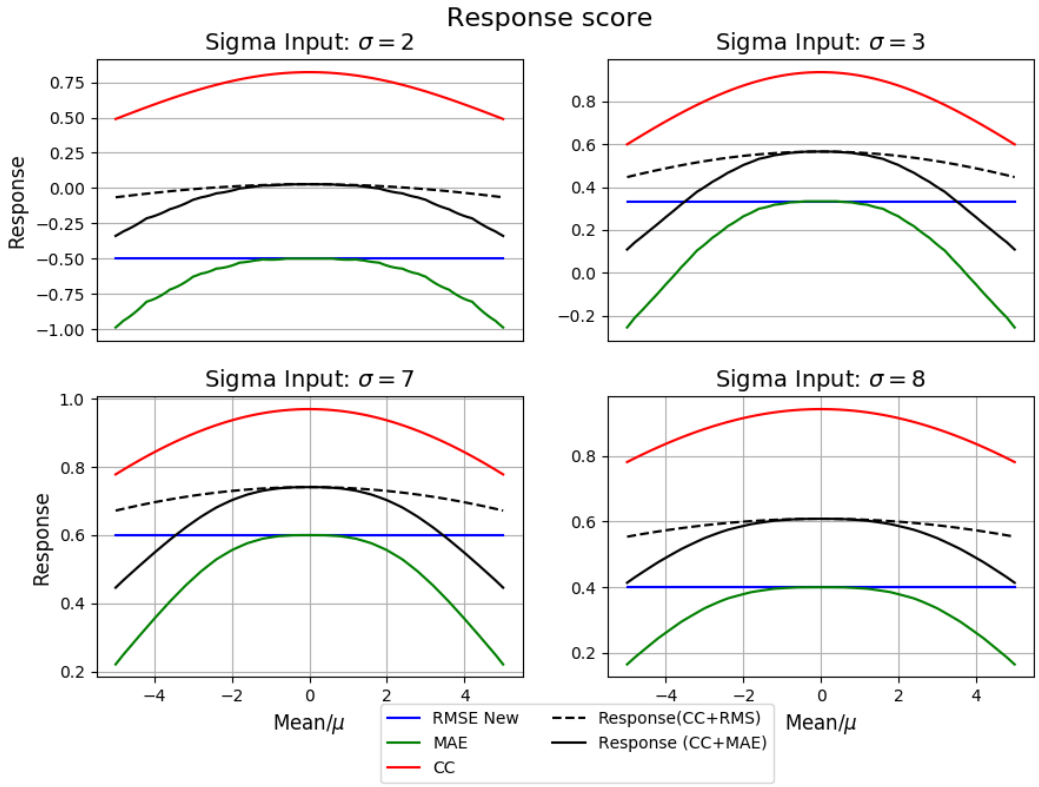

Figure 4: *Shows a series of values for the respective response functions (CC (Red), RMSE (Blue) and MAE (Green) when two Gaussians with different σ are compared by sliding one of them over the other (Target: $\sigma = 5$). The two black lines show the weighted average for RMSE+CC (dashed) and MAE+CC (line).*

in **Sec.4.2** is used. In **Fig.4** it is shown how the value from the different response functions are affected when the mean is varied for a fixed σ when comparing two Gaussians. **Fig.6** shows how the response score when the two Gaussians have the same standard deviation (left), $\sigma = 5$, and the same mean (right), $\mu = 0$. The right part in **Fig.5** show the same thing as **Fig.4** but where the target has been swapped to a linear function instead of a Gaussian.

The reason the collection of response values was done is in order to find the response functions that will yield the best landscape for the program to use when comparing mechanisms to aid the reduction process. For this to be the case the response landscape needs to be as smooth as possible and have a predictable behaviour, because if it is not then the ACO-algorithm might modify the wrong weights and end up with a mechanism that is actually worse then the previous one. By analysing the behaviour of the different response functions it can be deduced whether or not they are suitable to be used in the program.

