

**Automated rate equation screening for heterogeneous
catalytic reactions based on experimental data**

Vasco António Correia Saltão

Thesis to obtain the Master of Science Degree in

Chemical Engineering

Supervisor: Prof. Filipe José Da Cunha Monteiro Gama Freire

Supervisor: Dr. Pedro Simão Freitas Mendes

Examination Committee

Chairperson: Prof. Carlos Manuel Faria de Barros Henriques

Supervisor: Dr. Pedro Simão Freitas Mendes

Member of the Committee: Prof. José Manuel Félix Madeira Lopes

December 2020

Acknowledgements

First of all, I would like to thank the Laboratory for Chemical Technology and Ghent University for receiving me very well as a master thesis student. It was a great opportunity and experience (a short one, unfortunately), for which I am grateful.

There are two people in particular that I must thank – Pedro Mendes and Laura Pirro. I am very grateful for the great number of hours you spent on our progress meetings, for the immense feedback and help you provided me when developing this work, for reading several times the chapters of this manuscript and giving me your feedback on them, and for keeping me motivated to continue working. With your coaching, the quality of this work surely improved, and I could not have asked for better coaching. Therefore, thank you very much for your support.

I would also like to thank Prof. Joris Thybaut and Prof. Filipe Gama Freire. Your comments and suggestions on my presentations along this year were very important to guide my work, and your feedback on this manuscript was also very important to improve its quality.

Next, I would like to thank my friends. In these peculiar times, it is very important to be there for each other, and their support surely kept me motivated to continue working through the challenges of this pandemic situation.

Last but not least, I must thank my family. With their support, I was able not only to do my master thesis abroad in Belgium, but also to study in Instituto Superior Técnico. Without them by my side, I would not be where I am today. Thank you very much.

Resumo

Os modelos cinéticos permitem a formulação de novos catalisadores e otimização de sistemas reacionais de forma racional. No entanto, o desenvolvimento tradicional e manual destes modelos cinéticos baseia-se no método de tentativa e erro, tornando-o muito demorado. Também exige ao investigador experiência em cinética e catálise. Estes fatores criam um obstáculo, que pode ser ultrapassado através automatização da construção desses mesmos modelos cinéticos. Assim, este trabalho teve como objetivo o desenvolvimento de uma ferramenta que propõe automaticamente equações de velocidade para reações catalíticas baseando-se em dados experimentais.

Esta ferramenta, desenvolvida em Python, contém uma biblioteca de equações de velocidade, de onde são geradas curvas teóricas de velocidade inicial vs. pressão. As equações constantes na biblioteca foram obtidas a partir dos mecanismos típicos em catálise heterogénea. A ferramenta compara as características destas curvas com as características dos dados experimentais, fazendo uma triagem das equações de velocidade para eliminar as que não podem ser modelos possíveis. Os modelos possíveis são classificados com base na semelhança de características. Esta ferramenta foi testada analisando cinco casos típicos tendo chegado às mesmas equações de velocidade determinadas pelos investigadores, na maioria dos casos. Conclui-se que a ferramenta consegue propor os mesmos modelos cinéticos que um investigador experiente proporia.

Este é um primeiro passo decisivo no desenvolvimento de uma ferramenta de automatização para modelação cinética aplicada a reações de catálise heterogénea que, quando aperfeiçoada, permitirá determinar a cinética de reações catalíticas heterogéneas mais rapidamente e a um maior número de investigadores.

Palavras-chave: Catálise heterogénea, modelação cinética, automatização, dados experimentais, cinética Hougen-Watson

Abstract

To design catalysts and optimize reactive systems rationally, a kinetic model of the catalytic reaction is needed. However, the traditional manual construction of kinetic models has a trial-and-error nature, which makes it very time consuming. Also, it demands from the researcher a high degree of expertise in kinetics and catalysis. These factors create a bottleneck, which can be mitigated by automating the construction of kinetic model. Therefore, the goal of this work was to develop a tool that automatically proposes rate equations for a catalytic reaction based on kinetic experimental.

This tool, which was developed in Python, contains a library of rate equations, from which it generates theoretical initial rate vs. pressure curves. The rate equations in the library were deduced from typical mechanism in heterogeneous catalysis. Afterwards, the tool compares the features of these curves to the features of the experimental data, thus screening the rate equations to eliminate those that cannot be a possible model. Afterwards, the possible models are ranked based on feature similarity. This tool was tested through five different literature case studies, arriving to the same rate equations that the researchers proposed in most cases. It can be concluded that the tool can propose the same kinetic models an experienced researcher would, and in less time. This is an important first step in developing an automation tool for kinetic modelling applied to heterogeneous catalysis that, once fully perfected, will enable a larger number of researchers to achieve kinetic models for catalytic reactions in less time.

Keywords: Heterogeneous catalysis, kinetic modelling, automation, experimental data, Hougen-Watson kinetics.

Contents

Acknowledgements	ii
Resumo	iii
Abstract.....	iv
List of figures	vii
List of tables	x
List of symbols	xii
1. Introduction.....	1
1.1. Heterogeneous catalysis	1
1.2. Kinetic modelling of catalytic reactions.....	1
1.2.1. Reaction mechanism and elementary reaction rates	1
1.2.2. Types of rate equations.....	2
1.2.3. Obtaining a kinetic model from experimental data	3
1.3. Automating kinetic modelling.....	4
1.3.1. Automatic reaction network generation based on theoretical simulations	4
1.3.2. Automatic kinetic modelling based on experimental data	5
1.3.3. Conclusions	6
1.4. Scope of the proposed work.....	6
2. Conceptual design of the automatic rate equation proposal tool	7
2.1. Goal and concept of the tool	7
2.2. Curve features and the feature extraction algorithm	8
3. Theoretical branch of the tool.....	11
3.1. Basis of the library construction	11
3.2. Generic initial rate equations	14
3.3. Theoretical initial rate vs. pressure curves	17
3.4. Conclusion	21
4. Rate equation proposal algorithm and tool development.....	23
4.1. Solving issues with comparison of features	23
4.1.1. Influence of pressure values in extracted features.....	23
4.1.2. Influence of dataset scaling in extracted features	24
4.2. Comparing the theoretical features with the experimental features	25
4.2.1. Tolerance.....	26

4.2.2. Ranking proposed rate equations	26
4.3. Estimating the lumped kinetic constant (k').....	28
4.4. Reducing runtime	28
4.5. Output of the tool.....	28
4.6. Conclusions	32
5. Case studies.....	33
5.1. Oxidation of methane – General test of the tool.....	34
5.1.1. Test results	35
5.1.2. Discussion	37
5.2. Propylene hydrogenation – Influence of data distribution	37
5.2.1. Test results	39
5.2.2. Discussion	41
5.3. Ethanol dehydrogenation – Influence of data variability and noise	42
5.3.1. Test results	44
5.3.2. Discussion	48
5.4. Propylene disproportionation – Influence of repeated tests and noise	49
5.4.1. Test results – Full datasets.....	51
5.4.2. Test results –Averaged datasets.....	53
5.4.3. Discussion	55
5.5. Propyne hydrogenation – Determined rate equation not in the library.....	56
5.5.1. Test results	58
5.5.2. Discussion	59
5.6. Conclusions	59
6. Conclusions and future work	61
References	63
Appendix 1 – Influence of assuming only one surface reaction step	65
Appendix 2 – Deduction of generic initial rate equations	67
Appendix 3 – Datasets used in the case studies	77
Appendix 4 – User manual	79

List of figures

Figure 1 - Conceptual flowchart of the tool (trapezoid – user input; rectangle – process; oval – information included in tool; rounded rectangle – output of tool)	7
Figure 2 - Considered primitives for the feature extraction algorithm[20] (signs of first and second derivatives in brackets, respectively).....	8
Figure 3 - Graphical representation of example of output of the feature extraction algorithm[20].....	8
Figure 4 - Origin of values used to generate theoretical curves.....	18
Figure 5 - Flowchart of theoretical branch of the tool (trapezoid – user input; rectangle – process; oval – information included in tool; rounded rectangle – output of theoretical branch)	21
Figure 6 - Comparison of features extracted from two datasets that follow Eq. 10 for different x-values	23
Figure 7 - Extracted features of the same curve at two different scales	24
Figure 8 - Flowchart of the comparison algorithm (trapezoid – user input; rectangle – process; oval – output of previous branches; rhombus – decision; rounded rectangle – terminator)	25
Figure 9 - Flowchart describing the process of choosing the best curve of each initial rate equation (trapezoid – user input; rectangle – process; oval – output of previous branches; rhombus – decision; rounded rectangle – terminator)	27
Figure 10 - Example of output graph with experimental features.....	29
Figure 11 - Example of output graph with theoretical curve of proposed rate equation	29
Figure 12 - Example of output text string	30
Figure 13 - Full flowchart of the tool (trapezoid – user input; rectangle – process; information included in tool; rounded rectangle – output of tool).....	31
Figure 14 - Graphical representation of the methane oxidation dataset extracted at 320 °C (markers – experimental points; solid line – guideline drawn by researcher)[23]	34
Figure 15 – Graphical representation of the recognized features of the methane oxidation dataset ...	35
Figure 16 - Best curve of the rate equation of mechanism d_1 for the methane oxidation dataset.....	36
Figure 17 - Best curve of the rate equation of mechanism b_1 for the methane oxidation dataset.....	37
Figure 18 - Graphical representation of the propylene hydrogenation dataset[24] (markers – experimental points; vertical bars – error of each point; solid line – model fitting; dashed lines – 95% confidence curves). Reprinted with permission from [24]. Copyright 2004 Elsevier B.V. All rights reserved.....	38
Figure 19 - Graphical representation of the recognized features of the propylene hydrogenation dataset.....	39
Figure 20 - Best curve of the rate equation of mechanism h_1 proposed for the propylene hydrogenation dataset.....	41
Figure 21 - Best curve of the rate equation of mechanism p_1 proposed for the propylene hydrogenation dataset.....	41
Figure 22 - Graphical representation of the recognized features of the p_1 theoretical curve	42

Figure 23 - Experimental datasets of the ethanol dehydrogenation reaction[25] (markers – experimental points; solid line – model fitting). Reprinted with permission from [25]. Copyright 1964 Published by Elsevier Ltd.	43
Figure 24 - Graphical representation of the recognized features of the 225, 235 and 250 °C ethanol dehydrogenation datasets	44
Figure 25 - Graphical representation of the recognized features of the 265 °C ethanol dehydrogenation dataset.....	45
Figure 26 - Best curve of the rate equation mechanism d_1 proposed for the 265 °C ethanol dehydrogenation dataset.....	45
Figure 27 - Graphical representation of the recognized features of the 275 and 285 °C ethanol dehydrogenation datasets	46
Figure 28 - Best curve of the rate equation of mechanism d_1 proposed for the 275 °C ethanol dehydrogenation dataset.....	47
Figure 29 - Best curve of the rate equation of mechanism d_1 proposed for the 285 °C ethanol dehydrogenation dataset.....	48
Figure 30 - Graphical representation of the recognized features of the 225 °C ethanol dehydrogenation dataset with the (0, 0) point added	49
Figure 31 - Experimental datasets of the propylene disproportionation reaction[26] (markers – experimental points; solid line – model fitting). Reprinted with permission from [26]. Copyright 1973 Published by Elsevier Inc.	50
Figure 32 - Graphical representation of the recognized features of the propylene disproportionation datasets	51
Figure 33 - Best curve of the rate equation of mechanisms a_1 and e_1 proposed for the 399 °C propylene disproportionation dataset	52
Figure 34 - Best curve of the rate equation of mechanisms a_1 and e_1 proposed for the 427 °C propylene disproportionation dataset	52
Figure 35 - Best curve of the rate equation of mechanism d_1 proposed for the 454 °C propylene disproportionation dataset	53
Figure 36 - Graphical representation of the recognized features of the averaged propylene disproportionation datasets	53
Figure 37 - Best curve of the rate equation of mechanisms a_1 and e_1 proposed for the averaged 399 °C propylene disproportionation dataset.....	54
Figure 38 - Best curve of the rate equation of mechanism d_1 proposed for the averaged 427 °C propylene disproportionation dataset	55
Figure 39 - Best curve of rate equation of mechanism d_1 proposed for the averaged 399 °C propylene disproportionation dataset.....	55
Figure 40 - Experimental dataset of the propyne hydrogenation reaction[27] (markers – experimental points; vertical bars – error of each point; solid line – model fitting; dashed lines – 95% confidence curves). Reprinted with permission from [27]. Copyright 2007 American Chemical Society.	57

Figure 41 - Graphical representation of the recognized features of the propyne hydrogenation dataset	58
Figure 42 - Best curve of the rate equation of mechanism n_2 proposed for the propyne hydrogenation dataset	59
Figure 43 - Template of input file (with example of data)	79
Figure 44 - Example of output of feature extraction algorithm	80
Figure 45 - Example of output graph with best theoretical curve (left) and output text string (right)	84
Figure 46 - Example of summary of output	84

List of tables

Table 1 - Advantages and disadvantages of methods of obtaining an overall rate equation	3
Table 2 - Example of output of the feature extraction algorithm[20]	9
Table 3 - Mechanisms studied by Yang and Hougen[21] and their respective initial rate equations for reactions with 2 reactants	12
Table 4 – Library of generic initial rate equations	15
Table 5 - Comparison of limitations between the work of Yang and Hougen[21] and the tool	17
Table 6 - Calculation of K_g for mechanisms where the RDS is the desorption of R ('c', 'g', 'j', 'm' and 'p'), for two reactants and one product	19
Table 7 - Lists of possible values for the equilibrium constants	20
Table 8 - Motivations behind each case study	33
Table 9 - Recognized features of the methane oxidation dataset	35
Table 10 - Proposed rate equations for the methane oxidation dataset (in order of ascending values of ranking criteria), for each tolerance value (mechanisms with _1: A= CH ₄ , B=O ₂ ; mechanisms with _2: A=O ₂ , B=CH ₄)	35
Table 11 - Constants and ranking criteria for the best curve of the top 5 proposed rate equations for the propylene oxidation dataset (mechanisms with _1: A=CH ₄ , B=O ₂ ; mechanisms with _2: A=O ₂ , B=CH ₄)	36
Table 12 - Constants for the propylene hydrogenation dataset determined by Brandão et al.[24]	38
Table 13 - Recognized features of the propylene hydrogenation dataset	39
Table 14 - Proposed rate equations for the propylene hydrogenation dataset (in order ascending values of ranking criteria), for each tolerance value (mechanisms with _1: A=H ₂ , B=C ₃ H ₆ ; mechanisms with _2: A=C ₃ H ₆ , B=H ₂)	39
Table 15 - Constants and ranking criteria for the best curve of the top 5 proposed rate equations for the propylene oxidation dataset (mechanisms with _1: A=H ₂ , B=C ₃ H ₆ ; mechanisms with _2: A=C ₃ H ₆ , B=H ₂)	40
Table 16 - Kinetic and equilibrium constants for the datasets of the ethanol dehydrogenation reaction determined by Franckaerts and Froment[25]	43
Table 17 - Recognized features of the 225, 235 and 250 °C ethanol dehydrogenation datasets	44
Table 18 - Recognized features of the 265 °C ethanol dehydrogenation dataset	45
Table 19 - Constants and ranking criteria for the best curve of the proposed rate equation for the 265 °C ethanol dehydrogenation dataset (A=EtOH)	45
Table 20 - Recognized features of the 275 and 285 °C ethanol dehydrogenation datasets	46
Table 21 - Proposed rate equations for the 275 and 285 °C ethanol dehydrogenation datasets, for each tolerance value	47
Table 22 - Constants and ranking criteria for the best curve of the proposed rate equation for the 275 °C ethanol dehydrogenation dataset (A=EtOH)	47
Table 23 - Constants and ranking criteria for the best curve of the proposed rate equation for the 285 °C ethanol dehydrogenation dataset (A=EtOH)	47

Table 24 - Kinetic and equilibrium constants for the datasets of the propylene disproportionation reaction determined by Luckner and Wills [26].	50
Table 25 - Recognized features of the propylene disproportionation datasets	51
Table 26 - Constants and ranking criteria for the best curve of each proposed rate equation for the 399 °C propylene disproportionation dataset (A=C ₃ H ₆)	51
Table 27 - Constants and ranking criteria for the best curve of each proposed rate equation for the 427 °C propylene disproportionation dataset (A=C ₃ H ₆)	52
Table 28 - Constants and ranking criteria for the best curve of each proposed rate equation for the 454 °C propylene disproportionation dataset (A=C ₃ H ₆)	52
Table 29 - Recognized features of the averaged propylene disproportionation datasets	54
Table 30 - Constants and ranking criteria for the best curve of each proposed rate equation for the averaged 399 °C propylene disproportionation dataset (A=C ₃ H ₆)	54
Table 31 - Constants and ranking criteria for the best curve of each proposed rate equation for the averaged 427 °C propylene disproportionation dataset (A=C ₃ H ₆)	54
Table 32 - Kinetic and equilibrium constants for the dataset of the propyne hydrogenation reaction determined by Fritsch et al.[27].	57
Table 33 - Recognized features of the propyne hydrogenation dataset	58
Table 34 - Constants and ranking criteria for the best curve of the top 5 proposed rate equation for the propyne hydrogenation dataset (mechanisms with _1: A=H ₂ , B=C ₃ H ₄ ; mechanisms with _2: A=C ₃ H ₄ , B=H ₂)	58
Table 35 - Experimental dataset of the oxidation of methane case study	77
Table 36 - Experimental dataset of the propylene hydrogenation case study	77
Table 37 - Experimental datasets of the ethanol dehydrogenation case study	77
Table 38 - Full experimental datasets of the propylene disproportionation case study	78
Table 39 - Averaged experimental datasets of the propylene disproportionation case study	78
Table 40 - Experimental dataset of the propyne hydrogenation case study	78

List of symbols

Symbol	Meaning
a, b, r, s	Stoichiometric coefficients of species A, B, R and S, respectively
$\alpha, \beta, \rho, \sigma$	Order of reaction of species A, B, R, and S in the elementary reaction step, respectively
δ	Relative error
k_1	Kinetic constant of elementary reaction step
k'	Lumped kinetic constant
K_r	Equilibrium constant of surface reaction step
K_{Eq}	Equilibrium constant
K_g	Global equilibrium constant
MSE	Mean squared error
MSE_{norm}	Mean squared error of normalized data
n	Exponent of rate equation denominator
N	Number of points in dataset
\hat{r}_0	Normalized initial rate value
SRE	Sum of relative errors
y_i	Molar fraction of reactant i

Symbol	Meaning	Units
i	Concentration of species i (i = A, B, R, S)	concentration units
C_R, C_P	Concentration of reactants and products, respectively	concentration units
K_i	Adsorption constant of species i (i = A, B, R, S)	pressure units ⁻¹
P	Pressure	pressure units
P_{Exp}, P_{Theor}	Pressure values of extremes of experimental data and theoretical curve, respectively	pressure units
P_i	Partial pressure of species i	pressure units
r	Rate of reaction	rate units
r_0	Initial rate of reaction	rate units
r_1	Rate of elementary reaction step	rate units

1. Introduction

1.1. Heterogeneous catalysis

A catalyst is a substance added to a chemical reaction that increases its rate without modifying the overall standard Gibbs energy change of the reaction[1], while the substance itself is not consumed[2]. Catalysts are essential to the sustainability of several industries since they offer a more energetically favourable pathway for chemical reactions to occur[3]. Therefore, the research for new catalyst is of major importance.

In heterogeneous catalysis, the physical state of the catalyst is different from the one of the reactants and products. Very often the reactants and products are in a gaseous state or liquid state while the catalyst is a solid, porous or not. For these cases, the transformation of reactants into products takes place in the surface of the referred catalyst, be it inside the porous structures (when they exist) or not[4]. For that to happen, the reactant molecules must migrate through diffusion from the bulk of the fluid to the surface of the catalyst, adsorb onto the active sites present on surface and convert into products. Then these products must desorb from the active sites and migrate through diffusion to the bulk of the fluid. While the diffusion of species is a physical step, the adsorption and the reaction step itself are chemical. Due to the different nature of these elementary steps, each will be influenced by different factors in the reaction system. The migration of species depends on the mass velocity of the fluid, the size of the particles, the diffusional characteristics of the fluid and, for porous catalysts, on the porosity of the catalyst, the dimension of the pores and the interconnectivity between them. On the other hand, the adsorption and surface reaction steps depend on the character and extent of the catalyst surface and on the respective activation energy of the steps[4]. Due to these different influences, each elementary step will occur at a different rate. This factor will influence the rate of the overall reaction.

1.2. Kinetic modelling of catalytic reactions

To perform a catalytical reaction, one must design the whole process first, so it is as optimized as possible. For this, it is necessary to manipulate the rate of the reaction. However, that requires some knowledge on how the rate varies with reaction conditions since, as mentioned before, each of the elementary steps occurs at a different rate due to the different factors that influence them.

A kinetic model is a mathematical model that describes how the rate of a given reaction is affected by the reaction conditions[5], [6]. With such a model, a process involving said reaction can be optimized by controlling its rate[7]. Therefore, kinetic modelling is very important to the development of catalysts[8] and to the sustainable processes[9].

1.2.1. Reaction mechanism and elementary reaction rates

As mentioned before, a reaction of heterogeneous catalysis occurs through some elementary reaction steps, namely the adsorption and desorption of the reactants and products, respectively, on the active sites, and the surface reaction itself. These elementary steps combined form the reaction mechanism.

An example of a catalytical reaction is present in Eq. 1.



Assuming that all the species adsorb onto the surface of the catalyst and that the surface reaction takes place in a single step, the mechanism of this reaction is the one present as Eq. 2.



X represents an active site of the surface of the catalyst and AX and BX represent the intermediate species of A and B adsorbed to said active site, respectively.

From the mechanism in Eq. 2 results a set of 5 rate equations, one for each elementary step. These equations follow Guldberg and Waage's mass action law[10], present in Eq. 3.

$$r_1 = k_1 \cdot \prod_{\text{Reactants}} C_R^r - k_{-1} \cdot \prod_{\text{Products}} C_P^p \quad \text{Eq. 3}$$

In this equation, C_R and C_P are the concentrations of each reactant and product, respectively, and r and p are the stoichiometric numbers of each reactant and product, respectively. The parameters k_1 and k_{-1} are the kinetic constants of the forward and reverse reactions, respectively.

1.2.2. Types of rate equations

There are different ways to achieve a rate equation for a global reaction. The simplest is to simply fit a power law to data of the reaction rate. Following this method, the rate equation for the example reaction of Eq. 1 would be the one present as Eq. 4.

$$r = k_1 \cdot [A]^\alpha \cdot [B]^\beta - k_{-1} \cdot [R]^\rho \cdot [S]^\sigma \quad \text{Eq. 4}$$

For a multistep reaction, Eq. 4 is typically used with an empirical approach, by fitting its parameters to a set of experimental data. This empirical approach has no basis on any kinetic or catalytical concepts, ignoring the physical and chemical background. This can potentially result in models which are unrealistic or have very limited application[5]. A good kinetic model for multistep reactions must be based on its mechanism.

By applying Eq. 3 to each of the elementary steps of the mechanism (Eq. 2) a rate equation for each of the steps will be obtained. A mass balance for each of the species involved (including the intermediates) can then be written as a function of each of the elementary reaction rates. These mass balances will depend on the type of reactor (PFR, CSTR or batch)[11].

To achieve a kinetic model for the reaction the mass balances must be solved. Three different methods to solve the mass balances are enumerated by de Oliveira et al.[5]: the complete solution, the quasi-steady state (QSS) approximation and the rate-determining step (RDS) approximation. The complete solution is achieved by not making any assumptions regarding the elementary steps when solving the mass balances of all. This method, while being the most accurate, normally does not result in an analytical solution, requiring a numerical method to solve the balances. This makes it very time

consuming. On the other hand, in the QSS approximation the problem is simplified by assuming that the intermediates of the reaction have a null net rate of production, i.e., the production of these species is as fast as its consumption. This way, the concentration of these species can be obtained as a function of the concentrations of the reactants and products, as well as a function of the rate parameters. The advantages of this method are that it requires less kinetic parameters to achieve the overall reaction rate than the complete solution method, and that through this method an analytical solution is obtained for most cases. On the other hand, it cannot account for time dependent phenomena. A simpler method is the RDS approximation, which is based on the assumption that the slowest step of the mechanism determines the rate of the overall reaction, with all the other steps being fast enough to be considered at quasi-equilibrium, having no influence on the overall rate. Again, the concentrations of the intermediates can be obtained as a function of the concentrations of the reactants and products and of the rate parameters. The overall rate equation will be the same as the rate equation of the RDS. After writing the concentration of the intermediates as a function of reactants and products, the overall rate equation will depend on these concentrations, as well as the kinetic constant of the RDS and the equilibrium constants of the other steps of the mechanism. This means that even fewer kinetic parameters are necessary. However, this further simplification comes with the additional limitation of not accounting for a shift of the RDS due to altering reaction conditions. This method is the basis of Hougen-Watson[4] type of rate equations. A summary of the advantages and disadvantages of the possible methods described is present in Table 1.

Table 1 - Advantages and disadvantages of methods of obtaining an overall rate equation

Method	Advantages	Disadvantages
Complete solution	Solution accounts for time and reaction condition dependent phenomena	Very time consuming, no analytical solution
QSS approximation	Less kinetic parameters are required, analytical solution is often obtainable	Solution cannot account for time depending phenomena
RDS approximation	Even less kinetic parameters are required, analytical solution is always obtainable	Solution cannot account for time and reaction condition dependent phenomena
Power law	Simplest method	Usually not suited for multistep reactions

1.2.3. Obtaining a kinetic model from experimental data

The first step in kinetic modelling is to propose a mechanism for the reaction. Traditionally, this is done by extracting kinetic data of the reaction in question from experiments and then do a qualitative analysis of said data. This kinetic datasets usually contain a small number of points[12]. The analysis is done by an experienced researcher, that draws a curve that follows the trends of the experimental data, observes the features of said curve and, based on kinetic and catalytic knowledge, proposes the rate equations that are possible explanations for the reaction behaviour. Afterwards, the kinetic parameters of each

proposed equation are determined. This is usually done by obtaining experimental data from the reaction and fitting the rate equation to the data. Then the proposed rate equations are ranked accordingly to how well they fit the experimental data, and the best one is chosen. This model, however, might not be valid for reaction conditions outside of the ones for which the experimental data was obtained, so the reliability and robustness of the resulting best model must be studied. More recently, the kinetic parameters have been estimated on the basis of microkinetic analysis[5], through methods based on quantum chemical calculations, such as DFT methods[13]. These methods use knowledge from surface science, computational chemistry and kinetic studies to gain insight on the intermediate species and to determine the value of some kinetic parameters[5].

The manual extraction of kinetic information from the data requires expertise on catalysis, since the researcher performing it needs to recognize the features of the data, while keeping in mind which possible features are chemically intuitive and which are not, and also needs to propose a rate equation based on the recognized features. This means that the number of researchers that can perform research on kinetic modelling is reduced. In addition, this is a trial-and-error process, since the researcher will propose a new rate equation until reaching one that can explain the data in a satisfactory way. This makes this process very time consuming. Also, for a given catalytic reaction, the volume of available kinetic datasets is small[12]. These factors create a bottleneck in the kinetic modelling process.

1.3. Automating kinetic modelling

By automating the traditional kinetic modelling process utilising a software tool, the bottleneck mentioned above would be in part mitigated. Since the process is being done by a software tool, the trial and error process would be performed much faster. Also, since the process is automated, the level of expertise on kinetics and catalysis the researcher is required to have is reduced, which means that more researchers are able to work on this field. Therefore, a search in the literature for past work about the automatization of the kinetic modelling process for catalytic reactions was performed, to better understand what progress has been achieved on the subject.

1.3.1. Automatic reaction network generation based on theoretical simulations

To generate a reaction network, a software must be developed in such a way that it can represent chemical species uniquely and unambiguously and link these species with each other through elementary steps[14]. There are two large groups of methods to generate reaction networks automatically from simulations: linear algebraic methods and graph theory based methods[15]. In the former, a matrix is built from the known elementary steps and the species, and through basis transformations a reaction network is achieved. In the latter, the species are represented by nodes and the bonds that connect the nodes represent the elementary steps. In this method, all possible pathways are created, and then all that are non-essential are eliminated. In both methods, the elementary steps are an input to the algorithm.

Some strategies used with these methods are representing the molecules of the species using graph theory[16], [17], and a lumping strategy[5]. This latter one involves the agglomeration of species in families based on their similarities and of the reactions that involve said species. This helps in reducing

the complexity of the system but leads to a loss of detail. An example of a tool that uses these strategies is the one developed by Gaoa et al[14].

Another way to generate a reaction network is through machine learning. However, the tools that implement this method create models without any theoretical fundamentals, outputting only what it was trained to by the training data, which is obtained from different plants and different operators, having much uncertainty associated with it[13]. Also, machine learning tools require an high volume of data to be trained, and the volume of kinetic data of catalytic reactions is small[12].

Automating the kinetic modelling process through theoretical simulations has its advantages, such as greatly reducing the need for performing catalytic experiments to obtain new experimental data, and in some cases eliminating the need of experimental data. However, it also has its flaws, with the main one being that the theory in which the simulations are based on is built on some assumptions. On some cases, these assumptions can lead to incorrect models that are unrealistic. Even if these algorithms use experimental data as a comparison to the generated models or even to estimate its kinetic parameters, the proposal of the model is purely based on these theoretical simulations.

