

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

VASCO SALTÃO^{1,2,*}, FILIPE G. FREIRE¹, LAURA PIRRO², JORIS W. THYBAUT², PEDRO S. F. MENDES²

¹Department of Chemical Engineering, Instituto Superior Técnico, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal

²Laboratory for Chemical Technology, Department of Materials, Textiles and Chemical Engineering, Ghent University, Technologiepark 125, 9052 Ghent, Belgium

*Corresponding author: vasco.saltao@tecnico.ulisboa.pt

December 9, 2020

To design catalysts and optimize reactive systems, a kinetic model of the catalytic reaction is needed. However, the traditional manual construction of kinetic models is of trial-and-error nature and demands from the researcher a high degree of expertise in kinetics and catalysis, thus creating a bottleneck. This can be mitigated by automating the construction of kinetic models. Therefore, the goal of this work was to develop a software tool that automatically proposes rate equations for a catalytic reaction based on kinetic experimental data. This proposal is based on the similarity between the features of the experimental data and those predicted by theoretical initial rate equations. These equations were deduced with some assumptions from typical mechanisms in heterogeneous catalysis, and are present in a library, from which the tool generates theoretical curves of initial rate as a function of pressure curves. Afterwards, the tool compares the features of these curves to the features of the data, thus screening the rate equations to eliminate those that cannot be a model of the data. Afterwards, the possible models are ranked based on feature similarity. Selected literature datasets were processed to test the tool, with it successfully being able to deduce the same rate equations that the researchers proposed in most cases. Therefore, the tool fulfils its goal, although its applicability is limited by analysing only initial rates and by the assumptions used to develop the library. Even with its limitations, this is an important first step in the automation of kinetic modelling applied to heterogeneous catalysis that, once fully generalized, will enable a larger number of researchers to achieve kinetic models for catalytic reactions in less time.

1. Introduction

Catalysts are essential to the sustainability of several industries since they offer a more energetically favourable pathway for chemical reactions to occur[1]. Therefore, the research for new catalysts is of major importance.

In heterogeneous catalysis, the physical state of the catalyst, i.e. solid, is different from the one of the reactants and products, i.e. gas or liquid. Here, the transformation of reactants into products takes place on the surface of the catalyst. For that to happen, the reactant molecules must adsorb onto the active sites present on surface and convert into products, which must then desorb. Each of these elementary steps is influenced by different factors[2], thus potentially occurring at different rates, determining the overall reaction rate. The net production rate of the species involved in the reaction can be written as a function of the rates of the elementary steps.

To optimize a reactive process, it is necessary to control the reaction rate. That, in turn, requires knowledge on how the rate varies with reaction conditions. For that it is necessary to have a kinetic model, which is a mathematical model that describes the influence of reaction conditions and of the catalyst on the reaction rate[3].

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[4]. This analysis is performed by an experienced researcher that observes the trends in the experimental data and based on kinetic and catalytic knowledge proposes the models that are possible explanations for the behaviour of the experimental data. Then, it is necessary to solve the mass balances of the species involved in the reaction. There are different ways to do so, depending on which assumptions are made[3]. A common one is the rate determining step (RDS) approximation, which assumes that one of the elementary steps in the mechanism determines the rate of the overall reaction, while the other steps are in quasi-equilibrium[3]. This approximation ensures that an analytical solution is achieved. However, this solution cannot account for a

shift in the RDS due to changing reaction conditions and neither for time dependent phenomena.

This manual modelling process requires expertise on kinetics and catalysis from the researcher, which limits the number of researchers that can work on this field. Also, this is a trial-and-error process, since the researcher will derive a new rate equation until reaching one that can explain the data in a satisfactory way. This makes this process very time consuming. These factors create therefore a bottleneck in the kinetic modelling process. By automating it utilising a software tool, the bottleneck mentioned above would be 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 would be reduced, which means that more researchers would be able to work on kinetic modelling.