CC is according to the theory presented in **Sec.3.2** supposed to be sensitive to shapes as it is measuring the amount of overlap between to functions and data-sets, which by looking at the **Fig.4** and **Fig.6** is consistent with the results. In **Fig.6** for the case when the two data-sets are identical in shape, i.e $\sigma_{target} = \sigma_{input}$, it can be seen that the value smoothly increase as the Gaussians are slid closer to each other. When they overlap to a 100% the maximum is reached and as they are now slid further and further away the value decrease in the same smooth way. This same behaviour can be observed for other

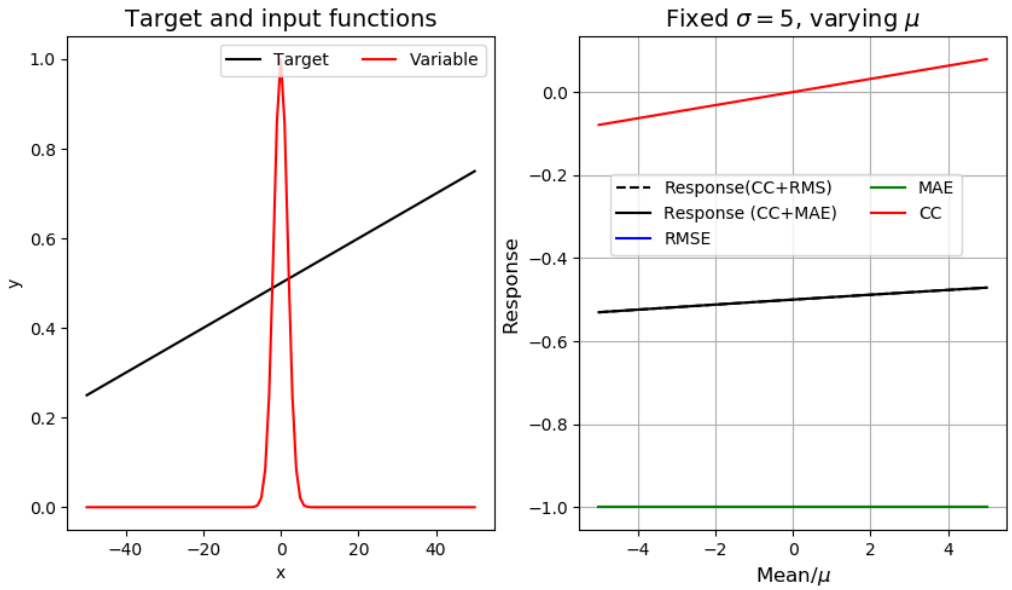

Figure 5: *Left figure shows the functions to be compared, where target (black) is a linear function and the input/variable (red) is a Gaussian with standard deviation $\sigma = 5$. The right figure shows the response values of CC (red), RMSE (blue), MAE (green), RMSE+CC (dashed), and MAE+CC (black) when the functions are compared.*

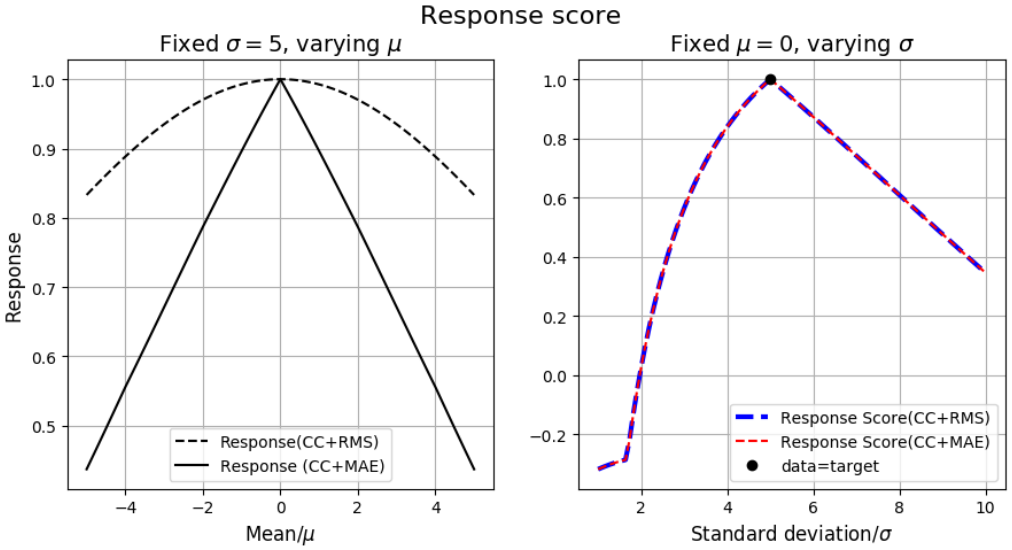

Figure 6: *Left figure shows the response values for RMSE+CC (dashed) and MAE+CC (black) when comparing two Gaussians with same standard deviation, $\sigma = 5$, while varying the mean, μ. The right figure shows the same RMSE+CC (blue) and MAE+CC (red) but with fixed $\mu = 0$ and varying σ. The red dot indicates where the input is the same as the target.*