1.3.2. Automatic kinetic modelling based on experimental data

Schaich, King and Becker[18], [19] have translated the rate equation proposal from recognized features into a tool, called TAM-C. This tool generates a curve that fits the experimental data of evolution of heat, recognizes the “episodes” (features) of said curve, picks the models that are compatible with the system from its library of models and then eliminates the models whose curve features are not qualitatively similar to those of the data. This greatly reduces the number of possible models to be quantitatively compared by means of regression and reduces the time spent on such comparison, by eliminating the models that would never fit adequately to the experimental data. However, the type of data analysed by TAM-C consists of timeseries of calorific data, which is different from the data necessary to propose a kinetic model, especially when it comes to the much larger number of points of timeseries when compared to kinetic data[12].

After searching the literature, no tool that automates the kinetic modelling process for catalytic reactions based on experimental data was found. However, some previous work useful to the development of such a tool was found.

The concept of screening models from recognized data features applied on the tool TAM-C[18], [19] can also be applied on this hypothetical tool. However, this means that a way to extract features from kinetic datasets of catalytic reactions is required, as well as a library of models for catalytic reactions.

Siradze[20] developed an algorithm capable of automatically recognizing the features of a given set of data points. This tool generates a curve which follows the data trends and is chemically realistic and then analyses it qualitatively, extracting its features, recognizing the maximums, minimums, plateaus, increasing and decreasing trends present on the data. This way, the algorithm mimics the work done by a researcher when recognizing the features of an experimental dataset. This algorithm can be useful for identifying the features of a set of experimental data, so a rate equation can be proposed based on them.

Yang and Hougen[21] studied the correlation between the evolution of initial rate vs. total pressure curves and the rate equations and reaction mechanisms that could describe that evolution. The reaction mechanisms studied in this paper are common ones in heterogeneous catalysis and cover all possibilities, within certain assumptions. The studied rate equations were all derived from the mechanisms through the RDS assumption. This work can be useful to develop the library of models of the previously mentioned hypothetical tool, since the mechanisms studied are applicable to most cases of heterogeneous catalysis.

1.3.3. Conclusions

From searching the literature, it has been determined that most researchers working on the automation of kinetic modelling have done so recurring to computer simulations based on theory, rather than experimental data (although some of these developed software use experimental data later in its algorithm). These developed softwares must generate a reaction network and determine the kinetic and thermodynamic parameters of the resulting model.

As mentioned before, automating the process of kinetic modelling through theoretical simulations has the advantage of reducing or even eliminating the need to obtain experimental data, while having the disadvantage of potentially proposing kinetic models which are unrealistic. On the other hand, the automatic proposal of models based on experimental data results in propose models that are more realistic, since it was based on real data, even if some assumptions are used. The disadvantage is that the performance of such a tool is dependent on the quality of the data itself (number of points, distribution of points, experimental noise, etc.).

While no tool that automates the kinetic modelling process for catalytic reactions based on experimental was found, some sources valuable for the development of such a tool were found.

1.4. Scope of the proposed work

The objective of this works is to develop a tool that automatically proposes initial rate equations that describe a set of initial rate vs total pressure experimental data. The development of this tool will contribute to the development of catalysts since it will mitigate in part the bottleneck present in the manual process of kinetic modelling by automating it. Similarly to the tool TAM-C[18], [19], this tool will perform a screening of the models in its library based on a comparison of the curve features between the experimental data and the curves theoretically predicted by the models in the library. The curve features will be extracted by the algorithm developed by Siradze[20], while the library of initial rate equations will be constructed based on the work developed by Yang and Hougen[21]. Once developed, this tool will be tested with some case studies to see how it performs with real experimental data and to determine its strengths and limitations.

2. Conceptual design of the automatic rate equation proposal tool

In this chapter the main concepts important to the automatic rate equation proposal tool will be discussed. The goal of the tool and the concepts behind it will be discussed. The concept of curve features will be introduced, and the feature extraction algorithm developed by Siradze[20], which is an integral part of this tool, will also be introduced.

2.1. Goal and concept of the tool

The goal of this work is to develop a tool that automatically proposes rate equations that can describe an experimental kinetic dataset. The concept behind this is that different initial rate equations produce curves with different features (i.e. trends), and the tool will propose the equations that result in curves with features similar enough to the features of the kinetic data. This proposal should be based on the elimination of rate equations that cannot generate curves with features similar to the features of the data.

Figure 1 contains a conceptual flowchart of the tool that describes the steps taken by the tool to arrive at the proposed rate equations.

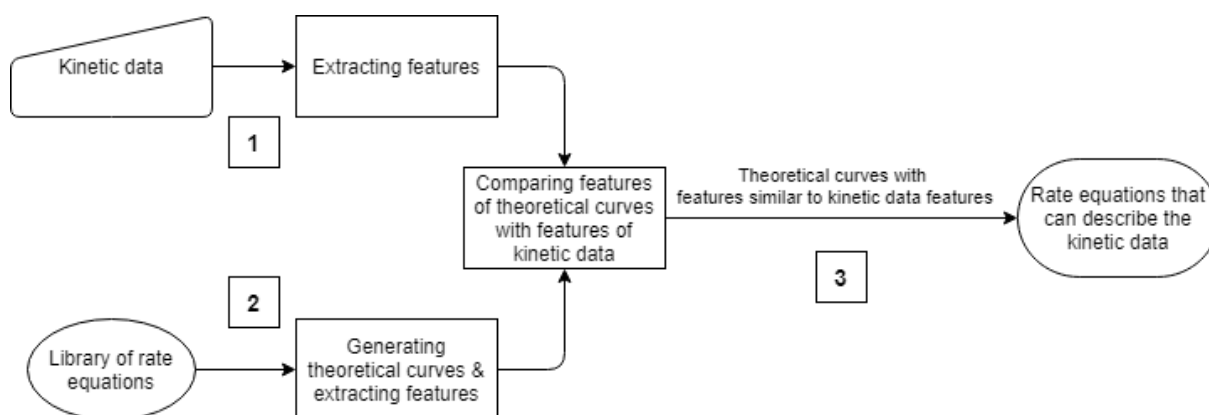


Figure 1 - Conceptual flowchart of the tool (trapezoid – user input; rectangle – process; oval – information included in tool; rounded rectangle – output of tool)

This process can be divided in three parts. The experimental branch (1) is where the features of the experimental kinetic data are extracted. In the theoretical branch (2), the theoretical curves are generated from a library of rate equations and their features are extracted. The third part is the comparison section(3), where the features of the theoretical curves are compared to the feature of the experimental data and the elimination of initial rate equations that cannot describe the data is done. To achieve this tool, several of its components must be developed. A library of rate equations needs to be developed, as well as the algorithms that generate the theoretical curves from said library. This will be covered in Chapter 3. The algorithms that compare the features and eliminate the rate equations that cannot describe the experimental data also need to be developed. This subject will be covered in Chapter 4. However, since the theoretical curves are compared to the experimental data on basis of their features, this means that in both the experimental and the theoretical branches there needs to be an algorithm that can extract the features of the curves in a way that they can be compared.

2.2. Curve features and the feature extraction algorithm

When analysing a set of experimental data, the trends of the data can contain useful information to the researcher. For this reason, a qualitative analysis of the data is useful to extract valuable information. Janusz and Venkatasubramanian[22] studied a way to perform this qualitative analysis. The first step is to draw a curve that follows the data. Afterwards, this curve is split in several intervals differentiated from each other by the signs of their first and second derivatives, denominated primitives. This way it can be determined if the experimental data has an increasing or decreasing behaviour, and if this behaviour is increasing or decreasing. The kinetic information can then be determined from the set of primitives of the data and from the x-values and y-values of the borders of the primitives.

This approach was chosen in the development of the feature extraction algorithm by Siradze[20], which performs this qualitative analysis automatically. The algorithm draws the trend following curve automatically by dividing the data in various sections along the x-axis and fitting splines to each section. The resulting curve is then split into primitives. The possible primitives are present in Figure 2.


						
A(+,-)	B(-,+)	C(+,+)	D(-,-)	E(+,0)	F(-,0)	G(0,+/-)

Figure 2 - Considered primitives for the feature extraction algorithm[20] (signs of first and second derivatives in brackets, respectively)

However, the curve initially generated by the code might not be a realistic one. Therefore, the algorithm developed by Siradze[20] also checks if the resulting sequence of primitives represents a chemically realistic curve. If this is not the case, a new curve will be redrawn. This will repeat until a curve that is chemically realistic is obtained. Afterwards, this curve is simplified as much as possible. This way, the behaviour of a researcher with expertise in the field can be mimicked.

Once the generated curve is fully simplified, the features of the dataset are extracted. They are the primitives, the x-values at the extremes and the y-values at the extremes. An example of the output of this algorithm is present in Table 2 and graphically represented in Figure 3.

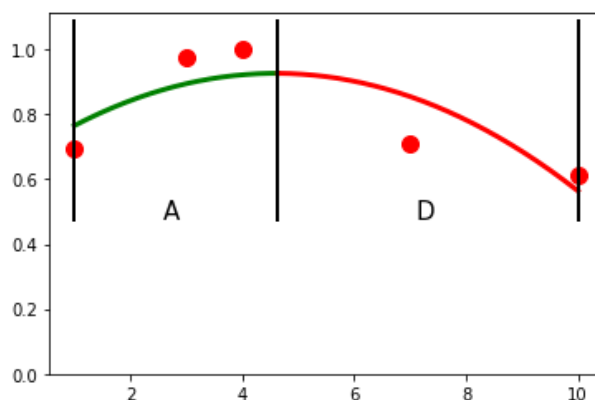


Figure 3 - Graphical representation of example of output of the feature extraction algorithm[20]

Table 2 - Example of output of the feature extraction algorithm[20]

Primitives	A, D		
x-values at extremes	1.00	4.64	10.0
y-values at extremes	1.08	1.31	0.80

This algorithm was developed to receive any kind of dataset as input. However, in the tool considered in the present thesis, the only datasets it will analyse are ones of kinetic data. This algorithm will be used to extract the features of the experimental dataset and of the theoretical curves, for them to later be compared. Important to note as well is the fact that in the tool developed in this thesis, for reasons discussed in Chapter 4, the datasets analysed by the feature extraction algorithm are normalized previously to said analysis. This makes the y-values at the extremes of the primitives void of important kinetic information, thus being discarded by the tool. Therefore, from this point of the manuscript onwards, any mention of features is referring only to the respective primitives and x-values at the extremes.

3. Theoretical branch of the tool

As mentioned before, the tool is composed by an experimental branch, which analyses the experimental data, and a theoretical branch, which creates the theoretical curves corresponding to the initial rate equations in the tool's library. The goal of this chapter is to explain how the theoretical branch was developed, what it consists of and what are its limitations. The correlation between the effect of total pressure on initial rates of reaction and the initial rate equation of said reaction will be discussed in this chapter. This subject was studied in the paper by Yang and Hougen[21], which was an important source for the development of the library. However, their work has its limitations regarding its applicability, which are also discussed in this chapter, as well as what was done to overcome them and increase the applicability of the tool to a wider range of reaction conditions. Also discussed in this chapter is the process of generating the theoretical curves from the library.

3.1. Basis of the library construction

As stated before, the goal of this work is to develop a tool that can propose rate equations that can describe an experimental kinetic dataset. Therefore, it is necessary for the tool to must contain a library of rate equations. From this library, theoretical curves will be generated to be compared with the kinetic data to determine which equations from the library can describe the data. Therefore, the library must be constructed in a way that makes the tool applicable to most catalytic reactions.

In 1950, Yang and Hougen[21] published their work in which they studied the effects of total pressure on the initial rates of catalysed gaseous reactions with different mechanisms. On their work, they established the correlation between the shape of initial rate vs. pressure curves and the possible rate equations that could result in such curves. These rate equations were deduced from their respective mechanisms for 4 types of reactions, varying on the number of reactants and species involved: 1 reactant and 1 product (Eq. 5), 1 reactant and 2 products (Eq. 6) , 2 reactants and 1 product (Eq. 7) and 2 reactants and 2 products (Eq. 8).



These four types of reaction were chosen because, as Yang and Hougen stated, "Complex chemical processes may be considered as consisting chiefly of a combination of separate bimolecular and unimolecular reactions"[21], and these four equations account for the possibilities of the existence of 1 or 2 reactants and 1 or 2 products.

The rate equations were deduced from the mechanisms utilizing the RDS approximation. The mechanisms are present in Table 3, as are the rate equations deduced for 2 reactants. These are valid for reactions of both the type of Eq. 7 and Eq. 8 (whenever a rate equation is different for these two types it is explicitly stated). The mechanisms of Table 3 are characterized by the adsorption of the species involved in the reaction, as well as its RDS.

Table 3 - Mechanisms studied by Yang and Hougen[21] and their respective initial rate equations for reactions with 2 reactants

Adsorption of species	RDS	Initial rate equation
A – All species adsorb molecularly	a – Adsorption of A controlling	$r_0 = \frac{k' \cdot y_A \cdot P}{1 + K_B y_B \cdot P}$
	b – Adsorption of B controlling	$r_0 = \frac{k' \cdot y_B \cdot P}{1 + K_A y_A \cdot P}$
	c – Desorption of R controlling	<p><u>Only 1 product (Eq. 7):</u></p> $r_0 = \frac{k' \cdot K_g \cdot y_A \cdot y_B \cdot P^2}{1 + K_R \cdot K_g \cdot y_A \cdot y_B \cdot P^2 + (K_A y_A + K_B y_B) \cdot P}$ <p><u>More than 1 product (Eq. 8):</u></p> $r_0 = \frac{k'}{K_R}$
	d – Surface reaction controlling	$r_0 = \frac{k' \cdot K_A y_A \cdot K_B y_B \cdot P^2}{(1 + (K_A y_A + K_B y_B) \cdot P)^2}$
B – A adsorbs dissociatively, all other species adsorb molecularly	e – Adsorption of A controlling	$r_0 = \frac{k' \cdot y_A \cdot P}{(1 + K_B y_B \cdot P)^2}$
	f – Adsorption of B controlling	$r_0 = \frac{k' \cdot y_B \cdot P}{1 + \sqrt{K_A y_A \cdot P}}$
	g – Desorption of R controlling	<p><u>Only 1 product (Eq. 7):</u></p> $r_0 = \frac{k' \cdot K_g \cdot y_A \cdot y_B \cdot P^2}{1 + K_R \cdot K_g \cdot y_A \cdot y_B \cdot P^2 + \sqrt{K_A y_A \cdot P} + K_B y_B \cdot P}$ <p><u>More than 1 product (Eq. 8):</u></p> $r_0 = \frac{k'}{K_R}$
	h – Surface reaction controlling	$r_0 = \frac{k' \cdot K_A y_A \cdot K_B y_B \cdot P^2}{(1 + \sqrt{K_A y_A \cdot P} + K_B y_B \cdot P)^3}$
C – B does not adsorb, all other species adsorb molecularly	i – Adsorption of A controlling	$r_0 = k' \cdot y_A \cdot P$
	j – Desorption of R controlling	<p><u>Only 1 product (Eq. 7):</u></p> $r_0 = \frac{k' \cdot K_g \cdot y_A \cdot y_B \cdot P^2}{1 + K_R \cdot K_g \cdot y_A \cdot y_B \cdot P^2 + K_A y_A \cdot P}$ <p><u>More than 1 product (Eq. 8):</u></p> $r_0 = \frac{k'}{K_R}$

Table 3 – Mechanisms studied by Yang and Hougen[21] and their respective initial rate equations for reactions with 2 reactants (continued)

Adsorption of species	RDS	Initial rate equation
C – B does not adsorb, all other species adsorb molecularly	k – Surface reaction controlling	$r_0 = \frac{k' \cdot K_A y_A \cdot y_B \cdot P^2}{(1 + K_A y_A \cdot P)^n}$ n = 1 for one product (Eq. 7), 2 for two products (Eq. 8)
D - A adsorbs dissociatively, B does not adsorb, all other species adsorb molecularly	l – Adsorption of A controlling	$r_0 = k' \cdot y_A \cdot P$
	m – Desorption of R controlling	<u>Only 1 product (Eq. 7):</u> $r_0 = \frac{k' \cdot K_g \cdot y_A \cdot y_B \cdot P^2}{1 + K_R \cdot K_g \cdot y_A \cdot y_B \cdot P^2 + \sqrt{K_A y_A \cdot P}}$ <u>More than 1 product (Eq. 8):</u> $r_0 = \frac{k'}{K_R}$
	n – Surface reaction controlling	$r_0 = \frac{k' \cdot K_A y_A \cdot y_B \cdot P^2}{(1 + \sqrt{K_A y_A \cdot P})^2}$
E – A does not adsorb, all other species adsorb molecularly	o – Impact of A controlling	$r_0 = \frac{k' \cdot K_B y_B \cdot y_A \cdot P^2}{1 + K_B y_B \cdot P}$
	p – Desorption of R controlling	<u>Only 1 product (Eq. 7):</u> $r_0 = \frac{k' \cdot K_g \cdot y_A \cdot y_B \cdot P^2}{1 + K_A \cdot K_B y_B \cdot y_A P^2 + K_R \cdot K_g \cdot y_A \cdot y_B \cdot P^2 + K_B y_B \cdot P}$ <u>More than 1 product (Eq. 8):</u> $r_0 = \frac{k'}{K_R}$
	q – Adsorption of B controlling	$r_0 = k' \cdot y_B \cdot P$
F – Uncatalyzed reaction	r – Uncatalyzed reaction	$r_0 = k' \cdot y_A \cdot y_B \cdot P^2$

To evaluate how useful this work was to the development of the library, an analysis of its assumptions and limitations was made. As discussed in Chapter 1, when deducing a rate equation from a mechanism, the RDS approximation, an analytical solution is guaranteed. However, this solution does not account for phenomena dependent on time and temperature. The assumption of the existence of a RDS is valid due to the fact that the adsorption steps and the surface reaction step of a given catalytic reaction mechanism typically involve large activation energies, which makes these steps very sensitive to temperature, which in turn means that the rates of these elementary steps will have widely different orders of magnitude[4]. Therefore, it is very likely for one of these steps to be significantly slower than the others, even if the slow step changes for different temperatures, making the RDS assumption valid

for most catalytic reactions. This combined with the fact that an analytical solution is always guaranteed makes the RDS approximation suitable for the rate equations in the library.

Another assumption made by Yang and Hougen was that there is only one reaction step on all mechanisms. This is a valid assumption to make if the RDS of the mechanism in question is the adsorption or desorption of one of the species. However, if the RDS is one of the surface reactions, then the resulting initial rate equation will be different than the one which results from assuming that all the surface reactions happen in one step (this is demonstrated in Appendix 1). However, this assumption is still satisfactory since it is impossible to determine from experimental data how many and which surface reactions take place. It is, however, one important limitation on the applicability of this work.

As stated before, Yang and Hougen studied the effect of total pressure on the initial rate values. The choice to deal with initial rates was made because by doing so the rate equations are simplified, since the reverse reaction is disregarded. This in turn makes it easier to identify the possible mechanisms of the reaction by analysing its initial rate vs. pressure curve. However, to assume initial rates, there must be no products in the reactor feed, which was the assumption made by Yang and Hougen. This assumption is satisfactory, since most experimental reactions take place without any products being fed to the reactor.

The previously mentioned assumptions taken by Yang and Hougen on their work are all satisfactory, which makes their work very useful to the development of the library of the tool, since it contains the equations for several possible reaction mechanisms in heterogeneous catalysis. However, this work still has some assumption that limit its applicability. Analysing Eq. 5 to Eq. 8 it can be observed that the possibility of more than one molecule of each species being involved in the reaction is not taken into account. Besides this, the assumption that the feed is composed solely by the reactants, without any inert species, and that these reactants are present in stoichiometric proportions is also made. This means that the possibility of inert species being present in the feed is not taken into account, with the same happening for the possibility of one of the reactants being proportionally present in excess in the feed. For these reasons, this work could not be directly used in the construction of the tool since these possibilities are not covered. In fact, the great majority of catalytical experiment is done using inert species in the reaction feed and/or with one of the reactants in exceeding its stoichiometric proportion. Also, it is very limiting to assume reactions where all species involved have a stoichiometric coefficient of 1.

3.2. Generic initial rate equations

As mentioned before, the library of rate equations must contain equations that makes the tool applicable to most catalytic reactions. The work developed by Yang and Hougen is a useful source for the development of this library since it contains the equations for several possible reaction mechanisms in heterogeneous catalysis. However, it cannot be directly implemented into the development of the library due to its applicability limitations regarding the stoichiometry and feed composition. Therefore, to develop the library from this work it is necessary to overcome these limitations. This is done by deducing generic initial rate equations for the mechanisms in Table 3, with generic stoichiometric coefficients for the involved species and generic molar fractions of reactants. These generic initial rate equations are

present in Table 4, derived for a reaction with 2 reactants (the number of products is irrelevant except when explicitly stated). To arrive to the equations for only 1 reactant all that is necessary is for the parameters relative to reactant B (y_B and b) to be 0. The deductions for these initial rate equations are present in Appendix 2.

Table 4 – Library of generic initial rate equations

ID	Description	Generic initial rate equation
a	All species adsorb molecularly, adsorption of A controlling	$r_0 = \frac{k' \cdot y_A \cdot P}{1 + K_B y_B \cdot P}$
b	All species adsorb molecularly, adsorption of B controlling	$r_0 = \frac{k' \cdot y_B \cdot P}{1 + K_A y_A \cdot P}$
c	All species adsorb molecularly, desorption of R controlling	<p><u>Only 1 product (R):</u></p> $r_0 = \frac{k' \cdot \sqrt[r]{K_g \cdot y_A^a \cdot y_B^b} \cdot P^{\frac{a+b}{r}}}{1 + K_R \cdot \sqrt[r]{K_g \cdot y_A^a \cdot y_B^b} \cdot P^{\frac{a+b}{r}} + (K_A y_A + K_B y_B) \cdot P}$ <p><u>More than 1 product:</u></p> $r_0 = \frac{k'}{K_R}$
d	All species adsorb molecularly, surface reaction controlling	$r_0 = \frac{k' \cdot (K_A y_A)^a \cdot (K_B y_B)^b \cdot P^{a+b}}{(1 + (K_A y_A + K_B y_B) \cdot P)^n}$ $n = \max(a + b, r + s)$
e	A adsorbs dissociatively, adsorption of A controlling	$r_0 = \frac{k' \cdot y_A \cdot P}{(1 + K_B y_B \cdot P)^2}$
f	A adsorbs dissociatively, adsorption of B controlling	$r_0 = \frac{k' \cdot y_B \cdot P}{1 + \sqrt{K_A y_A \cdot P}}$
g	A adsorbs dissociatively, desorption of R controlling	<p><u>Only 1 product (R):</u></p> $r_0 = \frac{k' \cdot \sqrt[r]{K_g \cdot y_A^a \cdot y_B^b} \cdot P^{\frac{a+b}{r}}}{1 + K_R \cdot \sqrt[r]{K_g \cdot y_A^a \cdot y_B^b} \cdot P^{\frac{a+b}{r}} + \sqrt{K_A y_A \cdot P} + K_B y_B \cdot P}$ <p><u>More than 1 product:</u></p> $r_0 = \frac{k'}{K_R}$
h	A adsorbs dissociatively, surface reaction controlling	$r_0 = \frac{k' \cdot (K_A y_A)^a \cdot (K_B y_B)^b \cdot P^{a+b}}{(1 + \sqrt{K_A y_A \cdot P} + K_B y_B \cdot P)^n}$ $n = \max(2a + b, r + s)$
i	B does not adsorb, adsorption of A controlling	$r_0 = k' \cdot y_A \cdot P$

Table 4 - Library of generic initial rate equations (continued)

ID	Description	Generic initial rate equation
j	B does not adsorb, desorption of R controlling	<p><u>Only 1 product (R):</u></p> $r_0 = \frac{k' \cdot \sqrt[r]{K_g \cdot y_A^a \cdot y_B^b} \cdot P^{\frac{a+b}{r}}}{1 + K_R \cdot \sqrt[r]{K_g \cdot y_A^a \cdot y_B^b} \cdot P^{\frac{a+b}{r}} + K_{Ay_A} \cdot P}$ <p><u>More than 1 product:</u></p> $r_0 = \frac{k'}{K_R}$
k	B does not adsorb, surface reaction controlling	$r_0 = \frac{k' \cdot (K_{Ay_A})^a \cdot y_B^b \cdot P^{a+b}}{(1 + K_{Ay_A} \cdot P)^n}$ $n = \max(a, r + s)$
l	A adsorbs dissociatively, B does not adsorb, adsorption of A controlling	$r_0 = k' \cdot y_A \cdot P$
m	A adsorbs dissociatively, B does not adsorb, desorption of R controlling	<p><u>Only 1 product (R):</u></p> $r_0 = \frac{k' \cdot \sqrt[r]{K_g \cdot y_A^a \cdot y_B^b} \cdot P^{\frac{a+b}{r}}}{1 + K_R \cdot \sqrt[r]{K_g \cdot y_A^a \cdot y_B^b} \cdot P^{\frac{a+b}{r}} + \sqrt{K_{Ay_A} \cdot P}}$ <p><u>More than 1 product:</u></p> $r_0 = \frac{k'}{K_R}$
n	A adsorbs dissociatively, B does not adsorb, surface reaction controlling	$r_0 = \frac{k' \cdot (K_{Ay_A})^a \cdot y_B^b \cdot P^{a+b}}{(1 + \sqrt{K_{Ay_A} \cdot P})^n}$ $n = \max(2a, r + s)$
o	A does not adsorb, impact of A controlling	$r_0 = \frac{k' \cdot K_B y_B \cdot y_A^{\frac{a}{b}} \cdot P^{1+\frac{a}{b}}}{1 + K_B y_B \cdot P}$
p	A does not adsorb, desorption of R controlling	<p><u>Only 1 product (R):</u></p> $r_0 = \frac{k' \cdot \sqrt[r]{K_g \cdot y_A^a \cdot y_B^b} \cdot P^{\frac{a+b}{r}}}{1 + K_A \cdot K_B y_B \cdot \sqrt[y_A^a]{P^{1+\frac{a}{b}}} + K_R \cdot \sqrt[r]{K_g \cdot y_A^a \cdot y_B^b} \cdot P^{\frac{a+b}{r}} + K_B y_B \cdot P}$ <p><u>More than 1 product:</u></p> $r_0 = \frac{k'}{K_R}$
q	A does not adsorb, adsorption of B controlling	$r_0 = k' \cdot y_B \cdot P$
r	Uncatalyzed reaction	$r_0 = k' \cdot y_A^a \cdot y_B^b \cdot P^{a+b}$

In this library, each mechanism is represented by an ID and is associated to a description of itself and to its resulting initial rate equation. The stoichiometric coefficients of the species A, B, R and S are

present in these equations as parameters a , b , r and s , respectively, and the molar fraction of each reactant in y_A and y_B . Regarding the descriptions of the mechanism, it is important to note that if the manner of adsorption of a species is not specified, then it means that that species adsorbs molecularly.

This way, the limitations of the work done by Yang and Hougen[21] regarding the stoichiometric coefficients and the molar fractions of the reactants were surpassed. Others, however still remain. In Table 5 there is a comparison between the limitations of their work and the limitations of the tool.

Table 5 - Comparison of limitations between the work of Yang and Hougen[21] and the tool

K. H. Yang and O. A. Hougen paper[21]	Tool
Initial rate of reaction only – No products in feed, conversion of reactants negligible	Initial rate of reaction only – No products in feed, conversion of reactants negligible
Reactants in stoichiometric proportions, no inert species in the feed	Works for any molar fractions of reactants and inert species
All mechanisms in library assume the existence of one RDS	All mechanisms in library assume the existence of one RDS
All mechanisms in library assume the existence of only one surface reaction step	All mechanisms in library assume the existence of only one surface reaction step
All mechanisms in library assume products adsorb molecularly	All mechanisms in library assume products adsorb molecularly
Maximum of 2 reactants and 2 products	Maximum of 2 reactants and 2 products
Stoichiometric coefficients of all species always 1	Works for any stoichiometry
Adsorption only on 1 type of active site	Adsorption only on 1 type of active site

3.3. Theoretical initial rate vs. pressure curves

With the library of generic initial rate equations, the theoretical initial rate vs. pressure curves of each mechanism can be created and its features extracted. To generate the theoretical curves from the generic rate equations, it is necessary to calculate the parameters and exponents of the equation. These are calculated using the stoichiometric coefficients of the species and the molar fractions of the reactants, which are both inputted by the user, and the equilibrium constant values, which come from a list of possible values for these constants. To obtain a curve it is also necessary to give some pressure values for which the initial rate values are calculated. This is schematically represented in Figure 4.

$$r_0 = \frac{k' \cdot (K_A y_A)^a \cdot (K_B y_B)^b \cdot P^{a+b}}{(1 + (K_A y_A + K_B y_B) \cdot P)^n}$$

Figure 4 - Origin of values used to generate theoretical curves

All the generic initial rate equations of the library fit the type of Eq. 9.

$$r_0 = \frac{c_1 \cdot P^{e_1}}{(1 + c_2 \cdot \sqrt{P} + c_3 \cdot P + c_4 \cdot P^{e_2} + c_5 \cdot P^{e_3})^n} \quad \text{Eq. 9}$$

As mentioned before, to generate the theoretical curves, the parameters c_1 , c_2 , c_3 , c_4 , c_5 , e_1 , e_2 , e_3 and n are calculated using the reactant fractions, the stoichiometric coefficients and the equilibrium constants. Since k' will only scale the resulting curve, it has no influence in the shape of the curve. Therefore, there is no need to permutate between values for this constant. Also, the resulting curves will be normalized (i.e., all the initial rate values will be divided by its maximum) for reasons related to the feature extraction that will be explained in detail in Chapter 4, so testing different values for k' is not necessary. The way the parameters are calculated from these values is different for each generic rate equation. Therefore, the formulas used depend on the descriptive tags associated with the generic rate equation in question (see Table 4).

The first step in the theoretical branch is to generate, for each generic initial rate equation, a specific equation for the reaction in question. For this, the stoichiometric coefficients of the species are inserted in the generic initial rate equations. After this step, it is possible that the exponent of the denominator (parameter n in the equations of Table 4) takes a value larger than 3. Since the number of this parameter is the same as the number of active sites involved in the RDS, having it be larger than three is chemically unrealistic[6], since that would mean that four or more adjacent active sites would need to be available, which is very unlikely. Therefore, to prevent the proposal of such equations by the tool, any mechanism which produces an equation with a denominator exponent larger than 3 is automatically eliminated from the list of possibilities.

After this, the initial rate vs. pressure theoretical curves can be generated, using the pressure values, the molar fractions of the reactants and the values of the equilibrium constants. For reasons discussed in Chapter 4 related to the comparison of theoretical and experimental features, the pressure values used to generate the curves are the same that are present in the experimental data. As mentioned before, the values of the equilibrium constants (K_{Eq}) are selected from a list of possible values. The reason for this is that these constants can have any value within a chemically realistic range, and

obviously not all values can be tested. So, the approach chosen was to generate several different curves for each equation by permutating between some possible values for these constants.

The chosen values must cover an adequate range and be distributed in way that ensures that the curves generated from an initial rate equation have among them as many different shapes that can result from said equation as possible. For the adsorption constants, the values used to generate the curves are 0.1, 0.3, 0.5, 0.7, 1, 10 and 100 (pressure units⁻¹). These values were chosen by analysing some examples of catalytic reactions[2] and observing the typical range of orders of greatness for these values. However, as it can be observed, there are more values in the range between 0.1 and 1 than between 1 and 100. This is because the shape of a curve depends on the values present on its denominator, where these constants appear, and this shape is more influenced by a variation of small values than for the same variation on larger values. Therefore, to ensure that as many different curve shapes as possible are generated, more values in this range of small values were included. For the equilibrium constant of the surface reaction step K_r a different list was used, with more small values. This is because it was assumed that if a species meaningfully adsorbs, then its adsorption constant will have a non-negligible value. However, the range of values for K_r can reach values not considered for the adsorption constants. Therefore, the values used for K_r are 1E-7, 0.01, 0.1, 0.3, 0.5, 0.7, 1, 10 and 100 (units dependant on equation).

Analysing Table 4, it can be observed that for most of the initial rate equations the only equilibrium constants present are related to the adsorption of one or both reactants. The exceptions to this are the equations of the mechanisms where the RDS is the desorption of a product, when only one product results from the equation. Besides containing the adsorption constant K_R , they also contain the global equilibrium constant K_g , which is calculated from the equilibrium constants of every step of the mechanism. The formulas to calculate K_g for these mechanisms are present in Table 6.

Table 6 - Calculation of K_g for mechanisms where the RDS is the desorption of R ('c', 'g', 'j', 'm' and 'p'), for two reactants and one product

Mechanism	K_g
'c': All species adsorb molecularly, desorption of R controlling	$\frac{K_r \cdot K_A^a \cdot K_B^b}{K_R^r}$
'g': A adsorbs dissociatively, desorption of R controlling	$\frac{K_r \cdot K_A^a}{K_R^r}$
'j': B does not adsorb, desorption of R controlling	$\frac{K_r \cdot K_A^a}{K_R^r}$
'm': A adsorbs dissociatively, B does not adsorb, desorption of R controlling	$\frac{K_r \cdot (K_A \cdot K_B)^b}{K_R^r}$
'p': A does not adsorb, desorption of R controlling	$\frac{K_r \cdot (K_A \cdot K_B)^b}{K_R^r}$

Each formula uses the equilibrium constants of adsorption (or impact, in the case of K_A in mechanism 'p') of all the species that adsorb, as well as K_r . When generating the theoretical curves, there will be permutations of the possible K_{Eq} values for each of these constants. Therefore, when there is only one product, the number of curves generated for these mechanisms will be higher than for the other mechanisms. This is problematic for reactions with two reactants since this drastically increases the

runtime of the tool. For this reason, whenever the reaction involves two reactants and one product, the tool offers the user the possibility of using reduced versions of the previously mentioned lists of possible K_{Eq} values to reduce the runtime of the tool. These reduced lists contain the values 0.1, 0.5, 1, 10 and 100 (pressure units⁻¹) for the adsorption constants list and 1E-7, 0.01, 0.1, 0.5, 1, 10 and 100 (units dependant on equation) for the K_r list. This brings the total number of lists to 4. These lists are present in Table 7.

Table 7 - Lists of possible values for the equilibrium constants

List	Values								
K_i (pressure units⁻¹): full list	0.1	0.3	0.5	0.7	1	10	100		
K_i (pressure units⁻¹): reduced list	0.1	0.5	1	10	100				
K_r: full list	1E-7	0.01	0.1	0.3	0.5	0.7	1	10	100
K_r: reduced list	1E-7	0.01	0.1	0.5	1	10	100		

In the end, the features of all the generated theoretical curves are extracted, using the algorithm developed by Siradze[20], and stored in a way that makes them traceable to the rate equation from which they were generated.

Also important to note is that most of the initial rate equations in Table 4, specifically the ones from mechanism 'e' to 'q', are "non-symmetrical" regarding the reactants, i.e., the reactants adsorb differently in these mechanisms, which means that the initial rate equations will change if reactants A and B are swapped. For this reason, when dealing with reactions with two reactants, the algorithms of the theoretical branch will be run twice, swapping the order of the reactants in the second run. This way, all possibilities are covered. To differentiate between the theoretical curves generated on the first run and those generated on the second run, '_1' is added to the mechanism IDs of the former, while '_2' is added to the IDs of the latter ('_1' is still added to the IDs when dealing with reactions with only 1 reactant). A practical example of the importance of this is present in the case study of Subchapter 5.2.

The process of the theoretical branch of the tool is summarized in Figure 5.

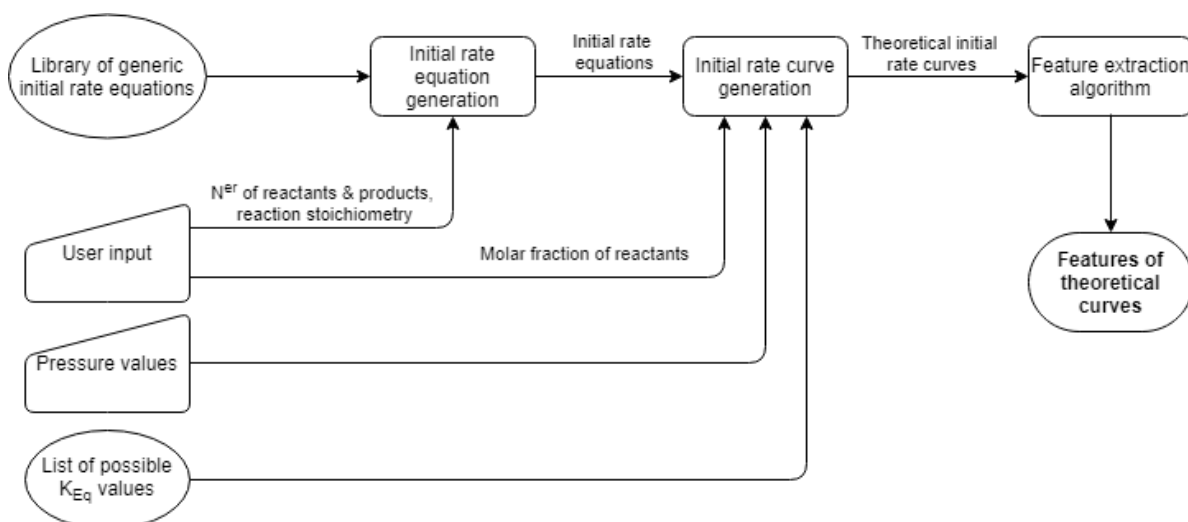


Figure 5 - Flowchart of theoretical branch of the tool (trapezoid – user input; rectangle – process; oval – information included in tool; rounded rectangle – output of theoretical branch)

3.4. Conclusion

To propose rate equations that can describe kinetic data, the tool must contain a library of rate equations to compare to said data. The wider the range of experimental conditions that the library is applicable to, the more useful the tool is.

The paper published by Yang and Hougen[21] was very useful to the development of the library, since it studied several common mechanisms in heterogeneous catalysis. However, its assumptions limit greatly the applicability of this work. Therefore, a generic initial rate equation was deduced for each of the studied mechanisms, with the whole set of equations forming the library. This eliminated some of the limitations, but several still remain in the tool, as seen in Table 5. From those limitations, it can be concluded that the tool is mainly suitable to analyse datasets of initial rate vs. pressure. Besides this, the reactions that the tool can analyse must have a maximum of two reactants and two products. While this is a limitation, most catalytic reactions follow this limitation[6]. Also, the initial rate equations proposed by the tool will always assume the existence of one RDS, as well as only one surface reaction step and that all adsorbing species adsorb competitively to the same type of active sites. This limits the applicability of the tool, since there are catalytic mechanisms that fall outside these assumptions.

From the generic initial rate equations in the library the theoretical curves are generated, using the stoichiometric coefficients of the species, the molar fractions of the reactants, the pressure values and the equilibrium constants. While the values of the stoichiometric coefficients, the molar fractions and the pressure values are inputted by the user, the equilibrium constants can have any value inside the chemically realistic values for these parameters. Therefore, the chosen approach was to select some values from this range and generate several curves for each generic initial rate equation by exchanging these values for each equilibrium constant. Due to the limited number of K_{Eq} values in these lists, it is still theoretically possible that the equation that best describes the data does not generate a curve with features similar to those of the experimental data. However, this is unlikely, since the values were chosen in a way that ensures that the typical range of these parameters is covered and that as many different curve shapes as possible are generated.

4. Rate equation proposal algorithm and tool development

The main topic of this chapter is the comparison between the experimental data and the generated theoretical curves, which has the goal of determining which initial rate equations can describe the experimental dataset. The first discussed subject are some issues related to how the feature extraction algorithm works that can affect the recognized features of a given dataset, and how that creates some issues when comparing the experimental features with the theoretical ones, and what has been done to resolve these issues. The decisions that can be taken to overcome these issues and their practical implementation in the code are also discussed. Then the comparison itself is discussed. It is discussed how the tool eliminates the curves that do not follow the experimental data trends and how it chooses the best curve of each initial rate equations that was determined as a possible model for the experimental data. Then it is discussed the final development and decisions implemented in the tool, specifically how the initial guess of kinetic constant k' is estimated for each of these curves, how the runtime of the tool can be reduced and how the proposed initial rate equations and their respective best theoretical curves are presented in the output of the tool.

4.1. Solving issues with comparison of features

Due to the nature of the feature extraction algorithm developed by Siradze[20], there are some issues related to the comparison of features that arise from utilizing this algorithm. These issues are the influence on the extracted features of the distribution of points along the x-axis and of the scaling of the dataset. To overcome these issues, some decisions can be made. These decisions are all related to the experimental and theoretical branches of the tool, which were already discussed from a more theoretical perspective.

4.1.1. Influence of pressure values in extracted features

As stated in Chapter 2, the feature extraction algorithm draws a curve that follows the trend of a given dataset and then extracts the features of said curve. Therefore, the distribution of the points along the x-axis will influence the shape of the drawn curve and consequently the extracted features. In Figure 6 there is an example of two datasets that follow Eq. 10 but each for a different set of x-values. Both sets of x-values are in the range of 1 to 10.

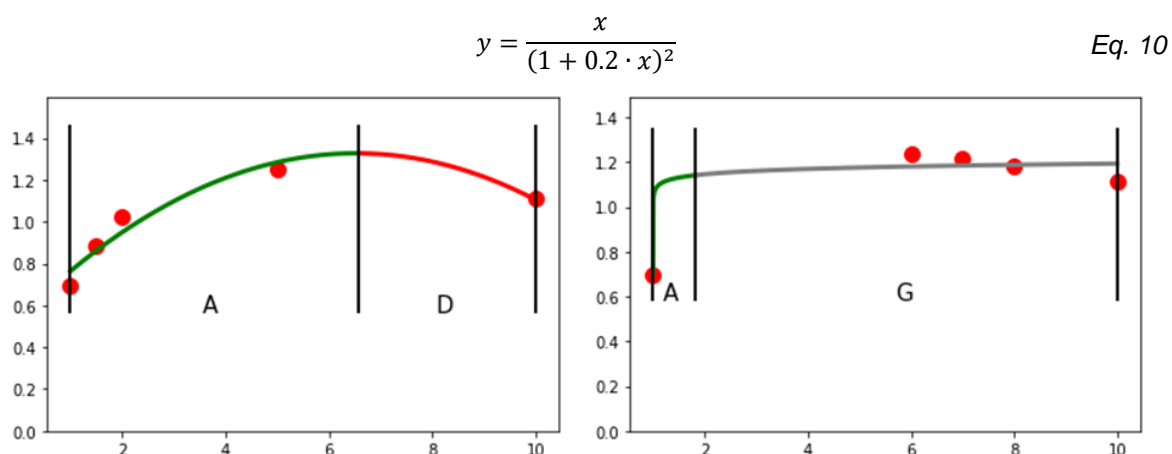


Figure 6 - Comparison of features extracted from two datasets that follow Eq. 10 for different x-values

As it can be seen, the recognized features are different for each dataset, even though they follow the same equation. The features will be as different as the x-values of the datasets are different, with the example in Figure 6 being an edge case since besides the first and last points, the points cover different sections of the x-values range. However, even if the x-values are not as different as the example in Figure 6, the recognized features could still be different, especially the extremes. This could be problematic for the performance of the tool: if the pressure points in the experimental data are different from the ones used to generate a theoretical curve, then the features extracted from both datasets could be different, even if the datasets come from very similar initial rate equations. Therefore, as mentioned before in Chapter 3, the pressure values used to generate the theoretical curves are the ones present in the experimental dataset. This way, the distribution of points along the x-axis in the theoretical curves is the same as in the experimental dataset, and this issue is avoided all together.

4.1.2. Influence of dataset scaling in extracted features

During the development of the tool, it has been observed that, on some occasions, the feature extraction algorithm identifies different features from the same set of data, depending on the scaling applied to this data. This was observed to happen mostly when the points had small y-values. An example of this is present in Figure 7, where both datasets follow Eq. 11 for the same x-values, varying only k, which does not affect the shape of the curve, scaling it only.

$$y = \frac{k \cdot x}{(1 + 2.75 \cdot x)^2} \tag{Eq. 11}$$

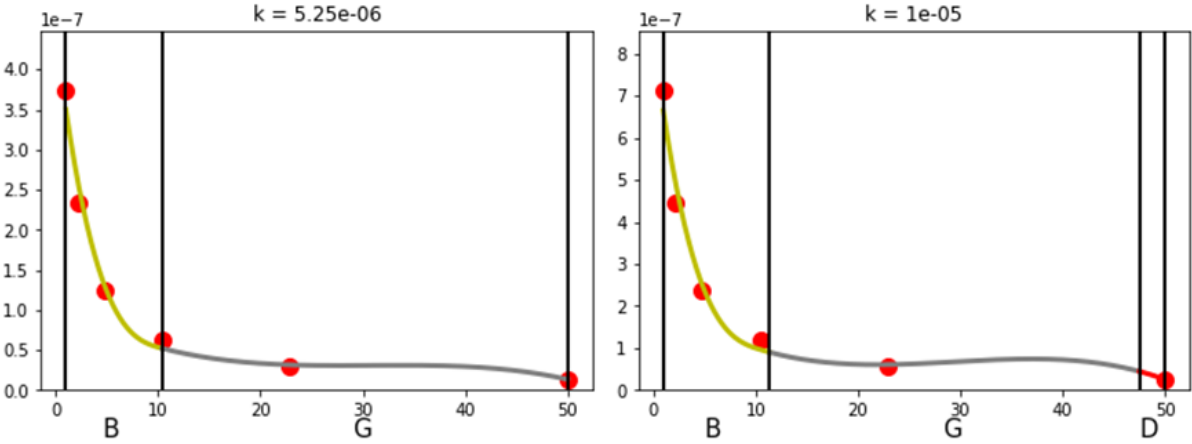


Figure 7 - Extracted features of the same curve at two different scales

As it can be seen, the recognized primitives are not the same for both cases, even though they theoretically should be, since only the scale of the curve was changed. This discrepancy exists because this algorithm is based on numerical methods, and therefore the values of the points themselves could influence the output of the algorithm. This is an issue because if the experimental data and the theoretical curve that best matches it are scaled differently, then the extracted features could be different, and the theoretical curve would not be recognized as having features similar to those of the experimental data.

The chosen approach to avoid this issue was to ensure that both the experimental data and the theoretical curves are at the same scale when their features are extracted. Therefore, the experimental dataset and all of the theoretical curves were normalized before having their features extracted. The normalization consists of dividing all the initial rate values by their maximum, according to Eq. 12.

$$\hat{r}_0(i) = \frac{r_0(i)}{\max(r_0)} \quad \text{Eq. 12}$$

This way, the normalized values will be within the range from 0 to 1. By normalizing the both the experimental data and the generated theoretical curves before extracting the features, it is ensured that both are on the same scale. Also, since all of the generic initial rate equations contain k' multiplying in the numerator (see Table 4), by normalizing the initial rate values the effect of this parameter is annulled, which means that there is no need to permutate between different values for k' when generating the theoretical curves (which, as mentioned before, does not influence the shape of the curve, only the scale), greatly reducing the runtime of the tool.

4.2. Comparing the theoretical features with the experimental features

With the theoretical curves generated and their features extracted, it is now possible to compare these features to the ones extracted from the experimental data to determine which initial rate equations cannot describe the experimental dataset. The comparison algorithm is present in Figure 8.

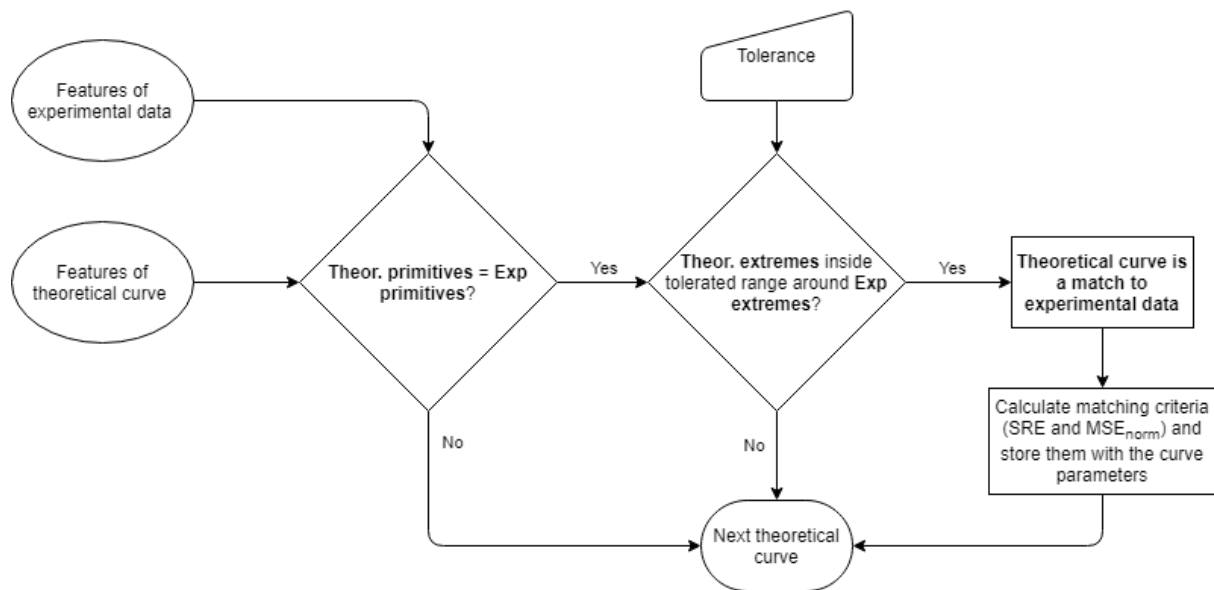


Figure 8 - Flowchart of the comparison algorithm (trapezoid – user input; rectangle – process; oval – output of previous branches; rhombus – decision; rounded rectangle – terminator)

For all curves generated from an initial rate equation, the first type of features to be compared are the primitives, since they indicate more clearly the trends of the experimental data. If the set of primitives of a given theoretical curve is not the same as the set of primitives of the experimental data, then this curve fails the test and is discarded, while the next curve generated from the same equation is tested. If both sets of primitives are the same, then the theoretical curve passes the test and continues to the next one.

4.2.1. Tolerance

After comparing the primitives, the following step is comparing the pressure values of the extremes of the primitives, to see if the trends of the theoretical curve are taking place at similar pressure values as the trends in the experimental data. For that, the relative error of each extreme of the theoretical curve in comparison with the corresponding value of the experimental data, according to Eq. 13.

$$\delta = \frac{|P_{Theor} - P_{Exp}|}{P_{Exp}} \quad Eq. 13$$

In this equation, P_{Theor} is the pressure value of the theoretical extreme and P_{Exp} is the pressure value of the equivalent extreme from the experimental data. After calculating this relative deviation, the tool checks if it does not exceed the tolerance chosen by the user. It was decided that the tolerance should be a user input so the user can decide how close the curves need to be to be considered to have features similar to those of the experimental data, as this also depends on the experimental error in the data. It is important to note that after the tool presents the output, the user can choose to run the comparison algorithm again for a different tolerance.

The relative error is calculated for every extreme, and the theoretical curve only passes this test if all relative errors are inferior to the tolerance. But, as stated before, the theoretical curves are generated using the same pressure values from the experimental dataset. Since the first and last extreme extracted from a dataset are always the smallest and largest pressure value in the dataset, respectively, this means that the first and last extremes of both the theoretical curve and the experimental data will be the same. Therefore, there is no need to test them. This also means that if the set of primitives extracted from the experimental data consists of only one primitive, then a theoretical curve with the same recognized primitive will automatically pass the test of extremes comparison, since there are no valid extremes to compare. The effect of the tolerance value chosen is studied in the case studies present in Chapter 5.

4.2.2. Ranking proposed rate equations

If the curve passes the primitives and extremes tests, then that means that it follows the trends of the experimental data at an acceptable level, which means that the initial rate equation that generated this curve is considered as a possible model for describing the experimental data. However, other theoretical curves may also pass the tests, so a ranking of all these curves must be made. Therefore, there need to be some ranking criteria that differentiates between the curves that pass these tests based on how well they follow the trends of the experimental data.

When comparing theoretical curves, the focus should be on how well they follow the trends of the experimental data, since the way the tool compares the theoretical curves to the experimental data is through the curve features, namely the primitives and the pressure values at the extremes. Since the curves that pass the tests all have the same primitives as the experimental data, the main ranking criterion should be based on the pressure values at the extremes. For these reasons, the principal ranking criteria is defined to be the sum of relative errors (SRE) of all the relative errors of the extremes previously calculated.

For the eventuality of two curves having the same SRE value, a secondary criterion is necessary. This criterion is the mean squared error of the normalized values (MSE_{norm}), which is calculated according to Eq. 14.

$$MSE_{norm} = \frac{\sum_{i=1}^N (\hat{r}_{0,Theor}(i) - \hat{r}_{0,Exp}(i))^2}{N} \quad \text{Eq. 14}$$

In this equation, $\hat{r}_{0,Theor}(i)$ and $\hat{r}_{0,Exp}(i)$ are the normalized initial rate values of the theoretical curve and the experimental data, respectively, and N is the total number of data points. This criterion is used to measure the goodness of fit of the theoretical curves to the experimental data. However, this is not the focus of the comparison between different curves, this serves only to differentiate between curves with the same SRE.

The comparison algorithm described in Figure 8 is repeated for every curve generated by each initial rate equation. Then, the best curve of each equation must be found. This process is described in Figure 9.

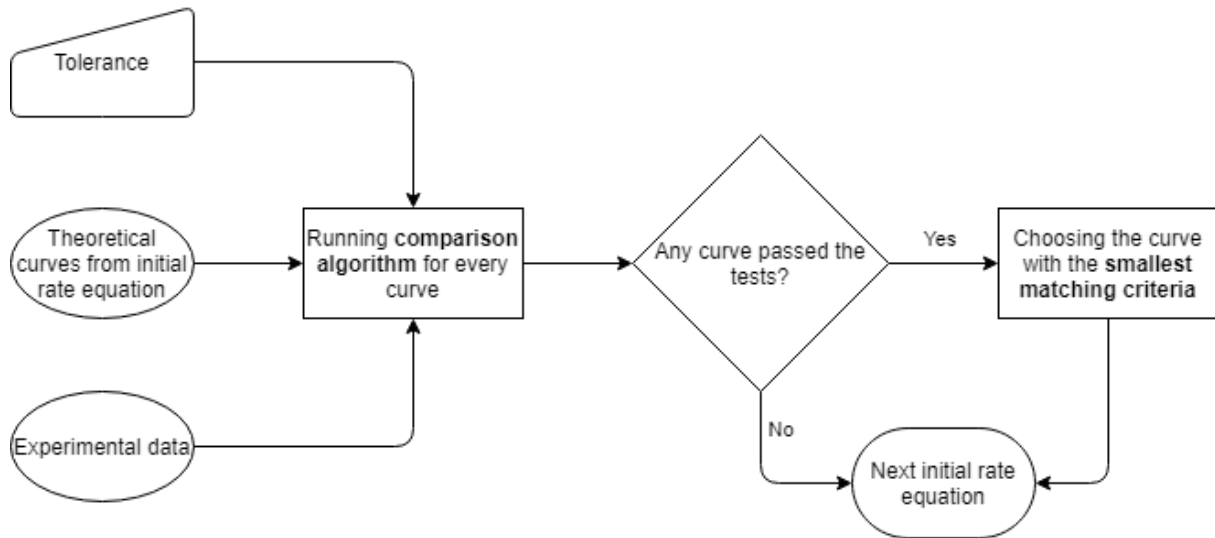


Figure 9 - Flowchart describing the process of choosing the best curve of each initial rate equation (trapezoid – user input; rectangle – process; oval – output of previous branches; rhombus – decision; rounded rectangle – terminator)

After testing every curve generated from an initial rate equation, the tool chooses the one that has the smallest SRE value (and the smallest MSE_{norm} , if necessary) from the ones that passed the tests. This curve is the one from this specific initial rate equation that best follows the experimental data trends, and therefore will be present in the output. This whole process of testing and choosing the best theoretical curve is reinitiated for the theoretical curves generated from the following initial rate equation. After running this algorithm for all initial rate equations, the final result will be a list of every initial rate equation that generated theoretical curves features similar to those of the experimental data, and the equilibrium constants and ranking criteria of the best curve generated by each initial rate equation. This list is then ordered by increasing order of SRE (and MSE_{norm} , if necessary), which results in a ranking of possible models for the experimental data.

4.3. Estimating the lumped kinetic constant (k')

The desired output of the tool was not only the proposed initial rate equations, but also some initial guesses for the parameters of the curves that could be used for an eventual regression of the curves. These guesses already exist for the equilibrium constants since the theoretical curves were generated by permutating different values for them. However, as stated before, no theoretical curves were generated by permutating values for k' , since all theoretical curves were normalized to have their features extracted. Therefore, a value for this parameter was estimated to also serve as an initial guess.

The initial k' value is estimated for the best curve of each proposed rate equation. The necessary information to do so is the initial rate equation that generated it (see Table 4) and its already determined guesses for the equilibrium constants. Then, the tool calculates the necessary k' value for the curve to intersect the first point of the experimental dataset. Then, a curve is generated using this calculated k' value, as well as the guesses for the equilibrium constants. Finally, the mean squared error (MSE) of this generated curve is calculated using Eq. 15.

$$MSE = \frac{\sum_{i=1}^N (r_{0,Theor}(i) - r_{0,Exp}(i))^2}{N} \quad \text{Eq. 15}$$

In this equation, $r_{0,Theor}(i)$ and $r_{0,Exp}(i)$ are the initial rate values of the new generated curve and the experimental data, respectively, and N is the total number of data points. After this, this process is repeated for the next point of the experimental data. Doing it for every point of the experimental dataset results in n possible k' values. Taking the average of all these values results in another possible value, for which a curve is also generated and its resulting MSE determined. Of all these values, the one with the smallest associated MSE is the one chosen as the initial guess. The smallest MSE value is also the one present in the tool's output.