cases when the target and input have different σ as seen in **Fig.4**. In **Fig.5** the target is swapped to a linear function instead and it can be seen that the CC reacts accordingly and as expected. These results confirm that CC is sensitive to the shape. It can therefore be concluded that CC is a solid choice for evaluating the shape of data.

RMSE is according to the theory presented in **Sec.3.2** supposed to look at the error in value is between the target and input. RMSE does therefore not explicitly depend on the shape of the functions. This can be observed to be the case in both **Fig.4**, **Fig.5**, and the left figure in **Fig.6** as the RMSE value is constant for every μ value, for a fixed σ. MAE on the otherhand is according to **Sec.4.3** also looking at the error in value between the target and the inpu but is slightly sensitive to outliers, i.e the shape. This can be observed to be the case as MAE has a shape that follows that of CC but always has it's maximum at the value of RMSE. This is an interesting behaviour as it is able to evaluate both the shape and the value, which is what is desired for the program. However, due to the slightly sporadic and wobbly behaviour for small σ, that can be observed in the top left figure of **Fig.4** MAE is not suitable for the program. For more examples of this behaviour see **Appendix B.1**. However, it is possible that this behaviour is due to the normalization process rather than the MAE itself. If μ is kept constant while varying σ, there is no difference between RMSE and MAE as can be deduced by **Fig.6**. This means that if it is found that the sporadic behaviour at low σ is occuring due to normalization and can be prevented, MAE could possibly be used alone in order to determine the response score and therefore increase the efficiency of the program. This needs to be studied further to concretely determine if possible though and is not done in this paper due to time constraints. Instead RMSE is chosen to evaluate the error in value between the data-sets.

5.2.2 Evaluation of reduced mechanism

The old method, see **Eq.10**, has a slightly different behaviour to the new one, where it is not constant over μ independent of σ_{input} as can be seen in **Fig.7** which shows the behaviour for the different methods for $\sigma input$ that is around $\sigma_{target} = 5$. This means that the response scores given during the reduction process are not comparable to each other since they depend somewhat arbitrarily set and not affected in the same manner for a given change.

The methods converge to a couple of mechanisms with different size and response scores. The mechanisms yielded from using the old and new methods in the reduction program, are presented in **Tab.1**. As can be seen the different methods give quite similar results both size wise and response wise. Which is an indication that the new method is, at least, as good as the old one meaning that they help facilitate the reduction process equivalently. However, since the responses are not directly comparable as mentioned before, there is still a possibility that one of the methods yield a better/worse result than the other. In order to put a numerical value to the performance of each of the methods the reduced mechanisms for each method they are simulated and compared to AramcoMech 2.0 detail mechanism with a separate evaluation algorithm. The results of the evaluation process can be seen in **Fig.8**, which shows how the simulation time is affected by the mechanism size, how the accuracy compared to AramcoMech 2.0 is affected by the size of mechanism size, and how the accuracy is affected by the response score of the mechanism, respectively. The figure does not show any major discrepancy between the new and the old method.