4.4. Reducing runtime

As stated before, the tool will generate initial rate curves for all initial rate equations in its library. However, it is possible that the user already has some knowledge on the reaction to be studied that indicates that some of the equations on the library cannot properly describe how the reaction occurs. In an effort to reduce the runtime, the tool offers the user the possibility to select these equations. The equations selected here will not be considered when generating the theoretical curves.

4.5. Output of the tool

The first output of the tool is related to the curve features extracted from the experimental data. This output consists of a graph with the experimental dataset and its extracted features represented on it, as well as the features themselves in text format. An example of one of these graphs is present in Figure 10.

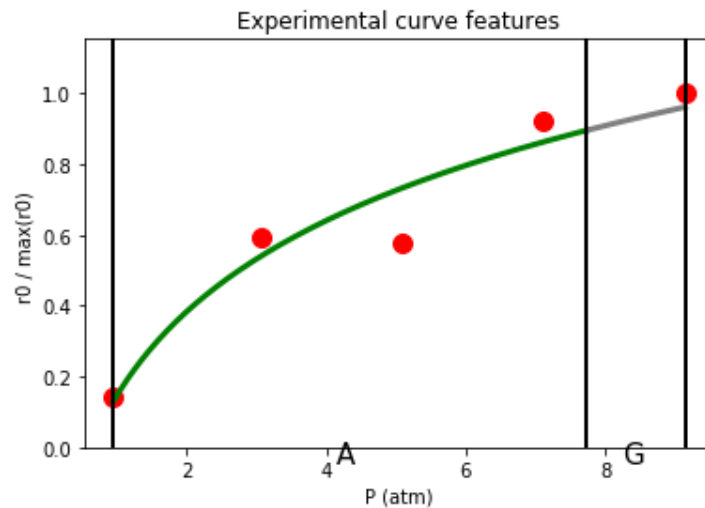


Figure 10 - Example of output graph with experimental features

The reason for the tool to output this graph is so the user can better understand the reason for the tool to choose the proposed initial rate equations. For example, if there are no equations proposed by the tool, it could be related to the complexity of the experimental features. On the other hand, if the tool proposes a high number of equations, it could be related to the simplicity of the experimental features.

After this graph, the tool outputs a graph for the best theoretical curve of each mechanism. The order of the graphs follows the order of the ranking of these curves. These graphs contain the theoretical curve overlapping the experimental data points. For clarity, instead of only presenting theoretical points at the same pressure values as the experimental points, the graph presents an uninterrupted curve covering the whole pressure range. An example of such a graph is present in Figure 11. Following each of these graphs is the initial rate equation that generated this curve as well as the initial guesses for the parameters and the ranking criteria (SRE and MSE), all in text format. An example of this is present in Figure 12.

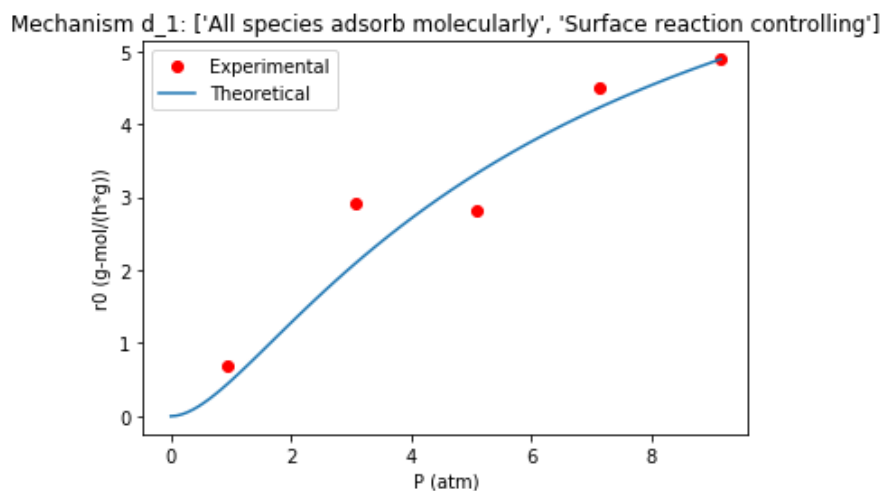


Figure 11 - Example of output graph with theoretical curve of proposed rate equation

Rate equation:

$$r = \frac{k' \cdot (KA \cdot y_A)^2 \cdot P^2}{(1 + KA \cdot y_A \cdot P)^2}$$

A = C3H6

Specific constants of this theoretical curve
(Units coincide with pressure and rate units):

k' = 9.102e+00

KA = 3.000e-01

Ranking criteria:

SRE = 2.140e-02

MSE = 2.089e-01

Figure 12 - Example of output text string

From observing these graphs, the user can better see how the theoretical curves follow the experimental data trends. It is important to note that the fitting of the curves to the experimental data was not performed, so these curves should not be discriminated on how well they fit the curve, but rather on how close their trends resemble the experimental data trends. From the text that follows each graph, the user obtains more information on these curves in order to better interpret the results.

A flowchart describing in more detail the steps taken by the tool from start to finish is represented in Figure 13.

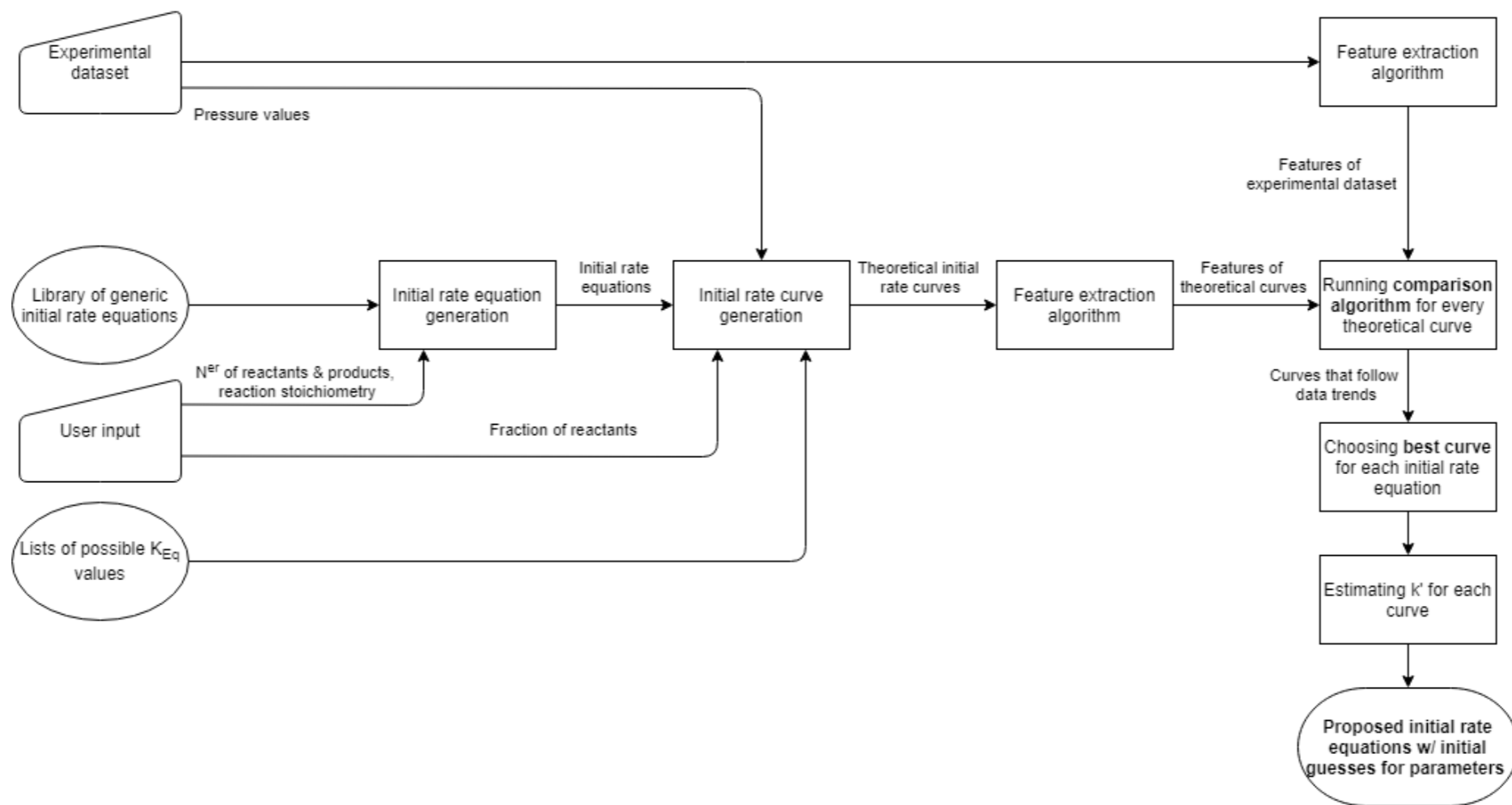


Figure 13 - Full flowchart of the tool (trapezoid – user input; rectangle – process; information included in tool; rounded rectangle – output of tool)

4.6. Conclusions

To determine which initial rate equations can describe the data, the theoretical curves generated by the models in the library are compared to the experimental data, based on their features. Due to the way the feature extraction algorithm works, two decisions had to be made to ensure that the features could be compared. One of these decisions was to generate the theoretical curves for the same pressure values of the experimental data, since the distribution of the points of a given dataset along the pressure axis influences the features extracted from it. The other decision was to normalize both the experimental data and the theoretical curves before extracting their features, since the scaling of the dataset also has influence on the extracted features, due to the numerical nature of the feature extraction algorithm. This nullifies the influence of k' from the equation, so there is no need to test different values for this constant. However, this makes it necessary to estimate its value for every proposed initial rate equation. While these issues do not arise in most cases, they can still be problematic, which makes it necessary to implement these decisions in the code.

The comparison of the experimental features with the theoretical ones is made through two tests. The first one tests whether the theoretical curve has the same primitives as the experimental data. If not, the curve is disregarded. If yes, then the curve moves on to the second test, which checks whether the pressure value of each extreme of the theoretical curve is within the tolerance range chosen by the user. If yes, then it is considered that the theoretical curve has features similar enough to the experimental data and therefore, the initial rate equation from which the curve was generated is proposed by the tool. By letting the user choose the tolerance for the second test, the user has more control on how similar a curve must be for it to be considered to have similar features to the experimental data. Also, since these tests can be repeated for different tolerance values, the user can search for other possible initial rate equations by increasing the tolerance.

To choose the theoretical curve from each initial rate equation most similar to the experimental data and also to rank the proposed initial rate equations, some ranking criteria are necessary. The SRE was chosen as the principal criterion, since it is the one that best translates how similar the features of a theoretical curve are to the experimental features. As a secondary criterion the mean squared error was chosen, be it of the normalized data (MSE_{norm}) to differentiate curves from the same equation, be it of the regular data (MSE) to rank the proposed equations. The MSE indicates the goodness of fit of a curve to a dataset. However, no curve fitting is performed by the tool. Therefore, when analysing this value and the best theoretical curve (see Figure 11) in the output it is important to remember that the tool does not propose the initial rate equations based on the goodness of fit, but based instead on how similar the features of their best curves are to the experimental features, which is represented by the SRE.

As previously mentioned, the tool estimates the k' value for the best curve of each proposed initial rate equation. The method used to do so is rudimentary. However, a more elaborated method is not necessary, since the objective at this point is not to determine as accurately as possible the value of k' , since the values of all the other parameters are still initial guesses for an eventual regression, and to get an initial guess for the value of k' is in fact the objective.

5. Case studies

To test the tool, experimental datasets of initial rate vs total pressure from five different sources were analysed. The chosen datasets were ones which follow the criteria required for data to be analysed by the tool: maximum of 2 reactants and 2 products, constant molar fractions of reactants in the feed, no products in feed, negligible conversion and constant temperature. Also, the source of the chosen datasets had to contain the rate equation determined by the researchers as the one that best describes the data. Besides these reasons, each case study had a specific motivation behind it. The first case study serves to test the generic performance of the tool. The second, third and fourth case studies all study the influence of the quality of the data in some way: the second studies the influence of the distribution of data points through the pressure axis, the third studies the influence of data variability and experimental noise, and the fourth studies the influence of doubling points from repeated experiments, as well as the influence of experimental noise again. This way, the influence of data quality in the performance of the tool is properly studied, which is important since this is a deciding factor in kinetic modelling, be it automatic or manual. The fifth case tests the tool with a dataset for which the determined rate equation is not present in the library of the tool. This serves to test if the tool can propose a rate equation that serves as a good approximation to the actual one not in its library. The motivations behind each case study are summarized in Table 8.

Table 8 - Motivations behind each case study

Case study	Feature to be tested
5.1. Oxidation of methane	Generic performance
5.2. Propylene hydrogenation	Influence of data distribution along pressure axis
5.3. Ethanol dehydrogenation	Influence of data variability on rate axis and noise
5.4. Propylene disproportionation	Influence of repeat tests and data noise
5.5. Propyne hydrogenation	Behaviour for rate equations not in tool's library

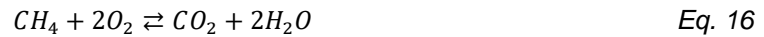
This way, the evaluation of the performance of the tool was made by checking if the rate equation determined was among the ones proposed by the tool and whether it is the one that generated the curve with the features most similar to those of the experimental data or not. Also, when the researcher had determined the kinetic constants of the rate equation, these were compared with the ones proposed by the tool.

For each dataset tested, the experimental curve features were compared with the features of the generated theoretical curves a total of five times, changing only the tolerance allowed for the extreme values of a given theoretical curve. The tolerance values used were 0.10, 0.20, 0.30, 0.40 and 0.50. Different tolerance values were chosen to test the influence of the tolerance in the rate equations proposed by the tool. More specifically, these five values were chosen to have a clear evolution of the proposed rate equations, up until 0.5, which was determined to be a tolerance value large enough to encompass all rate equations that could be potentially proposed. The role of tolerance in the screening of rate equations is present in Chapter 3. Since the purpose of these tests was to see which rate

equations were proposed from all the ones present in the toll library, no mechanisms were eliminated prior to the curve generation.

5.1. Oxidation of methane – General test of the tool

This case study is based on the oxidation of methane. The reaction in question is as follows:



The initial rate vs total pressure dataset of this reaction was extracted from the work of Ahuja[23]. The specific dataset utilized was obtained for constant molar fractions of reactants in the feed of 0.065 of CH_4 and 0.935 of O_2 , with no inert species present. A graphical representation of the dataset is present in Figure 14, while the actual values of the dataset are present in Appendix 3.

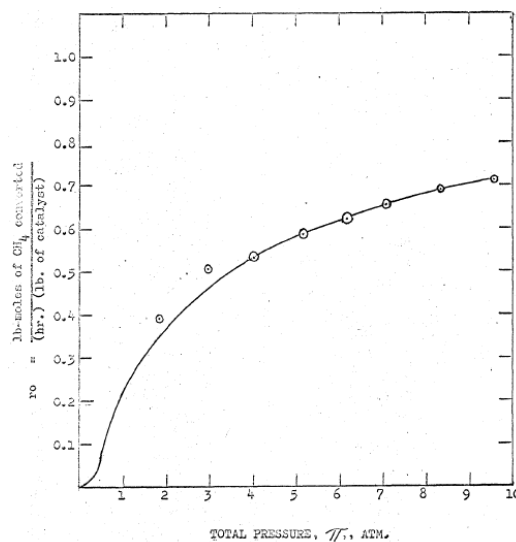
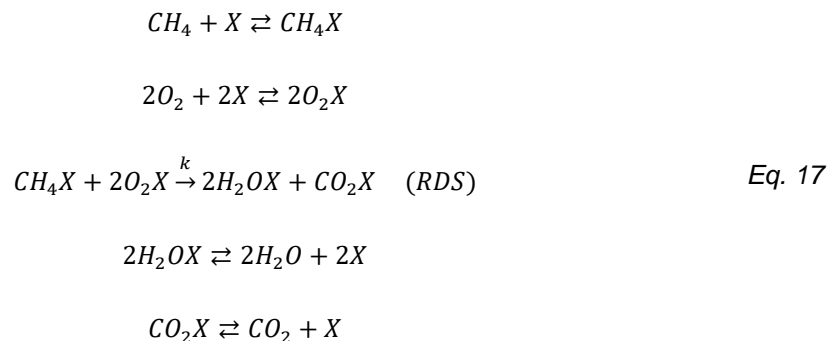


Fig. 4 INITIAL RATE OF REACTION VS TOTAL PRESSURE, 320°C

Figure 14 - Graphical representation of the methane oxidation dataset extracted at 320 °C (markers – experimental points; solid line – guideline drawn by researcher)[23]

The mechanism (Eq. 17) and rate equation (Eq. 18) determined by Ahuja for this reaction is of the following type:



$$r_0 = \frac{k' \cdot K_{CH_4} \cdot P_{CH_4} \cdot (K_{O_2} \cdot P_{O_2})^2}{(1 + K_{CH_4} \cdot P_{CH_4} + K_{O_2} \cdot P_{O_2})^3} \quad \text{Eq. 18}$$

This rate equation is equivalent that of mechanism 'd', where both reactants adsorb molecularly and the RDS is the reaction step on the catalyst surface (see Table 4). Therefore, this equation is expected to be among the ones proposed by the tool, if not the best one among all of them.

5.1.1. Test results

This dataset was used as an input of the tool. All the required user input was inserted: the name and stoichiometric coefficient of the reactants and products, as well as the fractions of each reactant in the feed. This reaction involves two reactants, so the second run with the order of reactants swapped was performed. These features are graphically represented in Figure 15 and present in Table 9.

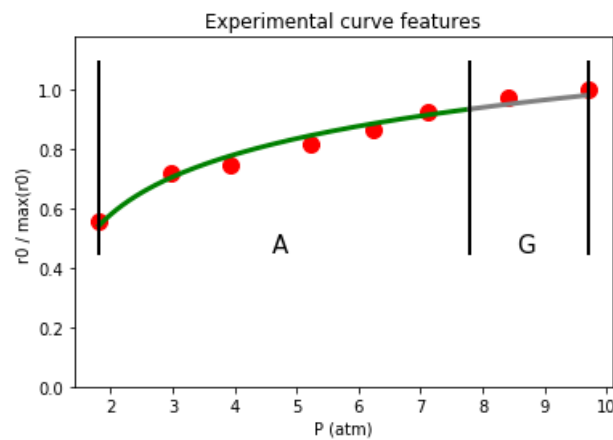


Figure 15 – Graphical representation of the recognized features of the methane oxidation dataset

Table 9 - Recognized features of the methane oxidation dataset

Primitives	A, G		
Pressure values of extremes (atm)	1.817	7.789	9.7

Comparing Figure 14 and Figure 15 it can be observed that the trend following curve generated by the tool is similar to the one drawn by the researcher. Therefore, the recognized features of the experimental data are a good representation of the data trends. The proposed rate equations for each tolerance value are present in Table 10 and the best curve for each of the top 5 proposed rate equations in Table 11.

Table 10 - Proposed rate equations for the methane oxidation dataset (in order of ascending values of ranking criteria), for each tolerance value (mechanisms with _1: A=CH₄, B=O₂; mechanisms with _2: A=O₂, B=CH₄)

Tolerance	Proposed rate equations
0.1	d_1, k_2, k_1, a_1, e_2
0.2	d_1, k_2, k_1, a_1, e_2, b_1
0.3 to 0.5	d_1, k_2, k_1, a_1, e_2, b_1, e_1

Table 11 - Constants and ranking criteria for the best curve of the top 5 proposed rate equations for the propylene oxidation dataset (mechanisms with _1: A=CH₄, B=O₂; mechanisms with _2: A=O₂, B=CH₄)

ID	Rate equation	k'	K _A (atm ⁻¹)	K _B (atm ⁻¹)	MSE	SRE
d_1	$r_0 = \frac{k' \cdot K_A \cdot y_A \cdot (K_B \cdot y_B)^2 \cdot P^3}{(1 + (K_A \cdot y_A + K_B \cdot y_B) \cdot P)^3}$	28 lb-mol·h ⁻¹ ·lb ⁻¹	0.5	0.7	9.20E-3	0.00
k_2	$r_0 = \frac{k' \cdot (K_A \cdot y_A)^2 \cdot y_B \cdot P^3}{(1 + K_A \cdot y_A \cdot P)^3}$	13 lb-mol·h ⁻¹ ·lb ⁻¹	0.7	N/A	1.02E-2	1.02E-2
k_1	$r_0 = \frac{k' \cdot (K_A \cdot y_A)^2 \cdot y_B \cdot P^3}{(1 + K_A \cdot y_A \cdot P)^3}$	0.61 lb-mol·h ⁻¹ ·lb ⁻¹	10	N/A	1.04E-2	1.02E-2
a_1	$r_0 = \frac{k' \cdot y_A \cdot P}{1 + K_B \cdot y_B \cdot P}$	4.4 lb-mol·h ⁻¹ ·lb ⁻¹ ·atm ⁻¹	N/A	0.3	1.05E-3	4.09E-2
e_2	$r_0 = \frac{k' \cdot y_A \cdot P}{(1 + K_B \cdot y_B \cdot P)^2}$	0.21 lb-mol·h ⁻¹ ·lb ⁻¹ ·atm ⁻¹	N/A	1	2.77E-3	7.16E-2

As it can be seen in Table 11, the rate equation considered by the tool as the best model is the one resulting from mechanism d_1, since the respective theoretical curve is the one with the smallest SRE – in fact, all the extremes of this theoretical curve are the same as the experimental ones. This rate equation is the one determined by the researcher. The best curve of this rate equation is present in Figure 16.

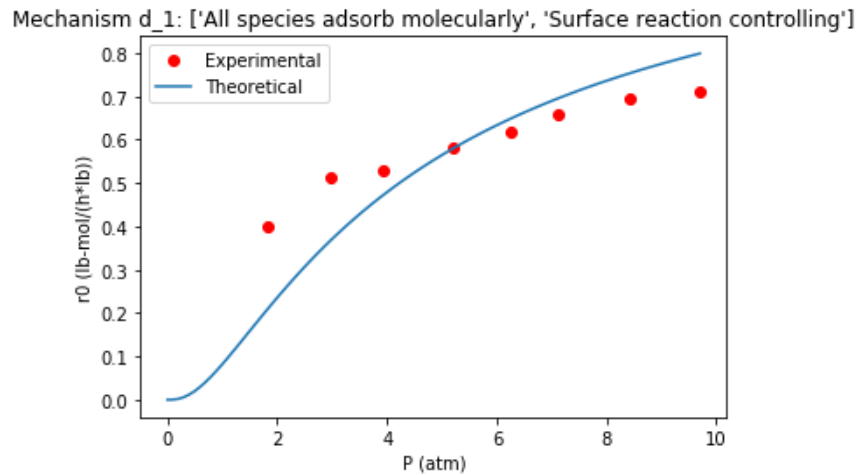


Figure 16 - Best curve of the rate equation of mechanism d_1 for the methane oxidation dataset

However, the possible rate equation which has the theoretical curve with the smallest MSE is in fact that of mechanism b_1. The best curve of this equation is present in Figure 17.

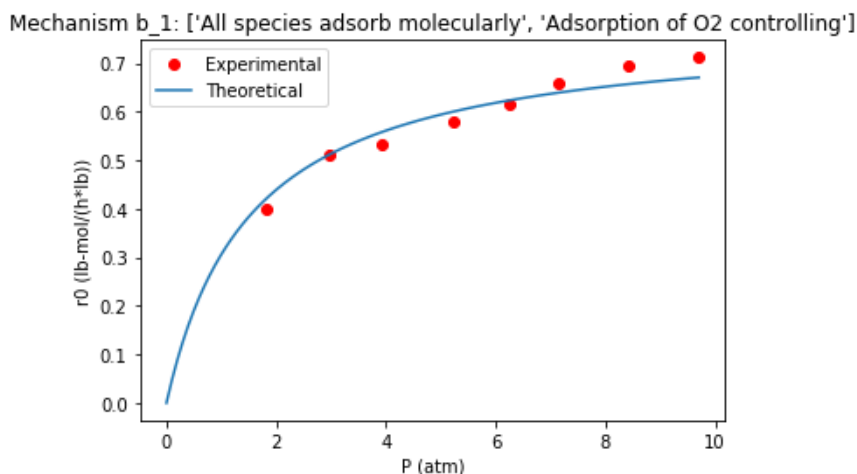


Figure 17 - Best curve of the rate equation of mechanism b_1 for the methane oxidation dataset

5.1.2. Discussion

As mentioned above, the rate equation identified by the researchers to best describe the experimental data (Eq. 18) is not only proposed as a possibility by the tool, but also as the best possibility (regarding SRE) among all proposed rate equations, since the value of this parameter for this rate equation is exactly 0. This means that the extreme between primitives A and G of this theoretical curve are exactly the same as in the experimental one. In contrast, the equivalent extreme of the rate equation of mechanism b_1 differs by 15.3% of the experimental value, which even though is not that high of a difference, it still was enough to make this rate equation as one of the worst possibilities (regarding SRE). Also, the best three proposed rate equations (from mechanisms d_1, k_1 and k_2) all have numerators and denominators with the same orders of magnitude regarding pressure as Eq. 18. Therefore, it can be said that the tool works properly for this case study.

However, the rate equation with the smallest MSE value is not the one of mechanism d_1 but of mechanism b_1. By analysing the curves of these rate equations (present in Figure 16 and Figure 17, respectively) it is easy to observe that the best fitting curve is in fact the latter. However, this is explainable with the fact that the tool does not perform a regression to the theoretical curves, so the constants used to create these curves were not fine-tuned to create the best fit possible to the experimental data and serve only as initial values for an eventual regression.

5.2. Propylene hydrogenation – Influence of data distribution

This case study is based on the reaction propylene hydrogenation on nanostructured palladium clusters, which has the following global reaction:



The initial rate vs total pressure dataset of this reaction was extracted from the work of Brandão *et al.*[24]. This dataset was obtained for a temperature of 308 K and a feed composition, in molar fractions, of 0.10 of H_2 , 0.35 of C_3H_6 and 0.55 of argon, which acted as an inert. A graphical representation of the dataset is present in Figure 18, while the actual values of the dataset are present in Appendix 3.

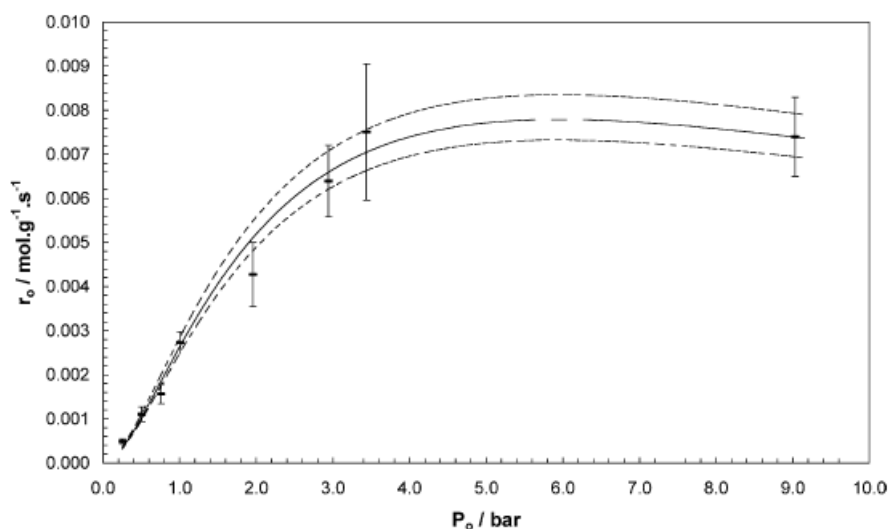
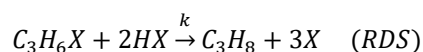
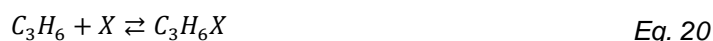
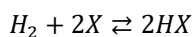


Figure 18 - Graphical representation of the propylene hydrogenation dataset[24] (markers – experimental points; vertical bars – error of each point; solid line – model fitting; dashed lines – 95% confidence curves). Reprinted with permission from [24]. Copyright 2004 Elsevier B.V. All rights reserved.

One particularity of this dataset is that there are no datapoints in the pressure range (from 3.439 and 9.034 bar) where one of its most important features is present – the maximum of initial rate values. The effect of this fact will be analysed in the results.

The mechanism (Eq. 20) and rate equation (Eq. 21) for this reaction determined by the researchers for this reaction was the following:



$$r_0 = \frac{k' \cdot K_{H_2} \cdot P_{H_2} \cdot K_{C_3H_6} \cdot P_{C_3H_6}}{(1 + \sqrt{K_{H_2} \cdot P_{H_2}} + K_{C_3H_6} \cdot P_{C_3H_6})^3} \quad \text{Eq. 21}$$

The researchers were also able to determine the values of the kinetic and equilibrium constants for this dataset, which are presented in Table 12.

Table 12 - Constants for the propylene hydrogenation dataset determined by Brandão et al.[24].

k' (mol.g ⁻¹ .s ⁻¹)	5.569
K_{H_2} (bar ⁻¹)	3.799E-02
$K_{C_3H_6}$ (bar ⁻¹)	0.996

The rate equation in Eq. 21 is the same as the one for mechanism 'h', where reactant A (in this case, H₂) dissociates when adsorbing and the rate controlling step is the reaction on the catalyst surface.

Therefore, if the tool works properly with this dataset, this rate equation will be among the proposed ones with curve constants as close as possible to the ones in Table 12.