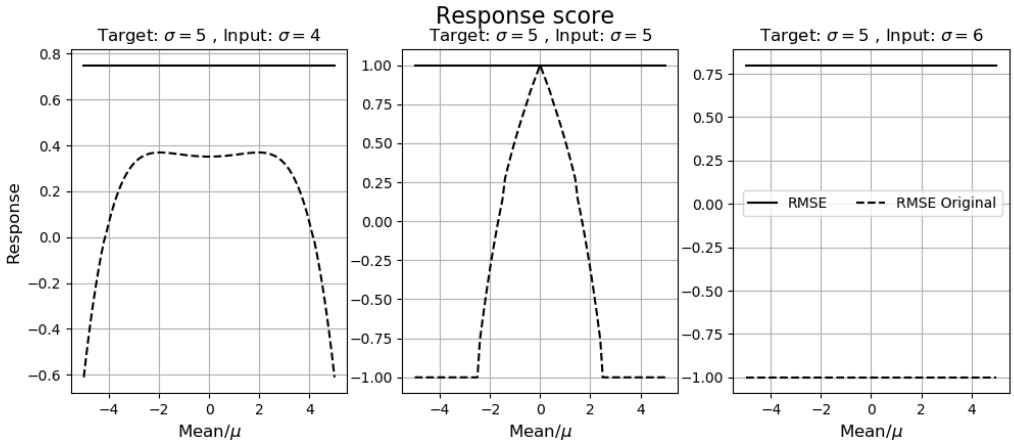

Figure 7: *Shows a series of response values for the old and new RMSE normalization method. The figures shows, from left to right,* $\sigma_{input} = 4, 5, 6$ *and* $\sigma_{target} = 5$.

The simulation time as affected by the reduced mechanism size appear to be more or less the same for both methods, however, it is difficult to give a definite conclusion due to the small amount of evaluated mechanisms and the somewhat sporadic distribution. The accuracy compared to AramcoMech 2.0 as affected by the size of the reduced mechanism are once again more or less the same for both the old and new method. This plot is less sporadic than the previous one and seem to follow a logarithmic behaviour, however, once again it is difficult to judge if this is the actual behaviour due to the few evaluated mechanisms. The accuracy as affected by the response score given by the response method is more slightly more differentiating between the old and new method, where the old method show a steeper linear behaviour than the new method for low accuracies. This means that for low accuracies the old method will show a higher response than the new method, which is not wanted. As with the other cases it is very difficult to draw any concrete conclusions from this result due to the low number of evaluated mechanisms, however, from the results that are available it seems as if the response score calculated with the new method is a slightly better representation of the accuracy of the reduced mechanism.

Table 1: *Table showing the size and response for different mechanisms found using the old and new method.*

Method	Response	Size
Old	0.574	30
Old	0.710	31
Old	0.831	35
Old	0.840	49
Old	0.861	52
Old	0.903	65
New	0.397	35
New	0.744	37
New	0.774	38
New	0.789	42
New	0.790	44
New	0.815	47
New	0.837	49
New	0.860	64

The reduction process takes a long time, and only two were able to be done during this project, one for the new method and one for the old method. The reduction process was made over 150 iterations (epochs) and the resulting time for the new and old method are presented in **Fig.9**. The old method took 12 hours and the new method just under 16 hours. As can be seen they are both following a linear behaviour for the most part with the exception of the new method getting an exponential increase for the last few epochs. From these results it seems as if the old method is slightly more effective time-wise, but worth noting is that from a sample size this small it is not possible to draw any concrete

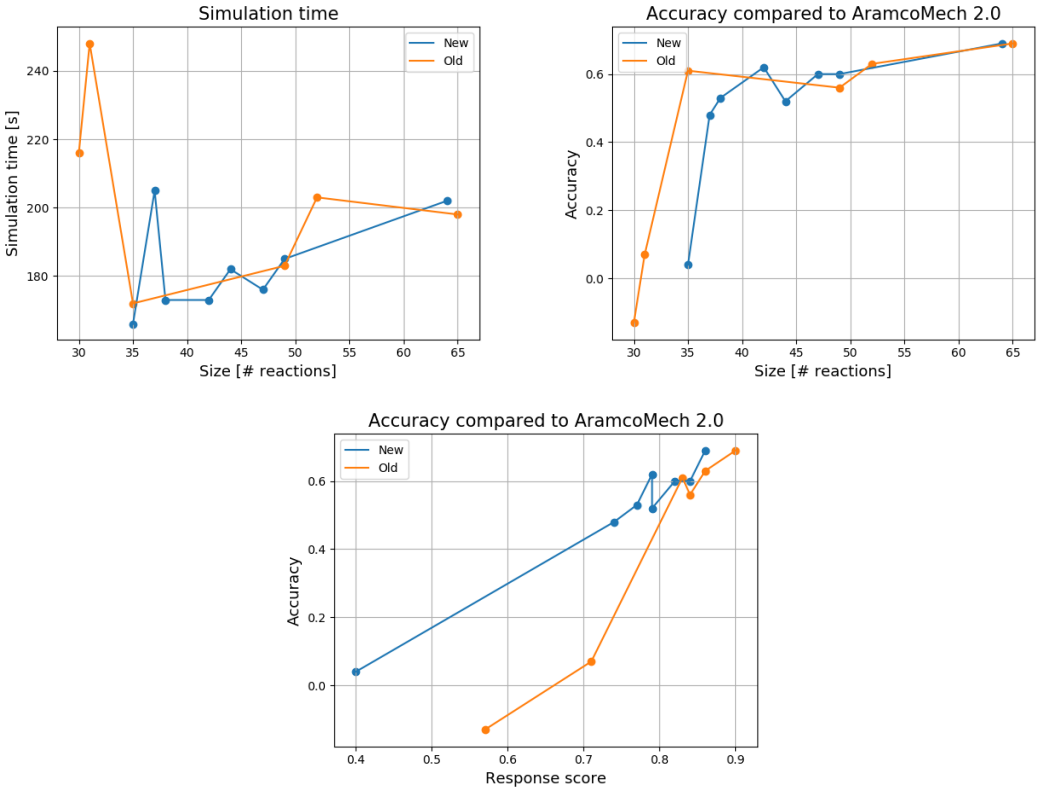

Figure 8: *Top left shows the simulation time as a function of mechanisms size, tp right shows the accuracy of the reduced mechanism compared to AramcoMech 2.0 as a function of the mechanism size and the bottom shows the accuracy of the reduced mechanism compared to AramcoMech 2.0 as a function of the response score of the mechanism. The accuracy is measured between* 1 *and* -1.

conclusions in this case. If the difference was much greater it would have been possible to assume that one of the methods are more effective than the other. In order to draw a more concrete conclusion the reduction process would have to be repeated multiple times to get a greater sample size to analyse.

The conclusion of the evaluation part of this project is that having a smooth response landscape produced by a response method will help facilitate the reduction of an N-dimensional kinetic mechanism as it allows the program to modify parameters in such a way as to converge onto an ideal reduced mechanism. The efficiency of the newly implemented method for calculating the response score can be concluded to be at least as good as the old method, with some features that indicate that it is gives a slightly better representation of accuracy. These features are difficult to base a certain conclusion around since there are very few reduced mechanisms available for analysis, however, from the available data it seems as if both methods are more or less equivalent. In order to come to a more certain conclusion, more reduced mechanisms facilitated by the two methods would be required. This is obtained by running the reduction process for each method several more times in order to find more reduced mechanisms which can be used in the analysis. With more mechanisms available for analysis the data points in **Fig.8** would be able to be fitted to a mathematical model and from it the efficiency of each method can be deduced with much greater accuracy.

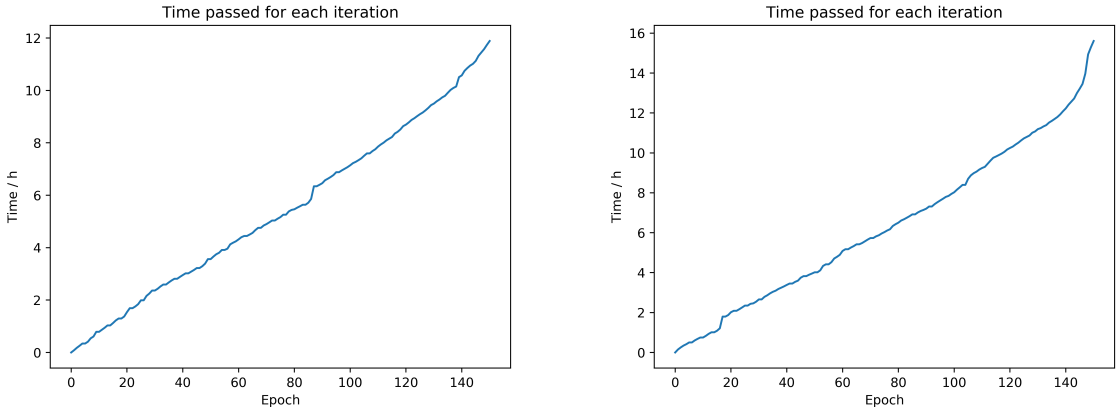

Figure 9: *Total time passed of reduction process after each iteration (epoch) for the old method (left) and new method (right).*

6 Summary and outlook

Combustion physics is and important research field in order to decrease the pollution and increase the efficiency of combustion engines to help minimize the effect they have on the climate. To be able to simulate real life combustion processes with CFD the reaction mechanisms need to be reduced significantly and there are multiple ways of doing this, for example skeletal reduction. To further reduced a mechanism it can be beneficial to separate a reversible reaction into two irreversible ones and analyse them separately. The method in developed in this paper was unsuccessful in determining the reverse reaction rates however possible leads as to why where discovered which will be summarized below.