5.2.1. Test results

This dataset was used as the tool input, and all the user input regarding the reactants, products and inert species was inserted. This reaction also has two reactants, so the second run with the swapped reactants was performed. In addition, it only produces one product, which drastically increases the runtime, as mentioned in Subchapter 3.3. So, to make this test more feasible, the reduced lists of K_{Eq} values were chosen when the tool offered the choice. The features recognized by the tool for this dataset are graphically represented in Figure 19 and are present in Table 13.

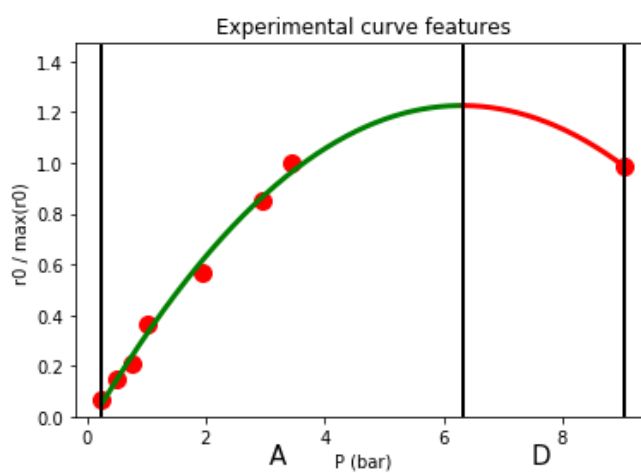


Figure 19 - Graphical representation of the recognized features of the propylene hydrogenation dataset

Table 13 - Recognized features of the propylene hydrogenation dataset

Primitives	A, D		
Extremes (atm)	0.257	6.330	9.034

Comparing Figure 18 with Figure 19, it can be observed that the maximum of the initial rate values recognized by the tool is in a pressure value similar to the one recognized by the researchers. Therefore, the absence of data points in that pressure range did not affect the feature recognition. The rate equations proposed by the tool in each comparison are present in Table 14, and the constants and ranking criteria of the best curve of top 5 proposed equations are present in Table 15.

Table 14 - Proposed rate equations for the propylene hydrogenation dataset (in order ascending values of ranking criteria), for each tolerance value (mechanisms with $_1$: $A=H_2$, $B=C_3H_6$; mechanisms with $_2$: $A=C_3H_6$, $B=H_2$)

Tolerance	Proposed rate equations
0.1	$h_{_1}, p_{_1}, p_{_2}, m_{_2}, g_{_2}, j_{_2}, m_{_1}, j_{_1}, g_{_1}, c_{_1}$
0.2 to 0.5	$h_{_1}, p_{_1}, p_{_2}, m_{_2}, g_{_2}, j_{_2}, m_{_1}, j_{_1}, g_{_1}, c_{_1}, h_{_2}, e_{_1}$

Table 15 - Constants and ranking criteria for the best curve of the top 5 proposed rate equations for the propylene oxidation dataset (mechanisms with _1: A=H₂, B=C₃H₆; mechanisms with _2: A=C₃H₆, B=H₂)

ID	Rate equation	k' (mol·g ⁻¹ ·s ⁻¹)	K _A (bar ⁻¹)	K _B (bar ⁻¹)	K _R (bar ⁻¹)	K _r (bar ⁻¹)	MSE	SRE
h_1	$r_0 = \frac{k' \cdot K_A \cdot y_A \cdot K_B \cdot y_B \cdot P^2}{(1 + \sqrt{K_A \cdot y_A \cdot P} + K_B \cdot y_B \cdot P)^3}$	2.3	0.1	1	N/A	N/A	1.33E-7	0.00
p_1	$r_0 = \frac{k' \cdot K_g \cdot y_A \cdot y_B \cdot P^2}{1 + K_R \cdot K_g \cdot y_A \cdot y_B \cdot P^2 + K_A \cdot y_A \cdot K_B \cdot y_B \cdot P^2 + K_B \cdot y_B \cdot P}$	3.8E-03	100	0.1	0.5	1	5.81E-6	0.00
p_2	$r_0 = \frac{k' \cdot K_g \cdot y_A \cdot y_B \cdot P^2}{1 + K_R \cdot K_g \cdot y_A \cdot y_B \cdot P^2 + K_A \cdot y_A \cdot K_B \cdot y_B \cdot P^2 + K_B \cdot y_B \cdot P}$	7.5E-01	10	1	100	1	6.05E-6	7.00E-3
m_2	$r_0 = \frac{k' \cdot K_g \cdot y_A \cdot y_B \cdot P^2}{1 + K_R \cdot K_g \cdot y_A \cdot y_B \cdot P^2 + \sqrt{K_A \cdot y_A \cdot P}}$	3.8E-03	100	N/A	0.5	1	4.90E-7	6.30E-2
g_2	$r_0 = \frac{k' \cdot K_g \cdot y_A \cdot y_B \cdot P^2}{1 + K_R \cdot K_g \cdot y_A \cdot y_B \cdot P^2 + \sqrt{K_A \cdot y_A \cdot P} + K_B \cdot y_B \cdot P}$	3.8E-03	100	0.1	0.5	10	4.93E-7	6.3E-02

It can be observed in Table 15 that there are two rate equations with theoretical curves which recognized features are exactly the same as the ones recognized for the experimental dataset. Also, one of these is the one determined by the researchers. The best curves of these two equations are present in Figure 20 and Figure 21, respectively.

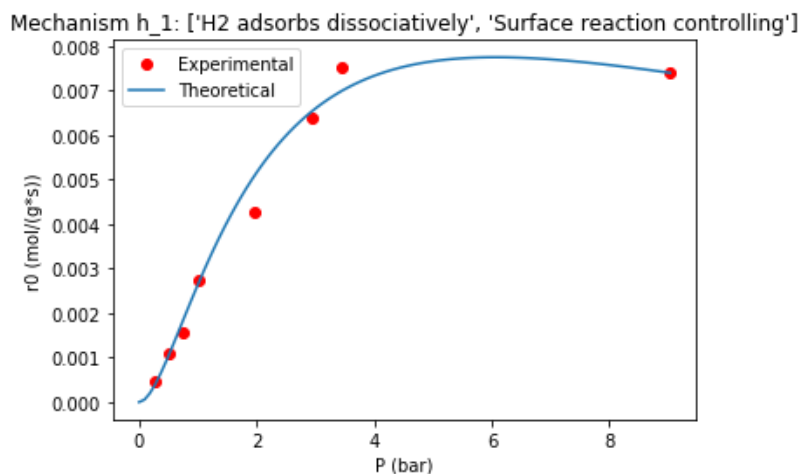


Figure 20 - Best curve of the rate equation of mechanism h_1 proposed for the propylene hydrogenation dataset

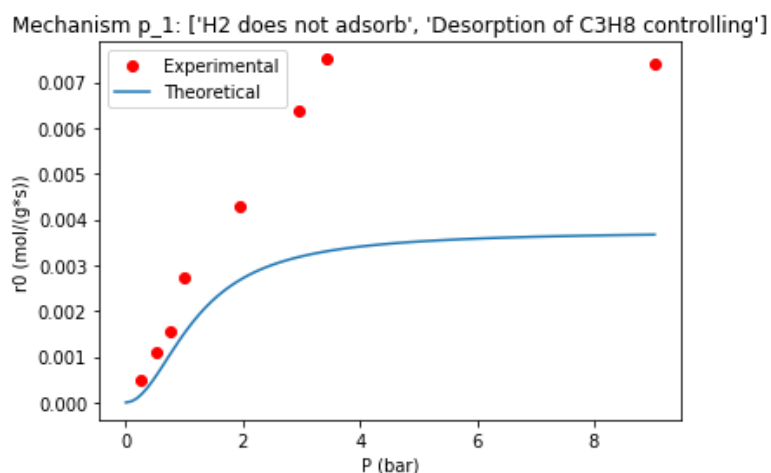


Figure 21 - Best curve of the rate equation of mechanism p_1 proposed for the propylene hydrogenation dataset

5.2.2. Discussion

As mentioned above, the rate equation determined by the researchers was recognized by the tool as one of the two with the curve features most similar to the ones from the experimental dataset. Also, the equilibrium constants of the best curve of this equation (present in Table 15) have the same order of magnitude as the values determined by the researchers (present in Table 12). Therefore, this indicates that the tool works properly for this case study.

The other mechanism with a rate equation with a SRE of 0 is mechanism p_1. However, by analysing the proposed rate equation (Table 15), it can be seen that the numerator and denominator have the same order of magnitude when regarding pressure, so the initial rate values will increase with increasing pressure until reaching a certain limit and will never decrease. However, by analysing Figure 18 and Eq. 21 it can be observed that the initial rate curve has a decreasing trend starting at a pressure value

between 3.439 and 9.034 bar. A graphical representation of the recognized features of the p_1 theoretical curve is present in Figure 22.

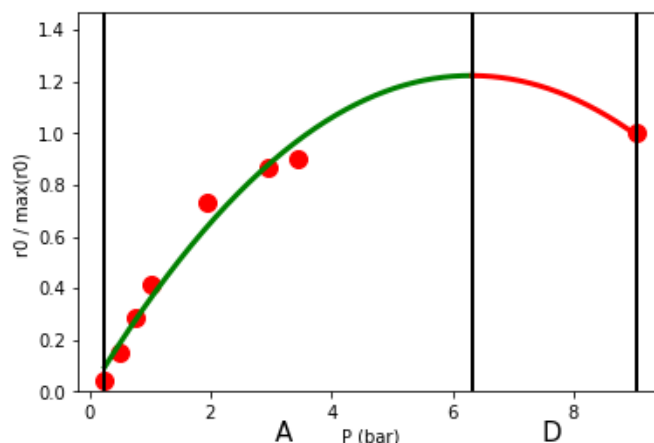


Figure 22 - Graphical representation of the recognized features of the p_1 theoretical curve

As mentioned in Subchapter 4.1.1, the features of the theoretical curves are extracted using only the points with the same pressure values as the experimental dataset. Therefore, since there are no points in the range between 3.439 and 9.034 bar, the recognized features assumed there was a decreasing trend in the data in this range, even though the rate equation that resulted in this data made it impossible. This is a negative effect of the absence of data points in the pressure range where important curve features are present.

It is also important to note that mechanism p_1 is one of the mechanisms affected by the choice of using the reduced list of possible values of the equilibrium constants. However, even if the full list was used, the resulting rate equation would always be a worse model when compared to the one of mechanism h_1, due to the reasons already stated.

When it comes to the effect of the tolerance, once again the first test with a value of 0.1 would be sufficient to identify the correct rate equation. Higher tolerances would only making it possible for rate equations that generate curves with features less similar to those of the experimental data to be proposed.

5.3. Ethanol dehydrogenation – Influence of data variability and noise

The next case study is focused on the ethanol dehydrogenation, which has the following global reaction:



The initial rate vs total pressure datasets of this reaction used to test the tool was obtained from the work of Franckaerts and Froment[25]. The total number of datasets from this paper that were used to test the tool was 6, each one of them having been obtained at a different temperature – 225, 235, 250, 265, 275 and 285 °C. For all of them the feed was composed by the molar fractions of 0.865 of EtOH and 0.135 of water, which acts as an inert component. A graphical representation of these datasets is present in Figure 23, while the actual values of the datasets are present in Appendix 3.

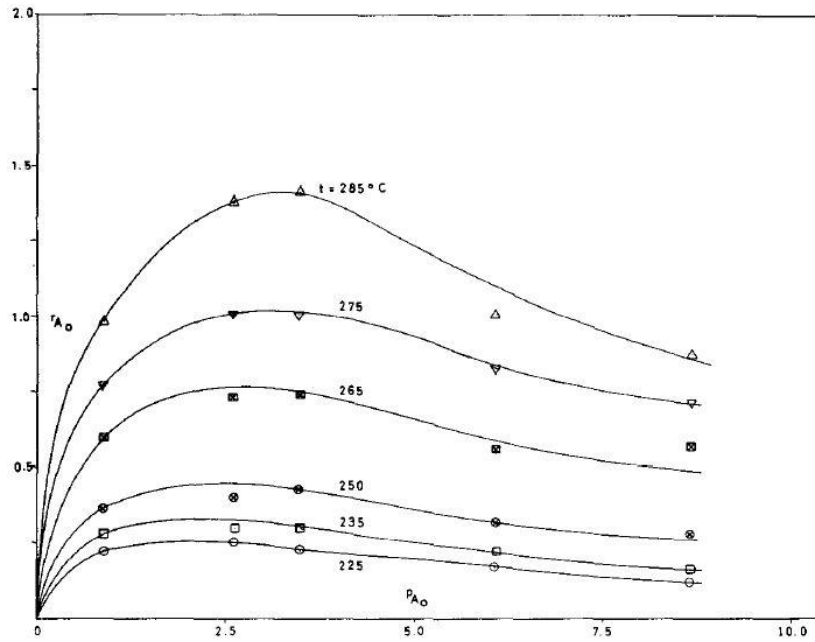


Figure 23 - Experimental datasets of the ethanol dehydrogenation reaction[25] (markers – experimental points; solid line – model fitting). Reprinted with permission from [25]. Copyright 1964 Published by Elsevier Ltd.

By observing Figure 23, it can be seen that some of the datasets have curves with very subtle features, i.e., have very low variability of initial rate values in the ranges of the important features (namely the maximum, in this case). Besides this, some datasets have significant noise in some of the points. The tests performed on these datasets will serve to see what the effect of these factor on the performance of the tool is.

The researchers who obtained this dataset described the mechanism of this reaction as being controlled by the “surface reaction on dual sites, without dissociation”[25], and determined the initial rate equation to be the following:

$$r_0 = \frac{k' \cdot K_{EtOH} \cdot P_{EtOH}}{(1 + K_{EtOH} \cdot P_{EtOH})^2} \quad \text{Eq. 23}$$

They also determined the values of the kinetic and equilibrium constants, which are present in Table 16.

Table 16 - Kinetic and equilibrium constants for the datasets of the ethanol dehydrogenation reaction determined by Franckaerts and Froment[25]

T (°C)	k' (mol·h ⁻¹ ·g ⁻¹)	K _{EtOH} (atm ⁻¹)
225	1.07	0.74
235	1.30	0.62
250	1.71	0.46
265	2.96	0.37
275	4.03	0.40
285	5.58	0.44

Eq. 23 corresponds to the same equation of mechanism 'd' (EtOH adsorbs molecularly and the RDS is the reaction step in the catalyst surface). Therefore, if the tool works properly with these datasets, it is expected that the rate equation of this mechanism is proposed as a possibility, if not the best one, and that the constants of the best curve of this mechanism are as close as possible to the ones in Table 16.

5.3.1. Test results

- 225, 235 and 250°C datasets:

The graphical representations of the recognized features of these three datasets are presented in Figure 24, while the features themselves are in Table 17.

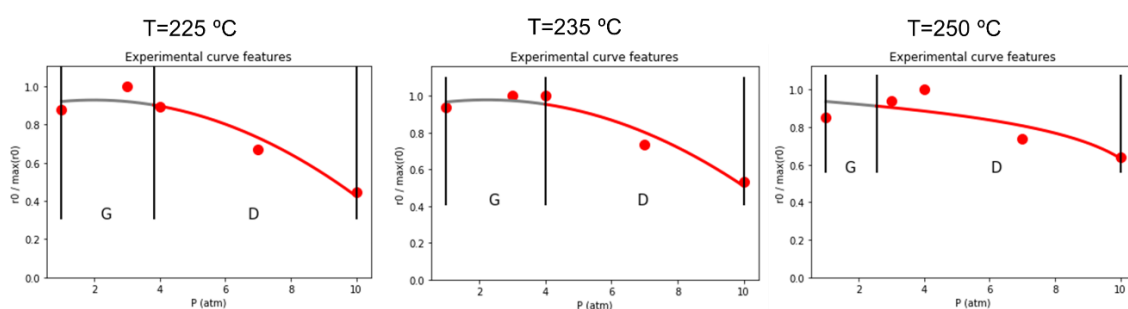


Figure 24 - Graphical representation of the recognized features of the 225, 235 and 250 °C ethanol dehydrogenation datasets

Table 17 - Recognized features of the 225, 235 and 250 °C ethanol dehydrogenation datasets

T (°C)	225			235			250		
Primitives	G. D			G. D			G. D		
Extremes (atm)	1	3.818	10	1	4	10	1	2.545	10

These datasets resulted in no rate equations being proposed by the tool, since no theoretical curves had the same primitives as the experimental data. This is because the rise in initial rate values for small pressures is very subtle in these datasets – so subtle that the different values are interpreted by the algorithm as noise in a near-constant value curve.

- 265 °C dataset:

The recognized features of this dataset are graphically represented in Figure 25 and present in Table 18.

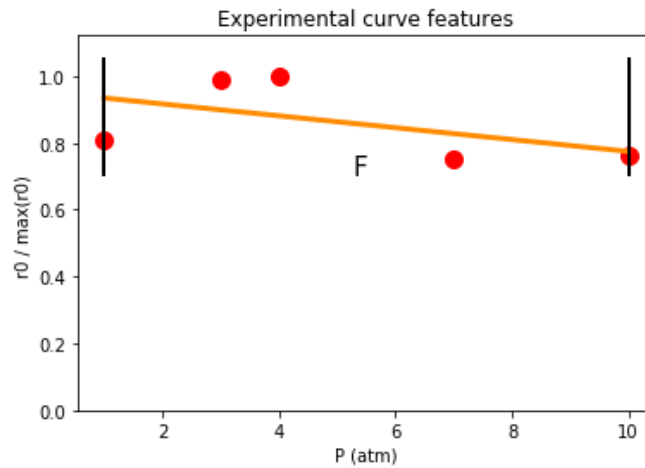


Figure 25 - Graphical representation of the recognized features of the 265 °C ethanol dehydrogenation dataset

Table 18 - Recognized features of the 265 °C ethanol dehydrogenation dataset

Primitives:	F	
Extremes (atm):	1	10

The only rate equation proposed as a possibility for this dataset is the one of mechanism 'd_1'. Since the set of primitives of this dataset consists of only one primitive, this equation was proposed for every tolerance value, since the theoretical and experimental extremes are the same. This equation, as mentioned before, is equal to Eq. 23. The constants and ranking criteria of the best curve of this rate equation are present in Table 19, and the curve itself is represented in Figure 26.

Table 19 - Constants and ranking criteria for the best curve of the proposed rate equation for the 265 °C ethanol dehydrogenation dataset (A=EtOH)

ID	Rate equation	k' ($\text{mol}\cdot\text{h}^{-1}\cdot\text{g}^{-1}$)	K_A (atm^{-1})	MSE	SRE
d_1	$r_0 = \frac{k' \cdot K_A \cdot y_A \cdot P}{(1 + K_A \cdot y_A \cdot P)^2}$	3.2	0.7	1.28E-2	0.00

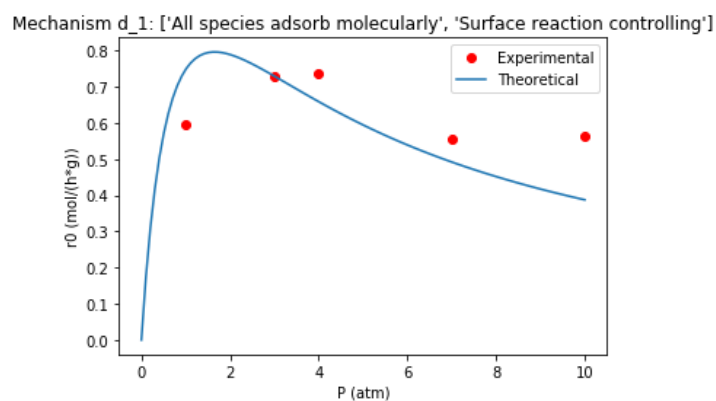


Figure 26 - Best curve of the rate equation mechanism d_1 proposed for the 265 °C ethanol dehydrogenation dataset

Even though the recognized features of the experimental data do not represent the rise in initial rate values for small pressure values, as it can be seen in Figure 25, there was one theoretical curve generated by the rate equation of mechanism 'd_1' that had similar features. However, since the recognized experimental features do not accurately reflect the evolution of the experimental data, the proposed curve for this rate equation does not properly follow the trends of the experimental data, as can be seen in Figure 26, and its associated value of K_{EtOH} is not close to the one determined by the researchers, present in Table 16. Therefore, the tool was successful in recognizing the correct rate equation, but the proposed curve of that equation was not the best one, due to the misidentified features.

- 275 and 285 °C datasets:

The recognized features of the 275 and 285 °C datasets are graphically represented in Figure 27 and are present in Table 20.

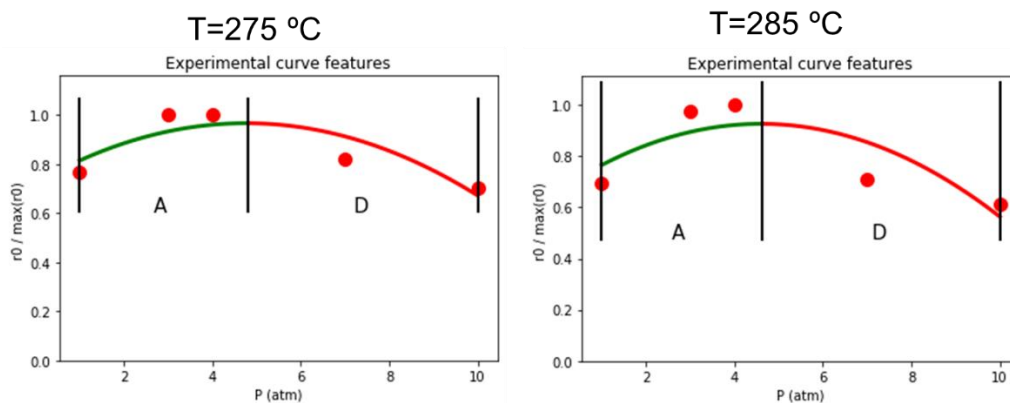


Figure 27 - Graphical representation of the recognized features of the 275 and 285 °C ethanol dehydrogenation datasets

Table 20 - Recognized features of the 275 and 285 °C ethanol dehydrogenation datasets

T (°C)	275			285		
Primitives	A, D			A, D		
Extremes (atm)	1	4.818	10	1	4.636	10

The only rate equation recognized as a possibility for these datasets was that of mechanism 'd_1', which is the same one determined by the researchers. However, this equation was only proposed as a possibility for a tolerance value equal and above 0.4 and 0.3, for $T=275$ °C and $T=285$ °C, respectively. This information is present in Table 21, while the constants and ranking criteria of the best curves of the equation for the datasets of $T=275$ °C and $T=285$ °C are present in Table 22 and Table 23 and graphically represented in Figure 28 and Figure 29, respectively.

Table 21 - Proposed rate equations for the 275 and 285 °C ethanol dehydrogenation datasets, for each tolerance value

Tolerance	Proposed rate equations: 275 °C	Proposed rate equations: 285 °C
0.1	(none)	(none)
0.2	(none)	(none)
0.3	(none)	d_1
0.4	d_1	d_1
0.5	d_1	d_1

Table 22 - Constants and ranking criteria for the best curve of the proposed rate equation for the 275 °C ethanol dehydrogenation dataset (A=EtOH)

ID	Rate equation	k' (mol·h ⁻¹ ·g ⁻¹)	K _A (atm ⁻¹)	MSE	SRE
d_1	$r_0 = \frac{k' \cdot K_A \cdot y_A \cdot P}{(1 + K_A \cdot y_A \cdot P)^2}$	4.2	0.5	4.31E-3	3.21E-1

Mechanism d_1: ['All species adsorb molecularly', 'Surface reaction controlling']

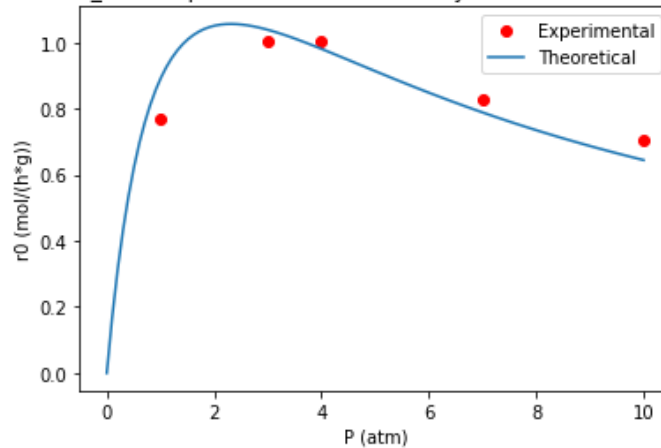


Figure 28 - Best curve of the rate equation of mechanism d_1 proposed for the 275 °C ethanol dehydrogenation dataset

Table 23 - Constants and ranking criteria for the best curve of the proposed rate equation for the 285 °C ethanol dehydrogenation dataset (A=EtOH)

ID	Rate equation	k' (mol·h ⁻¹ ·g ⁻¹)	K _A (atm ⁻¹)	MSE	SRE
d_1	$r_0 = \frac{k' \cdot K_A \cdot y_A \cdot P}{(1 + K_A \cdot y_A \cdot P)^2}$	5.5	0.5	1.03E-2	2.94E-1

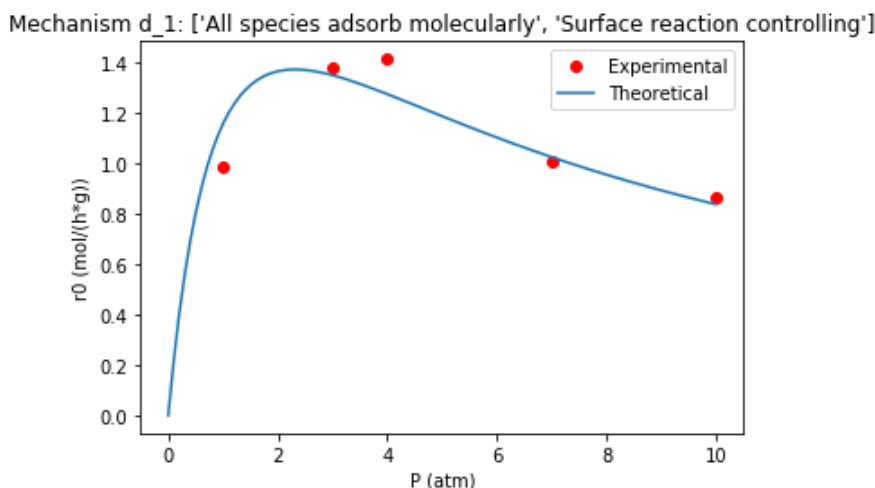


Figure 29 - Best curve of the rate equation of mechanism d_1 proposed for the 285 °C ethanol dehydrogenation dataset

With these datasets the tool was able to propose the correct rate equation, with a curve which equilibrium constants had the same order of magnitude as the ones determined by the researchers (present in Table 16), even though it was not proposed for tolerances below certain values. However, it can be said that the tool worked properly.

5.3.2. Discussion

- Low variability in data:

By analysing the features recognized in the 225, 235 and 250 °C datasets in Figure 24, we can see that these features were extracted recognizing the rise in initial rate values for small pressure values as noise. This is because the algorithm that extracts the features was developed to deal with experimental data, which is expected to have some noise associated. Since the maximum of each of these datasets was very subtle, i.e., since there was not enough variability to the initial rate values in that pressure range, a near-constant value curve was recognized for this pressure range, which resulted in experimental features that were not similar to any of the features from the theoretical curves. Therefore, if a dataset does not clearly represent the data trend, the recognized features may not reflect the important trends in the data. One proposed way of solving this was to add the point (0, 0) to the experimental dataset. However, this was not a good solution, since adding this point can alter the recognized features to ones which do not represent the experimental data well. An example of this is present in Figure 30.

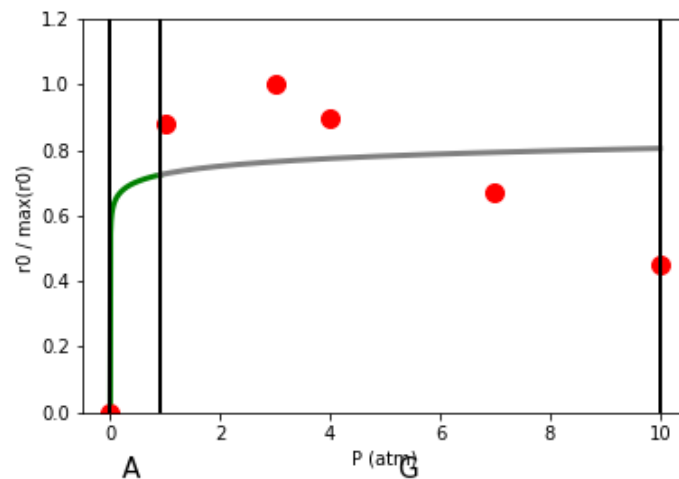


Figure 30 - Graphical representation of the recognized features of the 225 °C ethanol dehydrogenation dataset with the (0, 0) point added

- Effect of experimental noise

When analysing the features recognized from the 265 °C dataset (Figure 25), it can be observed that the recognized features do not accurately represent the data trends, mostly due to the fact that the point at $P=10$ atm has too much noise associated to it, and this is enhanced by the low variability of the dataset as a whole. The proposed rate equation is still the same that the researchers determined, but the proposed best curve of this rate equation does not properly follow the experimental data trends, as can be seen in Figure 26 and by comparing the equilibrium constant values from the proposed curve (Table 19) to the ones determined by the researchers (Table 16). One way of solving this is by eliminating any outliers in the dataset, be it manually by the researcher or possibly automatically done by a future version of the tool

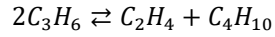
- Effect of tolerance value

When it comes to the 275 and 285 °C datasets, the tool worked properly, since it proposed the same rate equation that the researchers proposed. However, this only happened when the tolerance level was increased, while before that no rate equations generated curves with features similar enough to those of the experimental data to be proposed. This is because the determined value of K_{EtOH} is relatively small, and the shape of a curve is more sensitive to a variation over small values of equilibrium constants than to the same variation over bigger values.