The evaluation of a reduced mechanism is done by a response score function and this paper showed that having a smooth N-dimensional response landscape can be used to facilitate the reduction process. This paper also showed that it might be possible to use the MAE response function alone to evaluate the accuracy of a reduced mechanism. These are the concluding statements of this paper. Except for the concluding statement also opened up for new possible research directions. In **Sec.5** possible new directions are presented separately and will be summarize shortly in this section.

The secondary objective of the project, involving the development of a method to calculate the reverse reaction rates, did not yield results matching the expectation meaning that there are a few loose ends which further research can be done upon. The main potential direction is to do a thorough analysis and derivation of the equations used in the method based in fundamental physical expressions in order to find if some physical phenomena is unaccounted for or if physical relation have been missed in the method developed.

The primary objective of the project, involving the study of how a smooth response landscape can facilitate the reduction process, did yield results which indicated that the newly implemented method is at least as good, if not slightly better, than the old method. However, due to the few mechanisms analysed the only conclusion possible to be draw from the results are that a smooth response landscape does facilitate the reduction process, but the efficiency cannot be determined conclusively. This opens up for further study where many reduced mechanisms using different response methods are collected

and analysed in order to draw a quantitative conclusion of which response method yields the best reduced mechanisms. In this project only one reduction process per method was performed. However, in order to draw conclusions regarding the efficiency of the method in the reduction process more reductions would need to have been done. This is of course a potential continuation of this project. There is also a possibility to develop new response methods using other types of comparison methods, that yield a smooth response landscape, in order to find a method that require less computational power and hence lowering the time of reduction, for example study the behaviour of MAE more thoroughly and potentially use it to alone evaluate the response as it captures both the shape and value.

References

[1] BP Statistical Review of World Energy, [website], https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html, accessed 20th April 2018

[2] T. Lu and C.K. Law, 'Toward accomodating realistic fuel chemistry in large-scale computations', *Progress in Energy and Combustion Science*, 35:192-215, 2009, DOI: https://doi.org/10.1016/j.pecs.2008.10.002

[3] R. Hilbert et al., 'Impact of detailed chemistry and transport models on turbulent combustion simulations', *Progress in Energy and Combustion Science*, 30:61-117, 2004, DOI: https://doi.org/10.1016/j.pecs.2003.10.001

[4] N. Zettervall, C. Fureby and E.J.K Nilsson, 'Small Skeletal Kinetic Mechanism for Kerosene Combustion', *Energy Fuels*, 30(11):9801-9813, 2016, DOI:https://doi.org/10.1021/acs.energyfuels.6b01664

[5] ANSYS Chemkin Input Manual 17.0 (15151), ANSYS Reaction Design: San Diego, 2016

[6] ANSYS Chemkin Theory Manual 17.0 (15151), Reaction Design: San Diego, 2015

[7] B.J. McBride, S. Gordon and M.A Reno, 'Coefficients for calculating Thermodynamic and Transport Properties of Induvidual Species', *NASA Technical Memorandum 4513*, October 1993.

[8] K.J. Laidler, 'A glossary of terms used in chemical kinetics, including reaction dynamics (IUPAC Recommendations 1996)', *Pure and Applied Chemistry*, 68 (1):149–192, 1996, DOI: https://doi.org/10.1351/pac199668010149

[9] R. DeHoff. 'Thermodynamics in Materials Science', 2nd edn, CRC Press, 2006, ISBN:9780849340659

[10] W. Alsaeedan, M. El Menei and S. Al-Ahmadi, 'A hybrid genetic-ant colony optimization algorithm for the word sense disambiguation problem', *Information Sciences*, 417:20-38, 2017, DOI: https://doi.org/10.1016/j.ins.2017.07.002

[11] R.J. Hyndman and A.B. Koehler, 'Another Look at measures of forecast accuracy', *Internantional Journal of Forecasting*, 22 (4):679-688, 2006, DOI: https://doi.org/10.1016/j.ijforecast.2006.03.001

[12] T.Chai and R.R Draxler, 'Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments agains avoiding RMSE in the literature', *Geosci. Model Dev.*, 7 (3):1247-1250, 2014, DOI: :10.5194/gmd-7-1247-2014

[13] R.N. Bracewell, 'The Fourier Transform And it's Applications', 3rd edn., New York:McGraw-Hill, 1965, ISBN: 0070070156

[14] J. Grossman, M. Grossman and R. Katz, 'The First Systems of Weighted Differential and Integral Calculus', Archimedes Foundation, 1980, ISBN: 0977117014

Appendices

A File formats

A.1 NASA thermodynamic format

```
THERMO
   300.000  1000.000  5000.000
N2                G 8/02N   2    0    0    0G   200.000  6000.00  1000.00      1
 2.95257637E+00 1.39690040E-03-4.92631603E-07 7.86010195E-11-4.60755204E-15    2
-9.23948688E+02 5.87188762E+00 3.53100528E+00-1.23660988E-04-5.02999433E-07    3
 2.43530612E-09-1.40881235E-12-1.04697628E+03 2.96747038E+00 0.00000000E+00    4
H2                TPIS78H   2    0    0    0G   200.000  6000.00  1000.00      1
 2.93286575E+00 8.26608026E-04-1.46402364E-07 1.54100414E-11-6.88804800E-16    2
-8.13065581E+02-1.02432865E+00 2.34433112E+00 7.98052075E-03-1.94781510E-05    3
 2.01572094E-08-7.37611761E-12-9.17935173E+02 6.83010238E-01 0.00000000E+00    4
END
```

Figure 10: *An example look at the NASA thermodynamic data format for different species. This figure show the thermodynamic data for the species N_2 and H_2*

A.2 ACO93 example mechanism - reactions 1-6

```
REACTIONS
H2+O<=>H+OH                                      5.0800E+04   2.67000    6292.000
H2+OH<=>H+H2O                                    4.3800E+13   0.00000    6990.000
CH2OH+O2<=>CH2O+HO2                              2.4100E+14   0.00000    5017.000
CH2OH+HO2<=>CH2O+H2O2                            1.2000E+13   0.00000       0.000
O2+H<=>O+OH                                      1.0400E+14   0.00000   15286.000
H2O2(+M)<=>OH+OH(+M)                             2.0000E+12   0.90000   48749.000
LOW /     2.4900E+24        -2.30000       48749.000/
TROE /     4.3000E-01      1.0000E-30      1.0000E+30 /
H2O/7.650/ CO2/1.600/ N2/1.500/ O2/1.200/ H2O2/7.700/ H2/3.700/ CO/2.800/
HCO+H(+M)<=>CH2O(+M)                                1.0900E+12   0.48000    -260.000
LOW /     1.3500E+24        -2.57000        1425.000/
TROE /     7.8240E-01      2.7100E+02      2.7550E+03      6.5700E+03 /
```

Figure 11: *Shows first 6 reaction in the example ACO92 mechanism used to calculate the reverse reaction rates. The first column of numbers represent the pre-exponential factor A, the second column represents the temperature dependence n and the final column the activation energy E_a.*