In conclusion, the ideal dataset, when it comes to the performance of the tool, is one where the features of the experimental curve can be clearly identified and where the experimental noise is not enough to lead the tool to misidentify the features.

5.4. Propylene disproportionation – Influence of repeated tests and noise

Another performed case study was the one regarding the disproportionation of propylene, which has the following global reaction:



Eq. 24

The initial rate vs total pressure datasets of this reaction were recovered from the work of Luckner and Wills [26]. Three datasets were extracted from this source, each one at a different temperature: 399, 427 and 454 °C. These datasets were extracted for a feed purely composed of propylene. A graphical representation of the datasets is present in Figure 31, while the actual values are in Appendix 3.

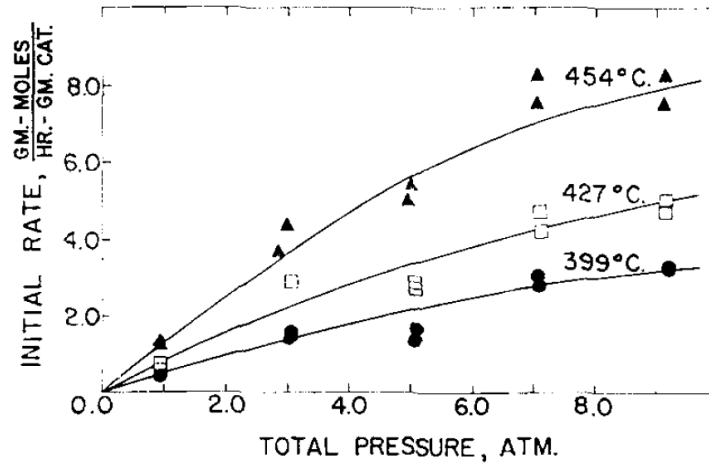


Figure 31 - Experimental datasets of the propylene disproportionation reaction [26] (markers – experimental points; solid line – model fitting). Reprinted with permission from [26]. Copyright 1973 Published by Elsevier Inc.

As it can be seen in Figure 31, these datasets have the particularity to have repeated experiments, which results in each point being paired with another one with very similar pressure and initial rate values. They also have experimental noise associated with them, more than in any of the previous case studies. The influence of these factors in the performance of the tool will be analysed.

The researchers who obtained these datasets determined that one of the possible rate equations that best fits the data is of the following type:

$$r_0 = \frac{k' \cdot (K_{C_3H_6} \cdot P_{C_3H_6})^2}{(1 + K_{C_3H_6} \cdot P_{C_3H_6})^2} \quad \text{Eq. 25}$$

This rate equation assumes that “a dual-site surface reaction was the rate-controlling step in the reaction mechanism” [26]. For this rate reaction, the determined kinetic and equilibrium constants are the ones present in Table 24.

Table 24 - Kinetic and equilibrium constants for the datasets of the propylene disproportionation reaction determined by Luckner and Wills [26].

Temperature (°C)	k' (g·mol·g ⁻¹ ·h ⁻¹)	$K_{C_3H_6}$ (atm ⁻¹)
399	5.1253	0.3933
427	7.9498	0.4087
454	12.5608	0.5116

Eq. 25 is the one resultant from mechanism 'd' (C_3H_6 adsorbs molecularly and the RDS is the reaction step on the catalyst surface). Therefore, if the tool works properly for these datasets then this rate equation will be the top proposed one, or at least among the ones proposed, and the constants of the best curves will be as close as possible to the ones in Table 24.

5.4.1. Test results – Full datasets

Each dataset was given to the tool, and all the user input regarding the reactants, products and inert species was given as well. The recognized features for each dataset are graphically represented in Figure 32 and present in Table 25.

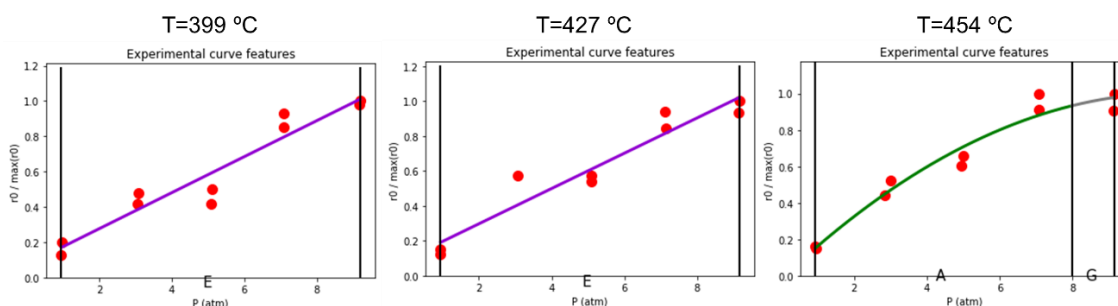


Figure 32 - Graphical representation of the recognized features of the propylene disproportionation datasets

Table 25 - Recognized features of the propylene disproportionation datasets

T (°C)	399		427		454		
Primitives	E		E		A, G		
Extremes (atm)	0.950	9.200	0.949	9.153	0.943	7.979	9.138

For the 399 and 427 °C datasets, the proposed rate equations do not depend on the tolerance value chosen, since both datasets have a set of primitives composed only by one primitive. For the 454 °C dataset, despite having a set of recognized primitives composed by two primitives, the proposed rate equations were the same for every tolerance value used. The constants and criteria of these mechanisms are present in Table 26, Table 27 and Table 28 for the 399, 427 and 454 °C datasets, respectively. The best curve of each proposed rate equation for each dataset are present in Figure 33, Figure 34 and Figure 35.

Table 26 - Constants and ranking criteria for the best curve of each proposed rate equation for the 399 °C propylene disproportionation dataset ($A=C_3H_6$)

ID	Rate equation	k' ($g \cdot mol \cdot g^{-1} \cdot h^{-1} \cdot atm^{-1}$)	MSE	SRE
a_1	$r_0 = k' \cdot y_A \cdot P$	3.6E-1	1.03E-1	0
e_1	$r_0 = k' \cdot y_A \cdot P$	3.6E-1	1.03E-1	0

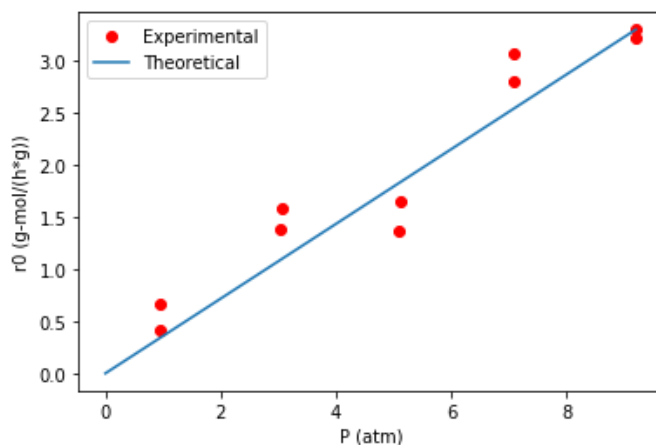


Figure 33 - Best curve of the rate equation of mechanisms a_1 and e_1 proposed for the 399 °C propylene disproportionation dataset

Table 27 - Constants and ranking criteria for the best curve of each proposed rate equation for the 427 °C propylene disproportionation dataset ($A=C_3H_6$)

ID	Rate equation	k' ($g \cdot mol \cdot g^{-1} \cdot h^{-1} \cdot atm^{-1}$)	MSE	SRE
a_1	$r_0 = k' \cdot y_A \cdot P$	5.7E-1	2.43E-1	0
e_1	$r_0 = k' \cdot y_A \cdot P$	5.7E-1	2.43E-1	0

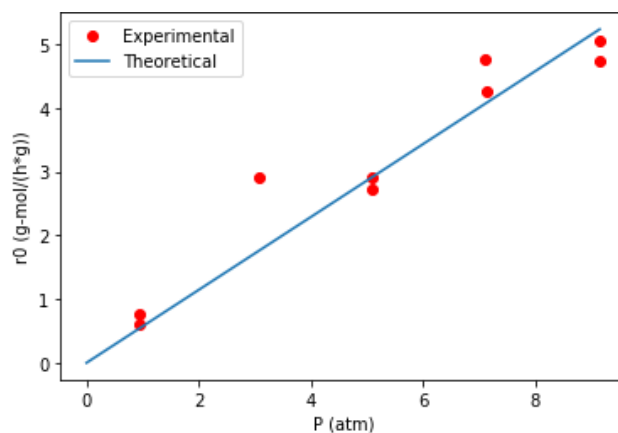


Figure 34 - Best curve of the rate equation of mechanisms a_1 and e_1 proposed for the 427 °C propylene disproportionation dataset

Table 28 - Constants and ranking criteria for the best curve of each proposed rate equation for the 454 °C propylene disproportionation dataset ($A=C_3H_6$)

ID	Rate equation	k' ($g \cdot mol \cdot g^{-1} \cdot h^{-1}$)	K_A (atm^{-1})	MSE	SRE
d_1	$r_0 = \frac{k' \cdot (K_A \cdot y_A)^2 \cdot P^2}{(1 + K_A \cdot y_A \cdot P)^2}$	1.5E+1	0.3	3.93E-1	9.34E-2

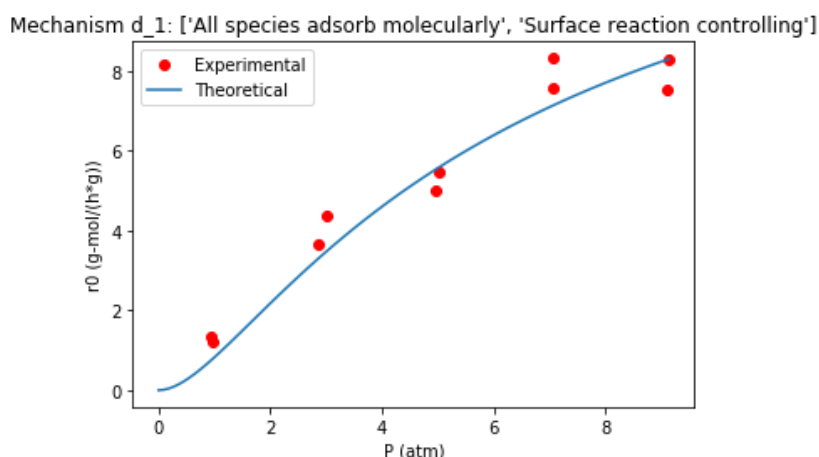


Figure 35 - Best curve of the rate equation of mechanism d_1 proposed for the 454 °C propylene disproportionation dataset

As it can be seen in Table 26 and Table 27, the best proposed rate equations for the 399 and 427 °C datasets were those of mechanism a_1 (C_3H_6 adsorbs molecularly and the RDS is the adsorption of C_3H_6 to the catalyst surface) and mechanism e_1 (C_3H_6 dissociates on adsorption and the RDS is the adsorption of C_3H_6 to the catalyst surface). These mechanisms resulted in the same proposed initial rate curve since they resulted in the same rate equation and proposed k' value. The rate equation of these mechanisms is not the one proposed by the researchers (Eq. 25). In fact, the mechanism which results in Eq. 25, mechanism 'd_1', is proposed as a possibility only for the 454 °C dataset, where it is the only mechanism proposed. It can be concluded that the tool only worked properly when analysing the 454 °C dataset.

5.4.2. Test results – Averaged datasets

The effect of the doubled points in the dataset was also a target of this case study. Therefore, the three previous datasets were manually treated by taking each pair of points and turning them into their average, both in terms of pressure and initial rate values. The tool then analysed these new datasets, which will be referred from now on as “averaged” datasets. The recognized features of these datasets are graphically represented in Figure 36 and present in Table 29.

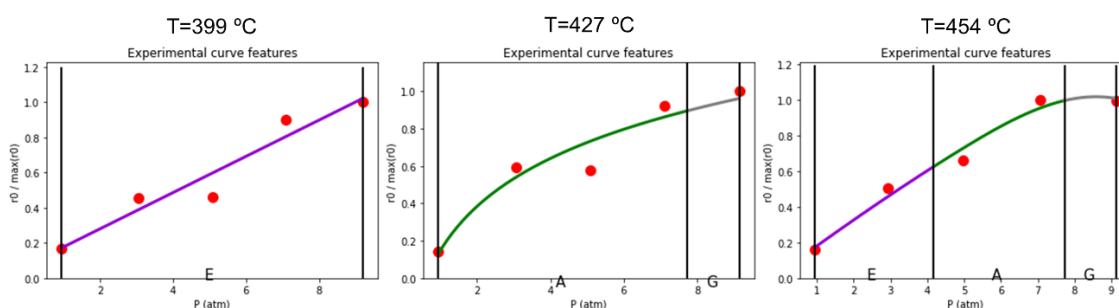


Figure 36 - Graphical representation of the recognized features of the averaged propylene disproportionation datasets

Table 29 - Recognized features of the averaged propylene disproportionation datasets

T (°C)	399		427			454			
Primitives	E		A, G			E, A, G			
Extremes (atm)	0.955	9.194	0.949	7.740	9.148	0.951	4.172	7.723	9.127

The proposed rate equations for these datasets were always the same, regardless of the tolerance value chosen. The constants and ranking criteria of the best curves of the rate equations proposed for the 399 and 427 °C averaged datasets are present in Table 30 and Table 31, respectively. The best curves of each proposed rate equation of each dataset are present in Figure 37 and Figure 38.

Table 30 - Constants and ranking criteria for the best curve of each proposed rate equation for the averaged 399 °C propylene disproportionation dataset (A=C₃H₆)

ID	Rate equation	k'	K _A (atm ⁻¹)	MSE	SRE
a_1	$r_0 = k' \cdot y_A \cdot P$	3.6E-1 g·mol·g ⁻¹ ·h ⁻¹ ·atm ⁻¹	N/A	9.39E-2	0
e_1	$r_0 = k' \cdot y_A \cdot P$	3.6E-1 g·mol·g ⁻¹ ·h ⁻¹ ·atm ⁻¹	N/A	9.39E-2	0
d_1	$r_0 = \frac{k' \cdot (K_A \cdot y_A)^2 \cdot P^2}{(1 + K_A \cdot y_A \cdot P)^2}$	1.4E+1 g·mol·g ⁻¹ ·h ⁻¹	0.1	1.88E-1	0

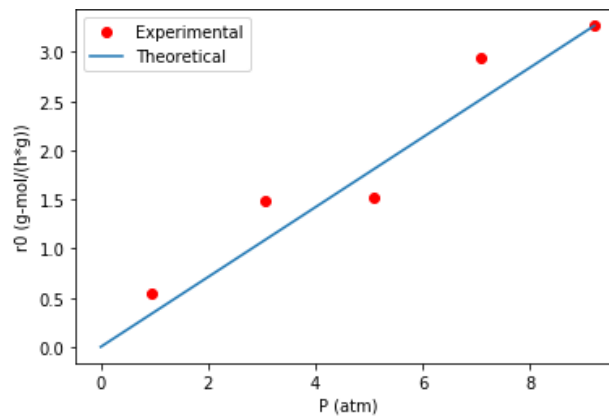


Figure 37 - Best curve of the rate equation of mechanisms a_1 and e_1 proposed for the averaged 399 °C propylene disproportionation dataset

Table 31 - Constants and ranking criteria for the best curve of each proposed rate equation for the averaged 427 °C propylene disproportionation dataset (A=C₃H₆)

ID	Rate equation	k' (g·mol·g ⁻¹ ·h ⁻¹)	K _A (atm ⁻¹)	MSE	SRE
d_1	$r_0 = \frac{k' \cdot (K_A \cdot P_A)^2}{(1 + K_A \cdot P_A)^2}$	9.1	0.3	2.09E-1	2.14E-2

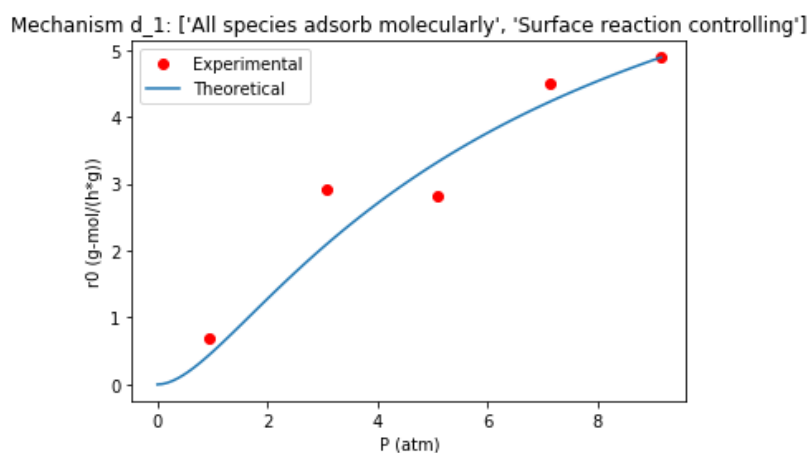


Figure 38 - Best curve of the rate equation of mechanism d_1 proposed for the averaged 427 °C propylene disproportionation dataset

When it comes to the 454 °C dataset, however, the tool did not generate any theoretical curve with a set of primitives that matched the experimental ones

As it can be seen in Table 30, the rate equations of mechanisms 'a_1' and 'e_1' are still proposed by the tool as the ones that generate the curves with the most similar features to those of the experimental data, among all of the proposed ones for the averaged 399 °C dataset. However, the rate equations of mechanism 'd_1' is now also proposed as a possibility for this dataset (curve present in Figure 39) and is the only one proposed for the 427°C dataset.

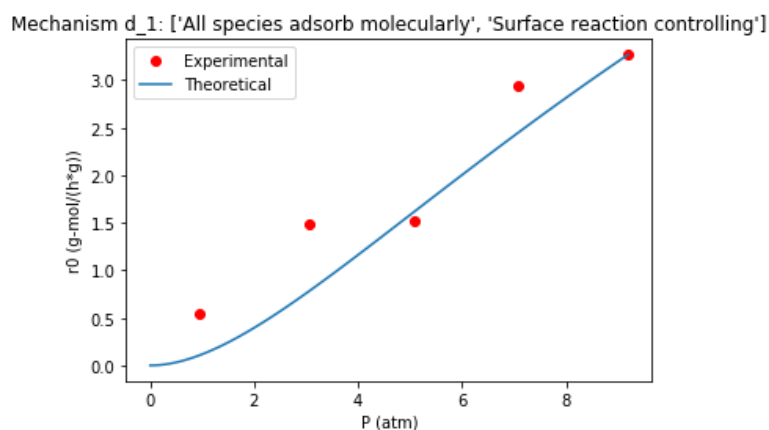


Figure 39 - Best curve of rate equation of mechanism d_1 proposed for the averaged 399 °C propylene disproportionation dataset

5.4.3. Discussion

Comparing the results of the test performed with the regular experimental data with the ones from the test with the averaged datasets, it can be seen that taking the average of each pair of points had different results for each temperature.

For 399 °C, the recognized features were very similar in both cases. This is because this dataset has much noise associated but also because the trend of the data itself is very near to being linear, even though it is not. However, while using the full dataset resulted in the rate equation of mechanism 'd_1'

not being proposed, using the averaged dataset made the tool propose this equation as a possibility. This is because the pressure values were reduced to half, and since the features of the theoretical curves are extracted using points at the same pressure values as the experimental data, there was at least one curve of the rate equation of mechanism 'd_1' with matching primitives (Figure 39). However, the features of the proposed curve of this equation, although similar to the ones recognized for the experimental dataset (Figure 36), do not reflect the trend recognized by the researchers (Figure 31). Therefore, using the averaged dataset did not bring many improvements to the tool performance.

For 427 °C, the features recognized for the averaged dataset are more similar to the ones determined by the researchers than the ones recognized by the tool for the full dataset. Not only that, but they are similar to what would be expected for a curve of the rate equation of mechanism 'd_1'. Therefore, this equation was proposed as a possibility for the averaged dataset. Therefore, it can be concluded that the tool worked better for the averaged dataset than for the full one, in this case.

For 454 °C, the tool still recognized the same stabilizing trend for the averaged dataset as it did for the full dataset. However, it also recognized in the range between 0.951 and 4.172 atm a linear trend, which resulted in a set of primitives with no match from any theoretical curve. The reason for this to happen was the noise present in the experimental data, specifically in the point at 4.982 atm.

In conclusion, the performance of the tool with these datasets was worsened by the noise associated with them, especially with the pairs of points at pressure values around 5 atm, which was larger than any dataset from the previous case studies. A possible solution for this is, as mentioned before, the elimination of outliers, be it manually by the researcher or possibly automatically done by a future version of the tool. Regarding the utilization of datasets with repeated experiments, there are advantages and disadvantages: while using the full pressure values will make it so that the features of the theoretical data are more correctly recognized, the recognition of the features of the experimental data will be more sensitive to noise in the data, leading to misidentified features, as was the case with the 427 °C dataset.

5.5. Propyne hydrogenation – Determined rate equation not in the library

The final case study regarded the reaction of propyne hydrogenation, which happens according to the following global reaction:



The dataset for this case study was recovered from the work of Fritsch et al.[27]. These researchers obtained the dataset at a temperature of 308K and a feed composed of H₂, C₃H₄ and Ar, which acts as an inert species, with the respective molar fractions being 0.03, 0.15 and 0.82. The dataset is graphically represented in Figure 40, while the actual values are present in Appendix 3.

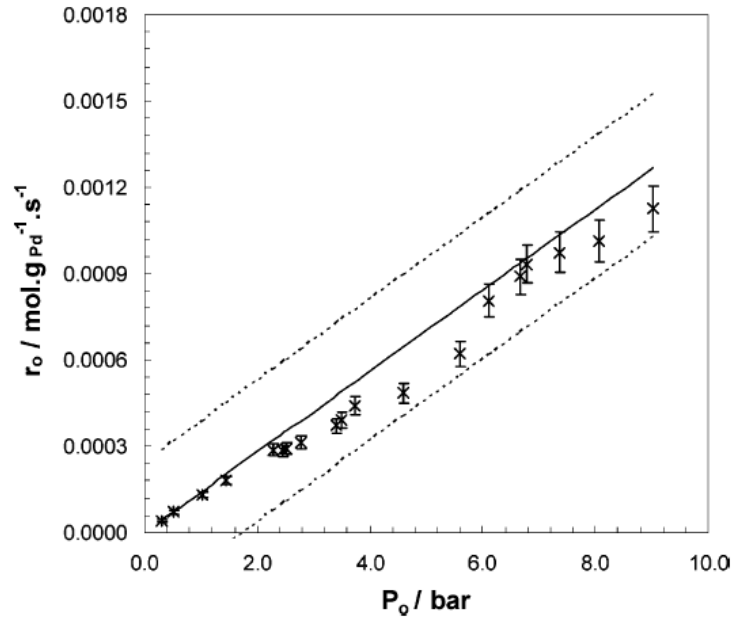
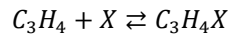
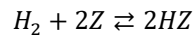
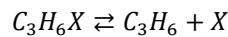
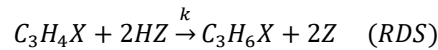


Figure 40 - Experimental dataset of the propyne hydrogenation reaction[27] (markers – experimental points; vertical bars – error of each point; solid line – model fitting; dashed lines – 95% confidence curves). Reprinted with permission from [27]. Copyright 2007 American Chemical Society.

The researchers who extracted this dataset determined that the mechanism (Eq. 27) and the initial rate equation (Eq. 28) that best suited it were the following:



Eq. 27



$$r_0 = \frac{k' \cdot K_{C_3H_4} \cdot P_{C_3H_4} \cdot K_{H_2} \cdot P_{H_2}}{(1 + K_{C_3H_4} \cdot P_{C_3H_4})(1 + \sqrt{K_{H_2} \cdot P_{H_2}})^2} \quad \text{Eq. 28}$$

As it can be seen in Eq. 27 and Eq. 28, this case study is particular, since this reaction uses a bifunctional catalyst, where the molecules of H_2 adsorb on metal sites, while the molecules of C_3H_4 adsorb on a different type of site[27]. The tool, as of now, does not support catalytic reactions that occur on bifunctional catalysts, so the resulting initial rate equation (Eq. 28) is not present in the library of the tool. The kinetic and equilibrium constants determined by the researchers are present in Table 32.

Table 32 - Kinetic and equilibrium constants for the dataset of the propyne hydrogenation reaction determined by Fritsch et al.[27].

k' (mol·g ⁻¹ ·s ⁻¹)	4.71E+02
$K_{C_3H_4}$ (bar ⁻¹)	8.70E+02
K_{H_2} (bar ⁻¹)	1.00E-05

5.5.1. Test results

This dataset was given to the tool as input, as well as all the required user input regarding the reactants, the products, and the inert species. The features recognized from the experimental dataset are graphically represented in Figure 41 and present in Table 33.

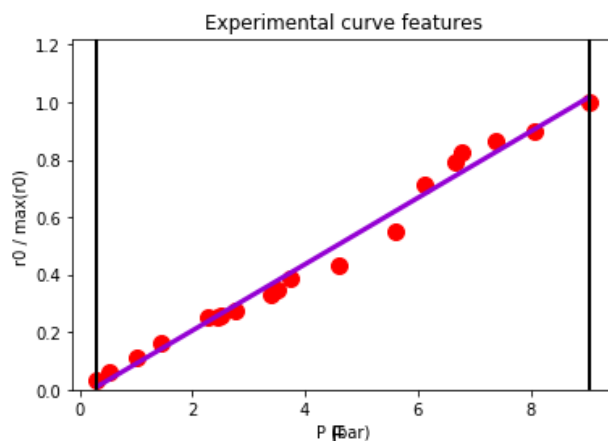


Figure 41 - Graphical representation of the recognized features of the propyne hydrogenation dataset

Table 33 - Recognized features of the propyne hydrogenation dataset

Primitives:	E	
Extremes (bar):	0.310	9.036

The set of primitives recognized for this dataset contains only one primitive, so all of the extremes of the theoretical curves will be the same as the ones in Table 33. Therefore, different tolerance values will result in the same proposed rate equations. For this dataset all the rate equations were proposed, for both permutations of reactants, except for the rate equation of mechanism 'r' (uncatalyzed reaction). The constants and ranking criteria of the top 5 proposed rate equations are present in Table 34, and the best curve of the top proposed rate equation in Figure 42.

Table 34 - Constants and ranking criteria for the best curve of the top 5 proposed rate equation for the propyne hydrogenation dataset (mechanisms with _1: A=H₂, B=C₃H₄; mechanisms with _2: A=C₃H₄, B=H₂)

ID	Rate equation	k' (mol·g ⁻¹ ·s ⁻¹)	K _A (bar ⁻¹)	K _B (bar ⁻¹)	MSE	SRE
n_2	$r_0 = \frac{k' \cdot K_A \cdot y_A \cdot y_B \cdot P^2}{(1 + \sqrt{K_A \cdot y_A \cdot P})^2}$	5.0E-03	100	N/A	1.60E-09	0
k_1	$r_0 = \frac{k' \cdot K_A \cdot y_A \cdot y_B \cdot P^2}{1 + K_A \cdot y_A \cdot P}$	8.9E-04	100	N/A	1.64E-09	0
o_2	$r_0 = \frac{k' \cdot K_B \cdot y_B \cdot y_A \cdot P^2}{1 + K_B \cdot y_B \cdot P}$	8.9E-04	N/A	100	1.64E-09	0
k_2	$r_0 = \frac{k' \cdot K_A \cdot y_A \cdot y_B \cdot P^2}{1 + K_A \cdot y_A \cdot P}$	4.2E-03	100	N/A	1.83E-09	0
o_1	$r_0 = \frac{k' \cdot K_B \cdot y_B \cdot y_A \cdot P^2}{1 + K_B \cdot y_B \cdot P}$	4.2E-03	N/A	100	1.83E-09	0

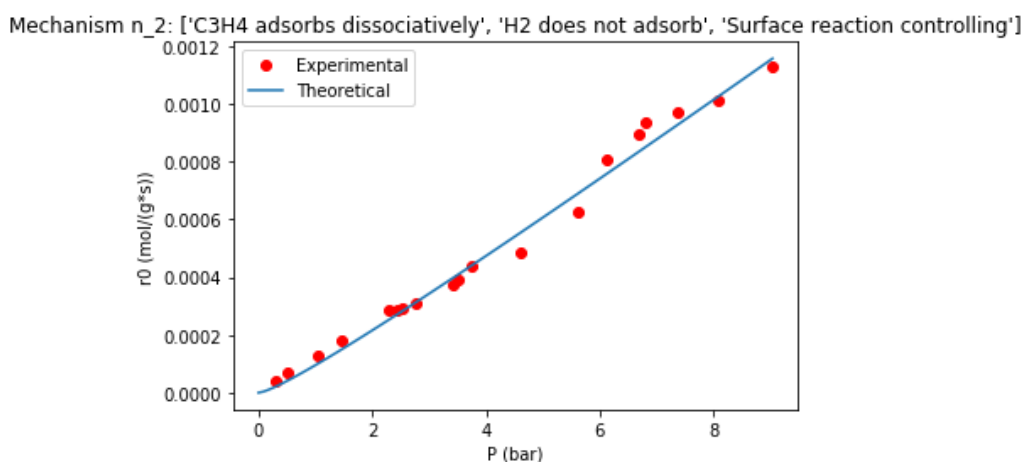


Figure 42 - Best curve of the rate equation of mechanism n_2 proposed for the propyne hydrogenation dataset

5.5.2. Discussion

As mentioned before, the rate equations of all the mechanism on the library, with exception of 'r' (uncatalyzed reaction) were proposed as possible models to describe the dataset. This is because all these mechanisms have at least one theoretical curve whose features are the same as those of the experimental dataset since the latter were in fact very simple. Therefore, the tool does not perform well when analysing this dataset.

By analysing Table 32 it can be observed that the both the value of K_{H_2} and the molar fraction of H_2 are very small. If they are considered negligible when compared to the rest of the denominator, then the initial rate equation will approximate to the following:

$$r_0 = \frac{k' \cdot K_{C_3H_4} \cdot P_{C_3H_4} \cdot P_{H_2}}{1 + K_{C_3H_4} \cdot P_{C_3H_4}} \quad \text{Eq. 29}$$

This rate equation is the one that results from mechanism 'k_2', where H_2 does not adsorb and the RDS is the reaction step on the catalyst surface. The value of $K_{C_3H_4}$ of the proposed curve of this rate equation (present in Table 34) is in the same order of greatness, although the actual value determined by the researchers is out of the range of the list of K_{Eq} values of the tool. Therefore, it can be said that the tool, in a way, could propose a rate equation that is a good approximation to the data, even if it does not have the actual rate equation in its library.

5.6. Conclusions

From the analysis of the case studies results it was possible to conclude that the tool had a satisfactory performance for most cases, since it was able to propose the same rate equations determined by the researchers for most cases, with kinetic and equilibrium constants similar to the ones determined by the researchers. In fact, in the cases that the tool was not successful this was due to the overall quality of the data. From those cases, it can be concluded that the tool can work properly for datasets that have a distribution of points that covers the pressure ranges with the important features, since the feature extraction algorithm will recognize the features that best describe the data trends more easily. The tool also works better for datasets that do not have too much noise associated with them. In these case

studies the tool worked properly for datasets with noise up to ca. 17%, but this can change with the point distribution of said data.

To combat the negative effect of noise, a proposed solution is the removal of outliers, either manually by the researcher or automatically by the tool, in an eventual future version of the tool where the algorithms to do that are implemented. A possible application of the latter approach is to insert a function in the feature extraction algorithm that determines outlier points through the interquartile range rule[28].

The existence of repeated experiments explicitly in the dataset will also make it difficult for the features extraction algorithm to recognize the correct features, if the dataset has noise associated with it. When dealing with repeated experiments, the best approach is to take the average of each set of repeated experiments. This generally leads to an improvement of the performance of the tool, as can be seen in the propylene disproportionation case study.

6. Conclusions and future work

A tool that automatically proposes initial rate equations for catalytic reactions, based on experimental data, namely initial rate as function of pressure, was developed. This tool contains a library of typical rate equations in heterogeneous catalysis, from which theoretical curves are generated. Then, the tool screens the rate equations its library based on the similarities between the features of the experimental data and the features of the theoretical curves generated from each rate equations. This way, only the rate equations that can generate curves with similar features to those observed in the experimental data are proposed as possible models. Once developed, the tool was tested. For this, five case studies were performed. From these case studies it can be concluded that this goal was achieved, since the tool performed well in most cases. When the tool did not perform so well, this was related to the characteristics of the data, more specifically the excessive amount of experimental noise, the presence of outliers, the lack variability of the data and the distribution of data points on pressure ranges that did not reflect meaningful features.

One of the achievements of this work was the fact that the tool can, in most cases, propose the same rate equations that an experienced researcher would, needing less time to do so. Another achievement is the fact that the tool can propose these equations based on typical catalytical kinetic datasets, which usually contain a small number of points. Therefore, the tool can make meaningful conclusions from relatively small datasets. Important to note as well is that the generic rate equation in the library were deduced from typical mechanisms for catalytic reactions in such a way that the library covers the whole range of possibilities, inside certain limitations and assumptions.

However, this tool has its limitations. The main one is the fact that it can only deal with initial rates of reaction. This limits the applicability of the tool since it reduces the number of experimental datasets it can analyse. Another limitation is the high sensitivity of feature extraction algorithm to outliers in the experimental data. As the feature extraction algorithm does not take their presence into account, the generated curve will be influenced by its presence, thus altering the recognized features from the data.

In the future, several developments could be made to the tool to mitigate its limitations. The influence of outliers on the performance of the tool could be mitigated by developing an algorithm that automatically removes outliers from data. The initial rates limitations could be surpassed by updating the library to be able to consider the presence of products in the feeds or in the reactor after conversion, so data other than initial rate of reaction can be inputted to the tool. The fully developed version of the tool should ideally analyse datasets of rate as a function of not only total pressure but other variables as well, such as molar fraction of reactants. This way, the tool would extract all the kinetic information from as many types of data as possible, which would improve the screening of the models.

The tool developed in this work is an important initial step for the automatization of the kinetic modelling process. Once fully developed and perfected, this tool will enable a larger number of researchers to achieve models for catalytic reactions in less time-consuming manner. This way, the design of new catalysts will be enhanced, which will be vital to the sustainability of the chemical industry.

References

- [1] A. D. McNaught and A. Wilkinson, "IUPAC. Compendium of Chemical Terminology," *Blackwell Scientific Publications*, 1997. <https://doi.org/10.1351/goldbook.C00876> (accessed Dec. 06, 2020).
- [2] M. A. Vannice, *Kinetics of catalytic reactions*. Springer Science & Business Media, 2005.
- [3] K. Kakaei, M. D. Esrafil, and A. Ehsani, "Introduction to Catalysis," in *Interface Science and Technology*, vol. 27, Academic Press, 2019, pp. 1–21.
- [4] O. A. Hougen and K. M. Watson, "Solid catalysts and Reaction rates," *Ind. Eng. Chem.*, vol. 35, no. 5, pp. 529–541, 1943.
- [5] L. P. De Oliveira, D. Hudebine, D. Guillaume, and J. J. Verstraete, "A Review of Kinetic Modeling Methodologies for Complex Processes," *Oil Gas Sci. Technol. – Rev. IFP Energies Nouv.*, vol. 71, no. 45, 2016, doi: 10.2516/ogst/2016011.
- [6] M. E. Davis and R. J. Davis, *Fundamentals of Chemical Reaction Engineering*. 2013.
- [7] J. W. Thybaut and G. B. Marin, "Single-Event MicroKinetics: Catalyst design for complex reaction networks," *J. Catal.*, vol. 308, pp. 352–362, 2013, doi: 10.1016/j.jcat.2013.08.013.
- [8] L. Pirro, P. S. F. Mendes, S. Paret, B. D. Vandegheuchte, G. B. Marin, and J. W. Thybaut, "Descriptor-property relationships in heterogeneous catalysis : exploiting synergies between statistics and fundamental kinetic modelling," *Catal. Sci. Technol.*, vol. 9, no. 12, pp. 3109–3125, 2019.
- [9] L. Pirro, P. S. F. Mendes, B. Kemseke, B. D. Vandegheuchte, G. B. Marin, and J. W. Thybaut, "From catalyst to process : bridging the scales in modeling the OCM reaction," *Catal. Today*, no. March, pp. 1–11, 2020, doi: 10.1016/j.cattod.2020.06.084.
- [10] E. W. Lund, "Guldberg and Waage and the Law of Mass Action," *J. Chem. Educ.*, vol. 42, no. 10, pp. 548–550, 1965, doi: 10.1021/ed042p548.
- [11] R. W. Missen, C. A. Mims, and B. A. Saville, *Introduction to Chemical Reaction Engineering and Kinetics*, 1st ed. New York: John Wiley & Sons, 1999.
- [12] P. S. F. Mendes, S. Siradze, L. Pirro, and J. W. Thybaut, "Open data in catalysis : from today's big picture to the future of small data," *ChemCatChem*, doi: 10.1002/cctc.202001132R3.
- [13] R. Van De Vijver, B. R. Devocht, K. M. Van Geem, J. W. Thybaut, and G. B. Marin, "Challenges and opportunities for molecule-based management of chemical processes," *Curr. Opin. Chem. Eng.*, vol. 13, pp. 142–149, 2016, doi: 10.1016/j.coche.2016.09.006.
- [14] C. W. Gao, J. W. Allen, W. H. Green, and R. H. West, "Reaction Mechanism Generator : Automatic construction of chemical kinetic mechanisms," *Comput. Phys. Commun.*, vol. 203, pp. 212–225, 2016, doi: 10.1016/j.cpc.2016.02.013.

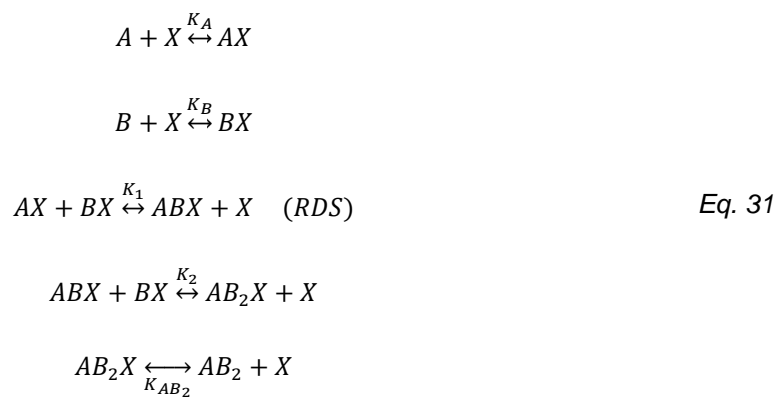
- [15] B. Bertok and L. Fan, "Review of methods for catalytic reaction-pathway identification at steady state," *Curr. Opin. Chem. Eng.*, vol. 2, pp. 487–494, 2013, doi: 10.1016/j.coche.2013.10.007.
- [16] C. F. Goldsmith and R. H. West, "Automatic Generation of Microkinetic Mechanisms for Heterogeneous Catalysis," *J. Phys. Chem. C*, vol. 121, no. 121, p. 9970–9981, 2017, doi: 10.1021/acs.jpcc.7b02133.
- [17] S. Katare, J. M. Caruthers, W. N. Delgass, and V. Venkatasubramanian, "An Intelligent System for Reaction Kinetic Modeling and Catalyst Design," *Ind. Eng. Chem. Res.*, vol. 43, no. 14, pp. 3484–3512, 2004, doi: 10.1021/ie034067h.
- [18] D. Schaich and R. King, "Qualitative modelling and simulation of chemical reaction systems," *Comput. Chem. Eng.*, vol. 23, pp. S415–S418, 1999, doi: 10.1016/S0098-1354(99)80102-1.
- [19] D. Schaich, R. Becker, and R. King, "Qualitative modelling for automatic identification of mathematical models of chemical reaction systems," *Control Eng. Pract.*, vol. 9, pp. 1373–1381, 2001.
- [20] S. Siradze, "Automated kinetic feature extraction from Open Access data," Universiteit Gent, 2019.
- [21] K. H. Yang and O. A. Hougen, "Determination of Mechanism of Catalyzed Gaseous Reactions," *Chem. Eng. Prog.*, vol. 46, no. 3, pp. 146–157, 1950.
- [22] M. E. Janusz and V. Venkatasubramanian, "Automatic Generation of Qualitative Descriptions of Process Trends for Fault Detection and Diagnosis," *Engng Applic. Artif. Intell.*, vol. 4, no. 5, pp. 329–339, 1991.
- [23] O. P. Ahuja, "Kinetics of catalytic oxidation of methane - Application of initial rate technique for mechanism determination," University of Windsor, 1966.
- [24] L. Brandão, D. Fritsch, L. M. Madeira, and A. M. Mendes, "Kinetics of propylene hydrogenation on nanostructured palladium clusters," *Chem. Eng. J.*, vol. 103, pp. 89–97, 2004, doi: 10.1016/j.cej.2004.07.008.
- [25] J. Franckaerts and G. F. Froment, "Kinetic study of the dehydrogenation of ethanol," *Chem. Eng. Sci.*, vol. 19, pp. 807–818, 1964, doi: 10.1016/0009-2509(64)85092-2.
- [26] R. C. Luckner, G. E. McConchie, and G. B. Wills, "Initial Rates of Propylene Disproportionation WO₃ on Silica Catalysts," *J. Catal.*, vol. 28, pp. 63–68, 1973.
- [27] L. Brandão, D. Fritsch, A. M. Mendes, and L. M. Madeira, "Propyne Hydrogenation Kinetics over Surfactant-Stabilized Palladium Nanoclusters," *Ind. Eng. Chem. Res.*, vol. 46, no. 2, pp. 377–384, 2007.
- [28] R. J. Freund, W. J. Wilson, and D. L. Mohr, "Data and Statistics," in *Statistical Methods*, 3rd ed., Academic Press, 2010, pp. 1–65.

Appendix 1 – Influence of assuming only one surface reaction step

This appendix will demonstrate how the assumption that the surface reaction occurs in one step can influence the overall rate equation when the RDS is a surface reaction step. To illustrate this the example of the heterogeneous catalytic reaction present in Eq. 30 will be used.



Assuming the reactants and product adsorb molecularly onto the catalyst surface and that each molecule of B reacts separately, the reaction mechanism is the one present in Eq. 31.



If it is assumed that the RDS is the first addition of B to A, as is explicit in Eq. 31, then the overall rate equation can be defined as Eq. 32.

$$r = k_1 \cdot [AX] \cdot [BX] - k_{-1} \cdot [ABX] \cdot [X] \quad \text{Eq. 32}$$

Defining the concentrations of intermediary species in terms of the concentrations of reactants and products through the RDS approximation and assuming initial rates (no products present in the system), Eq. 33 is obtained.

$$r_0 = k_1 \cdot K_A \cdot y_A \cdot K_B \cdot y_B \cdot P^2 \cdot [X]^2 \quad \text{Eq. 33}$$

The total number of active sites L is given by Eq. 34.

$$L = [X] + [AX] + [BX] + [ABX] + [AB_2X] \quad \text{Eq. 34}$$

Defining the concentrations of intermediary species in terms of the concentrations of reactants and assuming initial rates once again, Eq. 35 is obtained.

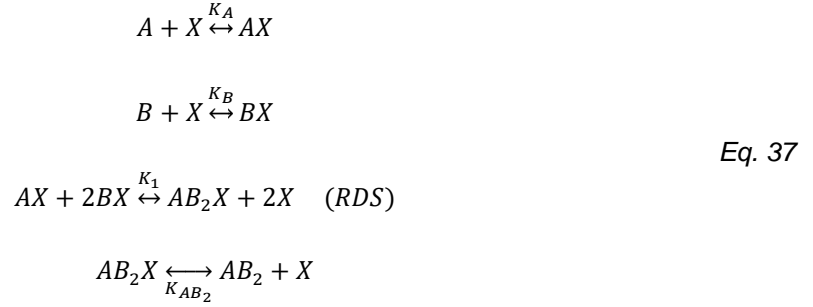
$$[X] = \frac{L}{1 + (K_A \cdot y_A + K_B \cdot y_B) \cdot P} \quad \text{Eq. 35}$$

Substituting [X] in Eq. 33 by its definition in Eq. 35, the initial rate equation present in Eq. 36 is obtained.

$$r_0 = \frac{k' \cdot K_A \cdot y_A \cdot K_B \cdot y_B \cdot P^2}{(1 + (K_A \cdot y_A + K_B \cdot y_B) \cdot P)^2} \quad \text{Eq. 36}$$

k' represents the product of k_1 with the number of total active sites and the probability of the reactant molecules involved in the RDS adsorbing into adjacent active sites.

However, if it is assumed that the addition of B molecules to A all happen in one step, the mechanism in Eq. 37 is obtained.



Assuming that the RDS is the surface reaction step, then the overall rate equation can be defined as Eq. 38.

$$r = k_1 \cdot [AX] \cdot [BX]^2 - k_{-1} \cdot [AB_2X] \cdot [X]^2 \quad \text{Eq. 38}$$

Defining the concentrations of intermediary species in terms of the concentrations of reactants and products through the RDS approximation and assuming initial rates Eq. 39 is obtained.

$$r_0 = k_1 \cdot K_A \cdot y_A \cdot (K_B \cdot y_B)^2 \cdot P^3 \cdot [X]^3 \quad \text{Eq. 39}$$

The total number of active sites for this scenario is given by Eq. 40.

$$L = [X] + [AX] + [BX] + [AB_2X] \quad \text{Eq. 40}$$

Defining the concentrations of intermediary species in terms of the concentrations of reactants and assuming initial rates, Eq. 35 is obtained once again. Substituting the definition of $[X]$ from that equation into Eq. 39, the initial rate equation for this scenario, present as Eq. 41, is obtained.

$$r_0 = \frac{k' \cdot K_A \cdot y_A \cdot (K_B \cdot y_B)^2 \cdot P^3}{(1 + (K_A \cdot y_A + K_B \cdot y_B) \cdot P)^3} \quad \text{Eq. 41}$$

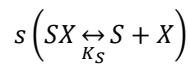
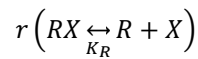
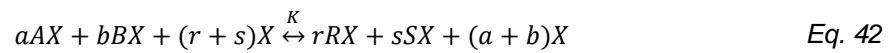
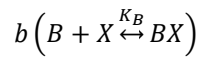
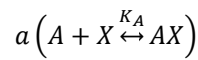
Comparing Eq. 36 with Eq. 41 it can be observed that the pressure exponent in the numerator and the exponent numerator are different. This is due to the different number of molecules of each reactant and active sites involved in the RDS, respectively, on each scenario. Therefore, it can be concluded that the number of surface reaction steps considered has influence on the overall rate equation whenever one of these steps is the RDS.

Appendix 2 – Deduction of generic initial rate equations

This appendix contains the deductions that led to the generic initial rate equations present in the library. These rate equations were all deduced using the RDS approximation and assuming initial rates (no products exist in the system). Another assumption made was that all products always adsorbed molecularly onto the catalyst surface.

Important to note that k' represents the lumped kinetic constant, i.e., the product of the kinetic constant of the RDS with the number of total active sites L and the probability of the reactant molecules involved in the RDS adsorbing into adjacent active sites.

- **A: All species adsorb molecularly**



$$L = [X] + [AX] + [BX] + [RX] + [SX] \quad \text{Eq. 43}$$

$$K_G = \frac{K \cdot K_A^a \cdot K_B^b}{K_R^r \cdot K_S^s} \quad \text{Eq. 44}$$

- **a – Adsorption of A controlling**

$$r = k_A \cdot y_A \cdot P \cdot [X] - k_{-A} \cdot [AX] \quad \text{Eq. 45}$$

Writing concentrations of intermediate species as a function of reactant and product species and applying the initial rates assumption to Eq. 43 and Eq. 45, respectively:

$$[X] = \frac{L}{1 + K_B \cdot y_B \cdot P} \quad \text{Eq. 46}$$

$$r_0 = k_A \cdot y_A \cdot P \cdot [X] \quad \text{Eq. 47}$$

Substituting $[X]$ in Eq. 47 by its definition in Eq. 46:

$$r_0 = \frac{k' \cdot y_A \cdot P}{1 + K_B \cdot y_B \cdot P} \quad \text{Eq. 48}$$

- **b – Adsorption of B controlling**

$$r = k_B \cdot y_B \cdot P \cdot [X] - k_{-B} \cdot [BX] \quad \text{Eq. 49}$$

Writing concentrations of intermediate species as a function of reactant and product species and applying the initial rates assumption to Eq. 43 and Eq. 49, respectively:

$$[X] = \frac{L}{1 + K_A \cdot y_A \cdot P} \quad \text{Eq. 50}$$

$$r_0 = k_B \cdot y_B \cdot P \cdot [X] \quad \text{Eq. 51}$$

Substituting $[X]$ in Eq. 51 by its definition in Eq. 50:

$$r_0 = \frac{k' \cdot y_B \cdot P}{1 + K_A \cdot y_A \cdot P} \quad \text{Eq. 52}$$

- **c – Desorption of R controlling**

$$r = k_{-R} \cdot [RX] - k_R \cdot y_R \cdot P \cdot [X] \quad \text{Eq. 53}$$

- Reaction with 2 products (S exists):

Writing $[RX]$ as a function of reactant and product species:

$$[RX] = K_R \cdot \sqrt{\frac{K_G \cdot y_A^a \cdot y_B^b}{y_S^s}} \cdot P^{\frac{a+b-s}{r}} \cdot [X] \quad \text{Eq. 54}$$

Writing concentrations of intermediate species as a function of reactant and product species to Eq. 43 and Eq. 53, respectively:

$$[X] = \frac{L}{1 + K_R \cdot \sqrt{\frac{K_G \cdot y_A^a \cdot y_B^b}{y_S^s}} \cdot P^{\frac{a+b-s}{r}} + (K_A \cdot y_A + K_B \cdot y_B + K_S \cdot y_S) \cdot P} \quad \text{Eq. 55}$$

$$r = k_R \cdot \sqrt{\frac{K_G \cdot y_A^a \cdot y_B^b}{y_S^s}} \cdot P^{\frac{a+b-s}{r}} \cdot [X] - k_R \cdot y_R \cdot P \cdot [X] \quad \text{Eq. 56}$$

Substituting $[X]$ in Eq. 56 by its definition in Eq. 55:

$$r = \frac{k' \cdot \sqrt{\frac{K_G \cdot y_A^a \cdot y_B^b}{y_S^s}} \cdot P^{\frac{a+b-s}{r}} - k_R \cdot y_R \cdot P}{1 + K_R \cdot \sqrt{\frac{K_G \cdot y_A^a \cdot y_B^b}{y_S^s}} \cdot P^{\frac{a+b-s}{r}} + (K_A \cdot y_A + K_B \cdot y_B + K_S \cdot y_S) \cdot P} \quad \text{Eq. 57}$$

Applying the initial rates assumption – y_S approximates to 0, therefore the fraction inside the root approximates infinity, becoming much larger than any other monomer. Thus, Eq. 57 can be approximated to Eq. 58:

$$r_0 = \frac{k' \cdot \sqrt[r]{\frac{K_G \cdot y_A^a \cdot y_B^b}{y_S^s} \cdot P^{\frac{a+b-s}{r}}}}{K_R \cdot \sqrt[r]{\frac{K_G \cdot y_A^a \cdot y_B^b}{y_S^s} \cdot P^{\frac{a+b-s}{r}}}} = \frac{k'}{K_R} \quad \text{Eq. 58}$$

The deduction for Eq. 58 is valid for any rate equation of a reaction with more than 1 product and when the RDS is the desorption of R.

- Reaction with 1 product (only R):

Writing [RX] as a function of reactant and product species:

$$[RX] = K_R \cdot \sqrt[r]{K_G \cdot y_A^a \cdot y_B^b \cdot P^{\frac{a+b}{r}}} \cdot [X] \quad \text{Eq. 59}$$

Writing concentrations of intermediate species as a function of reactant and product species to Eq. 43 and Eq. 53, respectively:

$$[X] = \frac{L}{1 + K_R \cdot \sqrt[r]{K_G \cdot y_A^a \cdot y_B^b \cdot P^{\frac{a+b}{r}}} + (K_A \cdot y_A + K_B \cdot y_B) \cdot P} \quad \text{Eq. 60}$$

$$r = k_R \cdot \sqrt[r]{K_G \cdot y_A^a \cdot y_B^b \cdot P^{\frac{a+b}{r}}} \cdot [X] - k_R \cdot y_R \cdot P \cdot [X] \quad \text{Eq. 61}$$

Substituting [X] in Eq. 61 by its definition in Eq. 60 and applying the initial rates assumption:

$$r_0 = \frac{k' \cdot \sqrt[r]{K_G \cdot y_A^a \cdot y_B^b \cdot P^{\frac{a+b}{r}}}}{1 + K_R \cdot \sqrt[r]{K_G \cdot y_A^a \cdot y_B^b \cdot P^{\frac{a+b}{r}}} \cdot P^{\frac{a+b-s}{r}} + (K_A \cdot y_A + K_B \cdot y_B) \cdot P} \quad \text{Eq. 62}$$

- **d – Surface reaction controlling**

- If $a+b > r+s$:

$$\begin{aligned} r &= k_+ \cdot [AX]^a \cdot [BX]^b - k_- \cdot [RX]^r \cdot [SX]^s \cdot [X]^{a+b-r-s} = \\ &= (k_+ \cdot (K_A \cdot y_A)^a \cdot (K_B \cdot y_B)^b \cdot P^{a+b} - k_- \cdot (K_R \cdot y_R)^r \cdot (K_S \cdot y_S)^s \cdot P^{r+s}) \cdot [X]^{a+b} \end{aligned} \quad \text{Eq. 63}$$

- If $r+s > a+b$:

$$r = k_+ \cdot [AX]^a \cdot [BX]^b \cdot [X]^{r+s-a-b} - k_- \cdot [RX]^r \cdot [SX]^s =$$

$$= (k_+ \cdot (K_A \cdot y_A)^a \cdot (K_B \cdot y_B)^b \cdot P^{a+b} - k_- \cdot (K_R \cdot y_R)^r \cdot (K_S \cdot y_S)^s \cdot P^{r+s}) \cdot [X]^{r+s}$$
Eq. 64

$$\therefore r = (k_+ \cdot (K_A \cdot y_A)^a \cdot (K_B \cdot y_B)^b \cdot P^{a+b} - k_- \cdot (K_R \cdot y_R)^r \cdot (K_S \cdot y_S)^s \cdot P^{r+s}) \cdot [X]^n$$

$$n = \max(a + b, r + s)$$
Eq. 65

Writing concentrations of intermediate species as a function of reactant and product species and applying the initial rates assumption to Eq. 43 and Eq. 65, respectively:

$$[X] = \frac{L}{1 + (K_A \cdot y_A + K_B \cdot y_B) \cdot P}$$
Eq. 66

$$r_0 = k_+ \cdot (K_A \cdot y_A)^a \cdot (K_B \cdot y_B)^b \cdot P^{a+b} \cdot [X]^n$$
Eq. 67

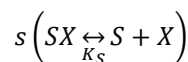
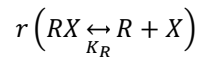
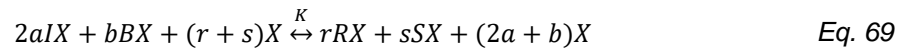
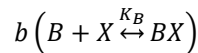
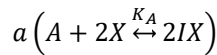
Substituting [X] in Eq. 67 by its definition in Eq. 66 and applying the initial rates assumption:

$$r_0 = \frac{\cdot (K_A \cdot y_A)^a \cdot (K_B \cdot y_B)^b \cdot P^{a+b}}{(1 + (K_A \cdot y_A + K_B \cdot y_B) \cdot P)^n}$$

$$n = \max(a + b, r + s)$$
Eq. 68

The denominator exponent n for rate equations where the RDS is the surface reactions step is always the maximum between the stoichiometric coefficients in the RDS of the reactants that adsorb and the stoichiometric coefficients in the RDS of the products that adsorb, since n is the number of active sites involved in the RDS

- **B: A adsorbs dissociatively**



$$L = [X] + [IX] + [BX] + [RX] + [SX]$$
Eq. 70

$$K_G = \frac{K \cdot K_A^a \cdot K_B^b}{K_R^r \cdot K_S^s} \quad \text{Eq. 71}$$

○ **e – Adsorption of A controlling**

$$r = k_A \cdot y_A \cdot P \cdot [X]^2 - k_{-A} \cdot [IX]^2 \quad \text{Eq. 72}$$

Writing concentrations of intermediate species as a function of reactant and product species and applying the initial rates assumption to Eq. 70 and Eq. 72, respectively:

$$[X] = \frac{L}{1 + K_B \cdot y_B \cdot P} \quad \text{Eq. 73}$$

$$r_0 = k_A \cdot y_A \cdot P \cdot [X]^2 \quad \text{Eq. 74}$$

Substituting [X] in Eq. 74 by its definition in Eq. 73:

$$r_0 = \frac{k' \cdot y_A \cdot P}{(1 + K_B \cdot y_B \cdot P)^2} \quad \text{Eq. 75}$$

○ **f – Adsorption of B controlling**

$$r = k_B \cdot y_B \cdot P \cdot [X] - k_{-B} \cdot [BX] \quad \text{Eq. 76}$$

Writing concentrations of intermediate species as a function of reactant and product species and applying the initial rates assumption to Eq. 70 and Eq. 76, respectively:

$$[X] = \frac{L}{1 + \sqrt{K_A \cdot y_A \cdot P}} \quad \text{Eq. 77}$$

$$r_0 = k_B \cdot y_B \cdot P \cdot [X] \quad \text{Eq. 78}$$

Substituting [X] in Eq. 78 by its definition in Eq. 77:

$$r_0 = \frac{k' \cdot y_B \cdot P}{1 + \sqrt{K_A \cdot y_A \cdot P}} \quad \text{Eq. 79}$$

When A dissociates and the RDS is not the adsorption of A, [IX] can be written as $\sqrt{K_A \cdot y_A \cdot P}$, since the adsorption of A requires 2 active sites. Therefore, for these cases, the adsorption term of A in the definition of [X] is $\sqrt{K_A \cdot y_A \cdot P}$

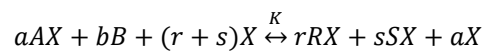
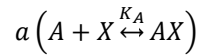
○ **g – Desorption of R controlling**

- Reaction with 2 products (S exists): Same as deduction for Eq. 58, only the adsorption term $K_A \cdot y_A \cdot P$ in the definition of [X] is replaced by $\sqrt{K_A \cdot y_A \cdot P}$.
- Reaction with 1 product (only R): Same as deduction for 'c' (Eq. 62), only the adsorption term $K_A \cdot y_A \cdot P$ in the definition of [X] is replaced by $\sqrt{K_A \cdot y_A \cdot P}$.