B Response landscape and reverse reaction rates

B.1 MAE for low σ_{input}

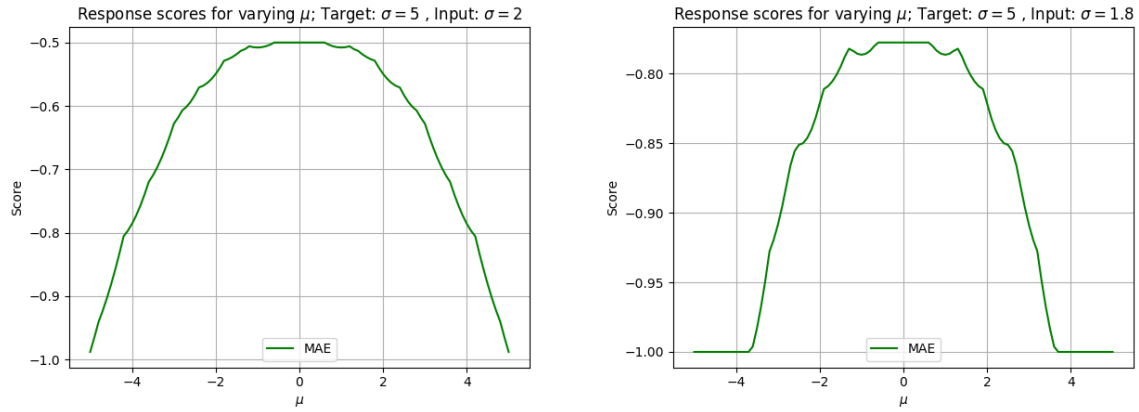

Figure 12: *Shows the response value for MAE when two Gaussians with different σ, target: 5 Input: 2 and 1.8, are compared by sliding one of them over the other.*

B.2 k_r plots for reactions 3-6 in ACO93

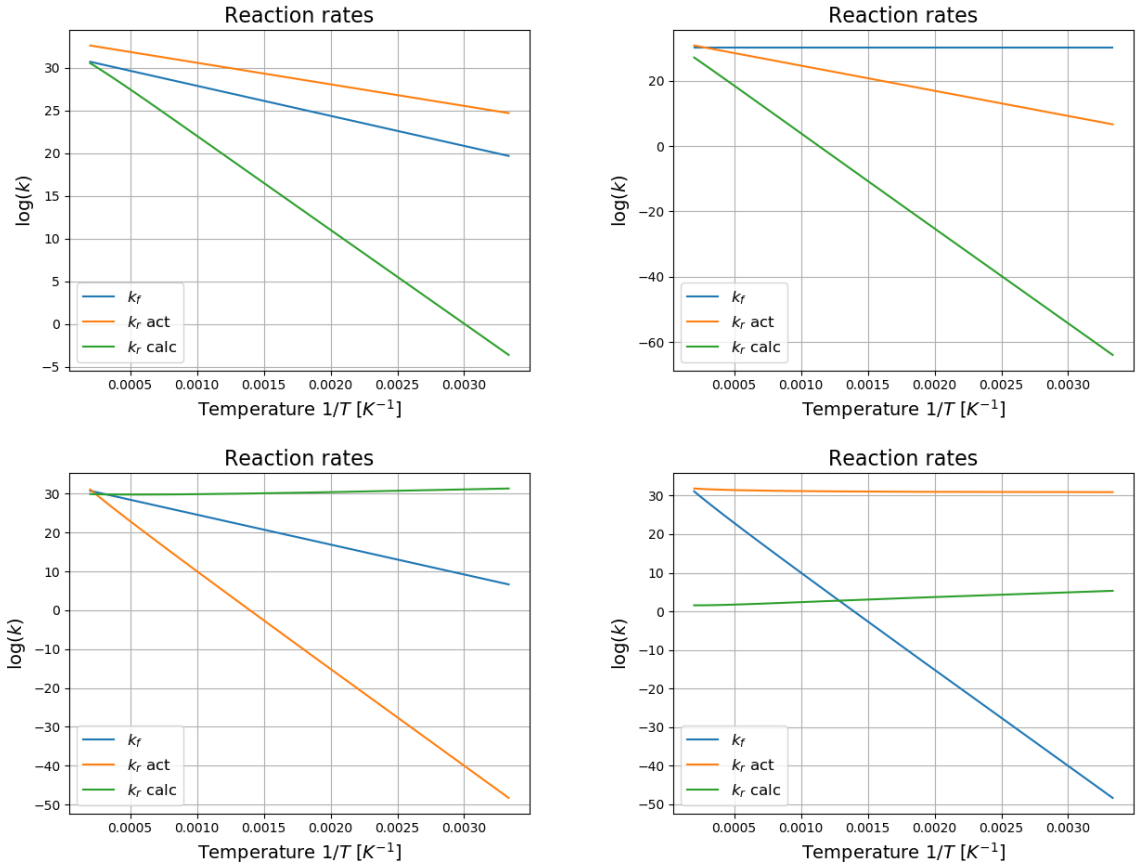

Figure 13: *Shows the forward, actual reverse, and calculated reverse reaction rates, k_f, k_r act, k_r calc, as a function of the inverse temperature, $1/T$. From top left top bottom right, shows the 3rd-6th reaction in ACO92 mechanism*