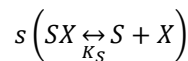
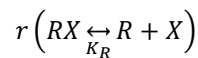
- **h – Surface reaction controlling:** Same as deduction for 'd' (Eq. 68), but the adsorption term $K_A \cdot y_A \cdot P$ in the definition of $[X]$ is replaced by $\sqrt{K_A \cdot y_A \cdot P}$ and the denominator exponent n is defined by Eq. 80.

$$n = \max(2a + b, r + s) \quad \text{Eq. 80}$$

- **C: B does not adsorb**



Eq. 81



$$L = [X] + [AX] + [RX] + [SX] \quad \text{Eq. 82}$$

$$K_G = \frac{K \cdot K_A^a}{K_R^r \cdot K_S^s} \quad \text{Eq. 83}$$

- **i – Adsorption of A controlling**

$$r = k_A \cdot y_A \cdot P \cdot [X] - k_{-A} \cdot [AX] \quad \text{Eq. 84}$$

Writing concentrations of intermediate species as a function of reactant and product species and applying the initial rates assumption to Eq. 82 and Eq. 84, respectively:

$$[X] = L \quad \text{Eq. 85}$$

$$r_0 = k_A \cdot y_A \cdot P \cdot [X] \quad \text{Eq. 86}$$

Substituting $[X]$ in Eq. 86 by its definition in Eq. 85:

$$r_0 = k' \cdot y_A \cdot P \quad \text{Eq. 87}$$

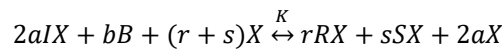
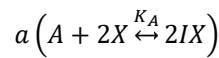
When a species does not adsorb it does not take up any active sites, thus it will not appear on the definition of $[X]$ nor on the denominator of the rate equation, consequently.

- **j – Desorption of R controlling**
 - Reaction with 2 products (S exists): Same as deduction for Eq. 58, only the adsorption term $K_B \cdot y_B \cdot P$ in the definition of $[X]$ is not present.

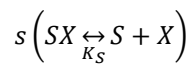
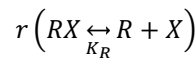
- Reaction with 1 product (only R): Same as deduction for 'c' (Eq. 62), only the adsorption term $K_B \cdot y_B \cdot P$ in the definition of $[X]$ is not present.
- **k – Surface reaction controlling:** Same as deduction for 'd' (Eq. 68), but since B does not adsorb, the adsorption constant K_B is not present in the numerator, the adsorption term $K_B \cdot y_B \cdot P$ is not present in the definition of $[X]$ and the denominator exponent n is defined by Eq. 88.

$$n = \max(a, r + s) \quad \text{Eq. 88}$$

- **D: B does not adsorb, A adsorbs dissociatively**



Eq. 89



$$L = [X] + [IX] + [RX] + [SX] \quad \text{Eq. 90}$$

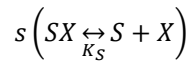
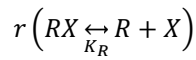
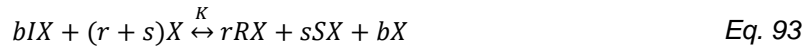
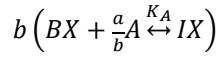
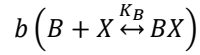
$$K_G = \frac{K \cdot K_A^a}{K_R^r \cdot K_S^s} \quad \text{Eq. 91}$$

- **l – Adsorption of A controlling:** Same as deduction for 'h' (Eq. 75), only the adsorption term $K_B \cdot y_B \cdot P$ in the definition of $[X]$ is not present.
- **m – Desorption of R controlling**
 - Reaction with 2 products (S exists): Same as deduction for Eq. 58, but in the definition of $[X]$ the adsorption term $K_A \cdot y_A \cdot P$ is replaced by $\sqrt{K_A \cdot y_A \cdot P}$ and the adsorption term $K_B \cdot y_B \cdot P$ is not present.
 - Reaction with 1 product (only R): Same as deduction for 'c' (Eq. 62), but in the definition of $[X]$ the adsorption term $K_A \cdot y_A \cdot P$ is replaced by $\sqrt{K_A \cdot y_A \cdot P}$ and the adsorption term $K_B \cdot y_B \cdot P$ is not present.
- **n – Surface reaction controlling:** Same as deduction for 'd' (Eq. 68), but the adsorption constant K_B is not present in the numerator and, in the definition of $[X]$, the adsorption term $K_A \cdot y_A \cdot P$ is

replaced by $\sqrt{K_A \cdot y_A \cdot P}$ and the adsorption term $K_B \cdot y_B \cdot P$ is not present. Also, the denominator exponent n is defined by Eq. 92.

$$n = \max(2a, r + s) \quad \text{Eq. 92}$$

- **E: A does not adsorb**



$$L = [X] + [BX] + [IX] + [RX] + [SX] \quad \text{Eq. 94}$$

$$K_G = \frac{K \cdot (K_A \cdot K_B)^b}{K_R^r \cdot K_S^s} \quad \text{Eq. 95}$$

- **o – Impact of A controlling**

$$r = k_A \cdot [BX] \cdot (y_A \cdot P)^{\frac{a}{b}} - k_{-A} \cdot [IX] \quad \text{Eq. 96}$$

Writing concentrations of intermediate species as a function of reactant and product species and applying the initial rates assumption to Eq. 94 and Eq. 96, respectively:

$$[X] = \frac{L}{1 + K_B \cdot y_B \cdot P} \quad \text{Eq. 97}$$

$$r_0 = k_A \cdot K_B \cdot y_B \cdot y_A^{\frac{a}{b}} \cdot P^{1+\frac{a}{b}} \cdot [X] \quad \text{Eq. 98}$$

Substituting $[X]$ in Eq. 98 by its definition in Eq. 97:

$$r_0 = \frac{k' \cdot K_B \cdot y_B \cdot y_A^{\frac{a}{b}} \cdot P^{1+\frac{a}{b}}}{1 + K_B \cdot y_B \cdot P} \quad \text{Eq. 99}$$

- **p – Desorption of R controlling**

$$\begin{aligned}
 [IX] &= K_A \cdot [BX] \cdot (y_A \cdot P)^{\frac{a}{b}} = \\
 &= K_A \cdot K_B \cdot y_B \cdot y_A^{\frac{a}{b}} \cdot P^{1+\frac{a}{b}}
 \end{aligned}
 \tag{Eq. 100}$$

- Reaction with 2 products (S exists): Same as deduction for Eq. 58, only the adsorption term $K_A \cdot y_A \cdot P$ in the definition of $[X]$ is replaced by the definition of $[IX]$ in Eq. 100.
 - Reaction with 1 product (only R): Same as deduction for 'c' (Eq. 62), only the adsorption term $K_A \cdot y_A \cdot P$ in the definition of $[X]$ is replaced by the definition of $[IX]$ in Eq. 100.
- **q – Adsorption of B controlling:** Same as deduction for 'b' (Eq. 52), but since A does not adsorb, the adsorption term $K_A \cdot y_A \cdot P$ is not present in the definition of $[X]$

• **F: Uncatalyzed reaction**



- **q – Uncatalyzed reaction**

$$r = k_+ \cdot y_A^a \cdot y_B^b \cdot P^{a+b} - k_- \cdot y_R^r \cdot y_S^s \cdot P^{r+s} \tag{Eq. 102}$$

Applying the initial rates assumption:

$$r_0 = k_+ \cdot y_A^a \cdot y_B^b \cdot P^{a+b} \tag{Eq. 103}$$

Appendix 3 – Datasets used in the case studies

This appendix contains the experimental datasets used as input in the case studies discussed in Chapter 5.

- Oxidation of methane case study

Table 35 - Experimental dataset of the oxidation of methane case study

P (atm)	1.817	2.975	3.925	5.22	6.24	7.13	8.41	9.7
r_0 (lb-mol·h ⁻¹ ·lb ⁻¹)	0.398	0.511	0.531	0.581	0.616	0.658	0.694	0.712

- Propylene hydrogenation case study

Table 36 - Experimental dataset of the propylene hydrogenation case study

P (bar)	r_0 (mol·g ⁻¹ ·s ⁻¹)
0.257	4.83E-04
0.508	1.11E-03
0.759	1.57E-03
1.01	2.73E-03
1.96	4.27E-03
2.94	6.39E-03
3.44	7.51E-03
9.03	7.40E-03

- Ethanol dehydrogenation case study

Table 37 - Experimental datasets of the ethanol dehydrogenation case study

P (atm)	r_0 (mol·h ⁻¹ ·g ⁻¹)					
	T=225 °C	T=235 °C	T=250 °C	T=265 °C	T=275 °C	T=285 °C
1	0.219	0.274	0.359	0.596	0.770	0.981
3	0.248	0.293	0.396	0.730	1.00	1.38
4	0.222	0.293	0.422	0.737	1.00	1.41
7	0.167	0.215	0.311	0.556	0.826	1.00
10	0.111	0.156	0.270	0.563	0.707	0.867

- Propylene disproportionation case study

Table 38 - Full experimental datasets of the propylene disproportionation case study

T = 399 °C		T = 427 °C		T = 454 °C	
P (atm)	r ₀ (g·mol ⁻¹ ·g ⁻¹)	P (atm)	r ₀ (g·mol ⁻¹ ·g ⁻¹)	P (atm)	r ₀ (g·mol ⁻¹ ·g ⁻¹)
0.950	0.416	0.949	0.755	0.943	1.32
0.960	0.666	0.949	0.612	0.960	1.24
3.04	1.38	3.08	2.91	2.86	3.67
3.07	1.59	5.08	2.90	3.00	4.37
5.08	1.37	5.09	2.72	4.95	5.02
5.11	1.66	7.11	4.75	5.01	5.46
7.08	3.08	7.13	4.25	7.06	8.31
7.09	2.81	9.14	4.73	7.06	7.58
9.19	3.23	9.15	5.05	9.12	7.54
9.20	3.30			9.14	8.29

Table 39 - Averaged experimental datasets of the propylene disproportionation case study

T = 399 °C		T = 427 °C		T = 454 °C	
P (atm)	r ₀ (g·mol ⁻¹ ·g ⁻¹)	P (atm)	r ₀ (g·mol ⁻¹ ·g ⁻¹)	P (atm)	r ₀ (g·mol ⁻¹ ·g ⁻¹)
0.955	0.541	0.949	0.683	0.951	1.28
3.06	1.49	3.08	2.91	2.93	4.02
5.10	1.51	5.08	2.81	4.98	5.24
7.08	2.94	7.12	4.50	7.06	7.95
9.19	3.27	9.15	4.89	9.13	7.92

- Propyne hydrogenation case study

Table 40 - Experimental dataset of the propyne hydrogenation case study

P (bar)	r ₀ (mol·g ⁻¹ ·s ⁻¹)	P (bar)	r ₀ (mol·g ⁻¹ ·s ⁻¹)	P (bar)	r ₀ (mol·g ⁻¹ ·s ⁻¹)
0.310	4.142E-05	2.78	3.122E-04	6.11	8.060E-04
0.515	7.009E-05	3.41	3.727E-04	6.67	8.920E-04
1.03	1.290E-04	3.50	3.919E-04	6.79	9.335E-04
1.45	1.816E-04	3.74	4.396E-04	7.38	9.733E-04
2.29	2.867E-04	4.60	4.858E-04	8.08	1.015E-03
2.45	2.867E-04	5.60	6.228E-04	9.04	1.128E-03
2.52	2.899E-04				

Appendix 4 – User manual

The goal of this tool is for it to analyse the trends of a set of initial rate vs. pressure data and then propose the initial rate equations present in its library that can describe this data. To determine which initial rate equations can describe the experimental data, the tool generates several initial rate theoretical curves from each equation on its library, and then discards the ones that cannot generate a curve that follows the trends of the experimental data. This user manual contains the instructions the user needs to be able to utilize the tool, as well as some information regarding the different steps taken by the tool to reach the final output.

User input

When the tool starts running, the first prompt to appear in the console asks the user to write the name of the Excel file where the experimental dataset is stored. This file should already be in the 'Input' folder present in the tool's files. The file must follow the template present in Figure 43.

P	r0
1,817	0,398
2,975	0,511
3,925	0,531
5,22	0,581
6,24	0,616
7,13	0,658
8,41	0,694
9,7	0,712

Figure 43 - Template of input file (with example of data)

The headers of the column must have exactly the same names (P and r0), and the columns must be of the same length.

The following prompt asks for the units of the pressure and initial rate values. The purpose of this is only so that the graphs generated have the correct units in them.

After this, the next prompt asks for the number of reactants. Afterwards, it asks for the name, the stoichiometric coefficient, and the molar fraction in the feed of each of the reactants. Then, it asks for the number of products formed, and for the name and stoichiometric coefficient of each one. It is important to note that, since the tool was built to support a maximum of 2 reactants and 2 products, the tool will halt if a number larger than 2 is typed by the user when asking for the number of reactants and the number of products. After this, the tool asks for the molar fraction of the inert species that may be present in the feed, checking if the sum of this with the molar fractions of the reactants is equal to 1.

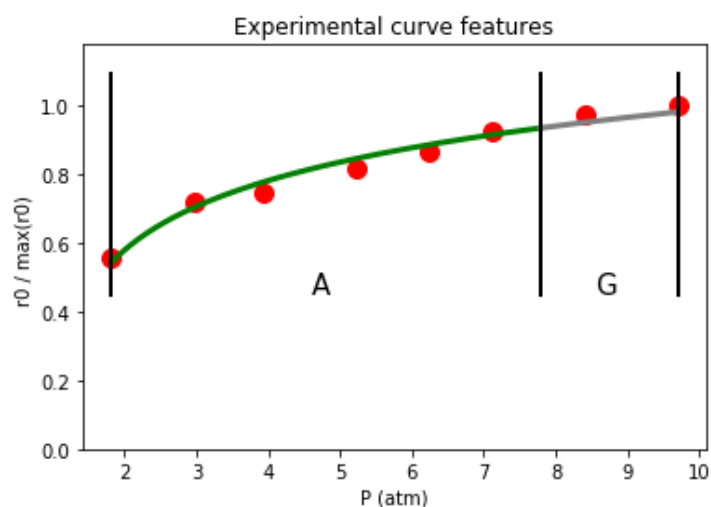
The following prompt asks for the user to write the desired name for the folder in which the output will be stored. Then, the tool presents the mechanisms in the library and asks the user for the mechanisms which are known *a priori* to not be possibilities. The mechanisms selected here will not be considered

when generating the theoretical curves. The purpose of this is to reduce the runtime of the tool, if that is possible.

If the reaction involves 2 reactants and 1 product, then the initial rate equations for all the mechanisms where the RDS is the desorption of the product will generate a high number of theoretical curves, since these initial rate equations contain the constants K_A , K_B , for K_R and K , for which the possible K_{Eq} values will be permuted. This will drastically increase the runtime of the tool if the full lists of possible K_{Eq} values are used. Therefore, for this case, the tool asks if the user wants to use a reduced list of possible values, which has less values than the full list, instead of the full list when generating the curves of these mechanisms. Using the reduced list will result in less generated curves for these mechanisms, and therefore will reduce the runtime.

Experimental branch

After all the initial user input is given the experimental data is analysed qualitatively by the tool. This qualitative analysis is done by determining the trends of the data by extracting its curve features. Therefore, the data is normalized and is analysed by the feature extraction algorithm. The output of this algorithm is then presented. It consists of the normalized experimental data with the graphical representation of its features overlapping. An example is present in Figure 44.



Experimental primitives: ['A', 'G']

Experimental extremes (atm): [1.817, 7.78896969697, 9.7]

Figure 44 - Example of output of feature extraction algorithm

The segmented curve present is drawn by the algorithm and follows the trends of the data. Each segment is called a primitive and is differentiated from other segments by the signs of their first and second derivatives. The borders of the primitives are called extremes, and the algorithm extracts the x-values of these extremes. The comparison between the experimental data and the theoretical curves is based on the primitives and the x-values (pressure values, in this case) of the extremes.

Theoretical branch

After extracting the features of the experimental data, the tool then proceeds to build the theoretical curves from each generic initial rate equation. For that, the reaction stoichiometry and the fractions of reactants in the feed inputted by the user are used.

All the initial rate equations that can be derived from the mechanisms in the library fit the type of Eq. 104.

$$r_0 = \frac{c_1 \cdot P^{e_1}}{(1 + c_2 \cdot \sqrt{P} + c_3 \cdot P + c_4 \cdot P^{e_2} + c_5 \cdot P^{e_3})^n} \quad \text{Eq. 104}$$

To generate the adequate initial rate equation for each mechanism, the parameters c_1 , c_2 , c_3 , c_4 , c_5 , e_1 , e_2 , e_3 and n are calculated using the reactant fractions, the stoichiometric coefficients and the equilibrium constants (k' would normally also be used in parameter c_1 , but this is not necessary due to the normalization of the curves before extracting their features, which cancels the influence of this constant). The way the parameters are calculated from these values depends on the descriptive tags associated to each mechanism.

Sometimes, it is possible that n takes a value larger than 3. Since the number of this parameter is the same as the number of active sites involved in the RDS, having it be larger than 3 is chemically unrealistic, since that would mean that 4 or more adjacent active sites would need to be available, which is very unlikely. Therefore, to prevent the proposal of such equations by the tool, any mechanism which produces an equation with a n value larger than 3 is automatically eliminated from the list of possibilities.

The equilibrium constants (K_{Eq}) in the initial rate equations can have any value. Therefore, the chosen approach was to insert lists of possible K_{Eq} values in the code, and make each dataset generated from an initial rate equation have a different combination these values.

If the reaction involves 2 reactants, then the algorithm that generates the rate equations is ran two times: in the first run, reactant A is the first reactant for which the user inputs its information, and reactant B is the second one inputted by the user. On the second run, the reactants are swapped. This is necessary because most of the mechanism (specifically mechanisms 'e' to 'q'), and consequently most of the generic initial rate equations in the library are "non-symmetrical" regarding the reactants, i.e., the reactants adsorb differently in most of the mechanism. This means that when reactants A and B are swapped the values of the parameters of Eq. 104 change for the initial rate equations of these mechanisms. Therefore, to ensure that these possibilities are considered by the tool, the second run with the reactants swapped is performed whenever the reaction has two reactants. When this happens, the rate equations generated in the first run have '_1' added to their associated mechanism ID, the rate equations generated in the second run have '_2' added instead (for example, the rate equation generated from mechanism 'e' in the first run becomes associated with the ID 'e_1', while the rate equation generated from the same mechanism but in the second run becomes associated with the ID 'e_2'; these IDs will be present in the tool). When the reaction only involves one reactant '_1' still is added to the IDs of the mechanisms.

With the equation parameters calculated, a theoretical curve is created for each equation and set of parameters. The pressure values used to generate these curves are the same as the ones in the experimental data. This is because the distribution of the points of a dataset has influence on the features extracted by the feature extraction algorithm. Therefore, the best way to nullify the influence of this factor on the comparison of experimental and theoretical features is to create the theoretical curves for the same pressure values present in the experimental dataset.

After generating the theoretical curves, their features are extracted, normalizing the initial rate values prior to the extraction. This way, the theoretical curves are at the same scale as the experimental curve when their features are extracted, which nullifies the influence of scale on the comparison of experimental and theoretical features.

Comparison of experimental and theoretical features

With the theoretical curves generated and their features extracted, it is now possible to compare these features to the ones extracted from the experimental data. All the theoretical curves generated are compared to the experimental data. The goal of this test is to discard all the curves that do not follow the trends of the experimental data.

The first test is the comparison of the primitives. If the set of primitives of the theoretical curve are different from the one of the experimental data, this curve fails the test and is discarded. If both sets of primitives are the same the curve passes onto the second test.

The second test is the comparison of the pressure values of the extremes. This is done by calculating the relative error of each extreme of the theoretical curve in comparison with the corresponding value of the experimental data, according to Eq. 105.

$$\delta = \frac{|P_{Theor} - P_{Exp}|}{P_{Exp}} \quad \text{Eq. 105}$$

In this equation, P_{Theor} is the pressure value of the theoretical extreme and P_{Exp} is the pressure value of the equivalent extreme from the experimental data.

For each extreme, the calculated relative error is compared to the tolerance value. This tolerance value is chosen by the user and is inputted when the prompt that asks for it appears. This value represents the maximum value that the relative errors of the extremes are allowed to have to pass the test. The value chosen for the tolerance should be between 0 and 1 (although values larger than 1 are still accepted).

For a theoretical curve to pass the test, the relative errors of all its extremes must not exceed the tolerance value. The first and last extremes are not compared since they are always the first and last pressure value of the analysed dataset. Since the theoretical curves are generated for the same pressure values of the experimental dataset, these two extremes will always have the same pressure values as the equivalent extremes of the experimental dataset, so there is no need to compare them.

If the curve passes this test, then it is considered to follow the trends of the experimental data to an acceptable extent. Therefore, its ranking criteria must be calculated. They are the sum of relative errors (SRE) and the mean squared error of the normalized values (MSE_{norm}). The SRE is simply the sum of all the relative errors of the extremes previously calculated, while the MSE_{norm} is calculated following Eq. 106.

$$MSE_{norm} = \frac{\sum_{i=1}^n (\hat{r}_{0,Theor}(i) - \hat{r}_{0,Exp}(i))^2}{n} \quad Eq. 106$$

In this equation, $\hat{r}_{0,Theor}(i)$ and $\hat{r}_{0,Exp}(i)$ are the normalized initial rate values of the theoretical curve and the experimental data, respectively, and n is the total number of data points. The principal ranking criterion is the SRE, since it is the one that best reflects how well the theoretical follows the trends of the experimental data. The MSE_{norm} serves more as a secondary criterion to distinguish between curves that happen to have the same SRE.

Any initial rate equation that generates at least one theoretical curve that passes the comparison tests is considered as a possible model for the experimental data and is proposed by the tool as such. For each of these equations the curve with the smallest ranking criteria is chosen, and the equilibrium constants of this curve are stored alongside with the values of its ranking criteria in a way that makes them retraceable to the initial rate equation that generated it. These curves are then ordered accordingly to their ranking criteria values, from smallest to largest.

Estimating k'

The desired output of the tool is not only the proposed initial rate equations, but also some initial guesses for the parameters of the curves that could be used for an eventual regression of the curves. These guesses already exist for the equilibrium constants since the theoretical curves were generated by permutating different values for them. However, no initial guess for k' has been made at this point. Therefore, the next step the tool makes is to estimate an initial guess for this parameter, for each of the best curves of the proposed initial rate equations.

For each of these curves, the necessary k' values for the curve to intersect each of the points of the experimental data separately were determined. Then, with each of those values, a curve was generated and its mean squared error (MSE) was calculated using Eq. 107. This was also done for the average of all the determined k' values.

$$MSE = \frac{\sum_{i=1}^n (r_{0,Theor}(i) - r_{0,Exp}(i))^2}{n} \quad Eq. 107$$

In this equation, $r_{0,Theor}(i)$ and $r_{0,Exp}(i)$ are the initial rate values of the theoretical curve and the experimental data, respectively, and n is the total number of data points. This results in $n+1$ possible k' values. Of all these values, the one with the smallest associated MSE is the one chosen as the initial guess. The smallest MSE value is the one present in the tool's output.

Output of the tool

The output of the tool is where the proposed initial rate equations are presented to the user, as well as the initial guesses for the constants of each equation. The order of appearance of the proposed initial rate equations in the output is from the one with the smallest ranking criteria to the one with the highest.

For each initial rate equation, the tool outputs a graph where the theoretical curve of that equation that best follows the experimental data trends is overlapping the experimental data points. For clarity, instead of only presenting theoretical points at the same pressure values as the experimental points, the graph presents an uninterrupted curve covering the whole pressure range. Following each of these graphs is the initial rate equation that generated this curve as well as the initial guesses for the parameters and the ranking criteria (SRE and MSE), all in text format. An example of both these outputs is present in Figure 45.

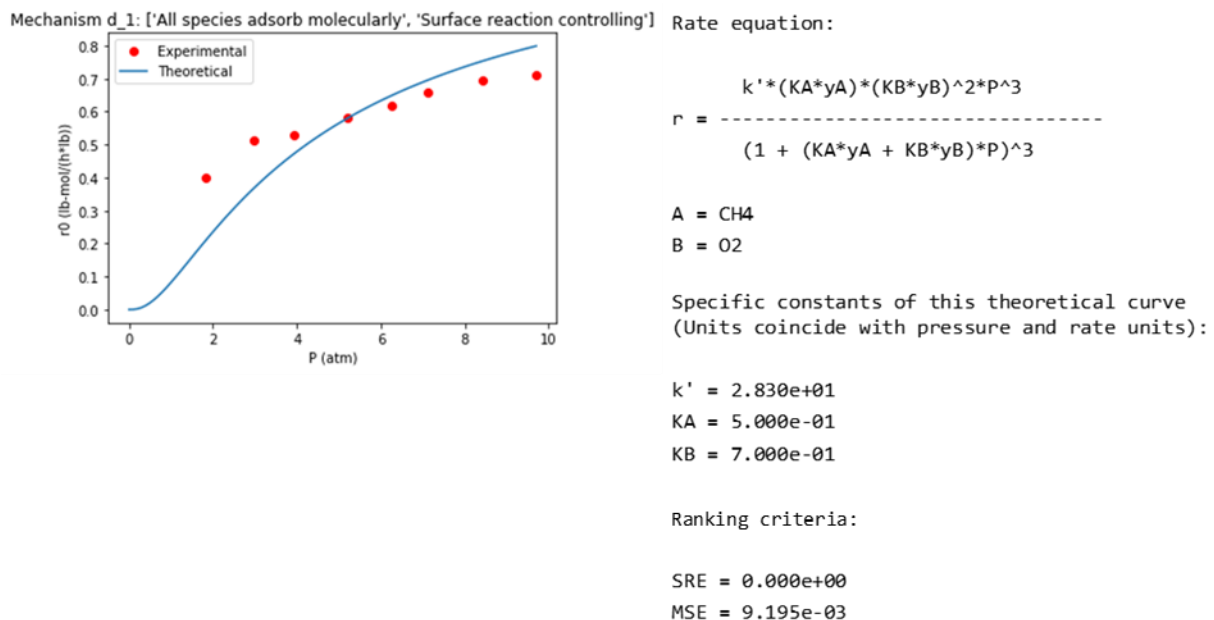


Figure 45 - Example of output graph with best theoretical curve (left) and output text string (right)

After all the graphs and text strings have appeared, a table containing all the information of the best curves of each proposed rate equation (mechanism ID, equilibrium constants, k' value and ranking criteria) appears in the console, as a summary of all the output. An example is present in Figure 46.

ID	k	KA	KB	KR	KS	Kr	MSE	SRE
d_1	28.30114	0.5	0.7	N/A	N/A	N/A	0.009195	0
k_2	12.63757	0.7	N/A	N/A	N/A	N/A	0.010232	0.010223
k_1	0.609413	10	N/A	N/A	N/A	N/A	0.010387	0.010223
a_1	4.372802	N/A	0.3	N/A	N/A	N/A	0.001046	0.040892
e_2	0.213526	N/A	1	N/A	N/A	N/A	0.002771	0.071561
b_1	0.540103	10	N/A	N/A	N/A	N/A	0.000642	0.153344
e_1	4.051128	N/A	0.1	N/A	N/A	N/A	0.00092	0.245351

Mechanisms with _1 : A = CH4, B = O2

Mechanisms with _2 : A = O2, B = CH4

Figure 46 - Example of summary of output

Re-run and saving output

After the output is presented, the tool offers the user the possibility of running the comparison algorithm once again, if the user wishes to try a different tolerance value. If the user chooses to do so, the tool restarts from the point where it asks the user for the tolerance value to use and runs until the new output is presented. The user can choose to do this as many times as desired.

If the user does not wish to test another tolerance value, the tool proceeds to save the output. It is important to note that the tool will only save the output resultant of the last tolerance value tested. The output is then saved in the folder with the name designated by the user at the start of the run, which will itself be located inside the 'Output' folder in the tool's files. The output consists of several .png files: the graph of the experimental features (see Figure 44) and the graphs of the best theoretical curves of the proposed initial rate equations (see Figure 45). In this folder will also be an Excel file with all of the text output (see Figure 45), as well as the table with the summary of the output (see Figure 46).