Concerning kinetic modelling automation, most work found based their automation on computer simulations, utilizing different resources such as linear algebraic methods[5], graph theory[5]–[7], lumping strategies[3] and machine learning[8]. These methods have some advantages, such as reducing the need for performing catalytic experiments to obtain new experimental data, and in some cases eliminating the need of experimental data. However, they also have their disadvantages, with the main one being that the theory on which the simulations are based is built on some assumptions, which can lead to incorrect models that are unrealistic. On the other hand, the automatic proposal of models based on experimental data results in proposing models that are more realistic, since they are based on real data, even if some assumptions are used. The disadvantage is that the performance of such a tool will strongly depend on the quality of the data itself (number of points, distribution of points, experimental noise, etc.).

Focusing on automatic kinetic modelling tools based on experimental data, the tool TAM-C, developed by Schaich, King and Becker[9], [10], was found. This tool generates a curve that fits the experimental data, recognizes the “episodes” (features) of said curve and then eliminates the models that are qualitatively dissimilar to the

data, reducing the number of possible models to be fitted 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 number of points[11].

In summary, no tool was found in the literature designed specifically for kinetic data of catalytic reactions.

Therefore, the goal of this work is to develop a tool that automatically proposes rate equations for catalytic reactions based on experimental kinetic data and on the screening of rate equations. This screening will be done by comparing the features of the experimental data with those theoretically predicted by different rate equations, since different rate equations result in rate curves with different features. This approach requires a way of automatically recognizing the features of datasets, as well as a library of typical rate equations for catalytic reactions from which to generate the theoretical curves. The algorithm developed by Siradze[12] can generate a curve which follows the data trends and is chemically realistic and then extract its features, much like an experienced researcher would do, so it fulfils the former requirement. For the library of rate equations to be developed, the equations themselves must be deduced from typical catalytic mechanisms. Some of these mechanisms can be found in the work developed by Yang and Hougen[13]. The rate equations will be deduced assuming initial rates of reaction (i.e., no products present), since rate equations deduced this way generate curves which features are more easily recognizable.

2. Concept of rate equation proposal tool

To develop a tool capable of automatically proposing rate equations that can describe an experimental kinetic dataset, a feature recognition approach was followed. The concept behind this is that different rate equations produce curves with different features 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. the conceptual flowchart of this tool is presented in Figure 1.

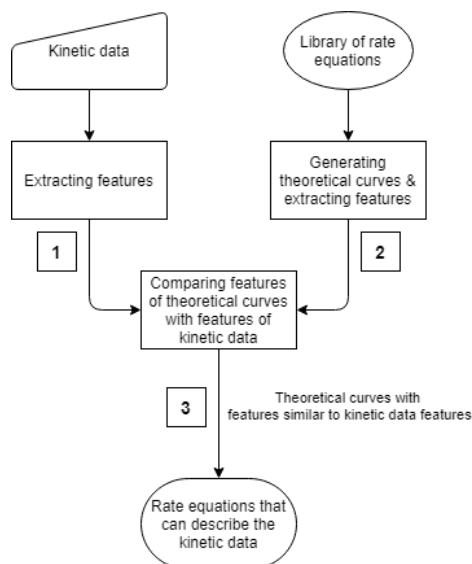


Figure 1 - Conceptual flowchart of the tool (trapezoid – user input; rectangle – process; oval – information included in tool; rounded rectangle – output of tool)

The process of the tool can be divided into three parts. In the experimental branch (1), the features of the experimental kinetic data are extracted. In the theoretical branch (2), theoretical curves are generated from the rate equations in the library and their features are extracted. The third and most downstream part of the tool 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. Of these three parts, most of the development was directed to the theoretical branch and the comparison section, since the feature extraction algorithm was already developed by S. Siradze[12] and this is the only element necessary for this branch, apart from some data treatment prior to the feature extraction.

The concept of the tool relies on curve features. Therefore, the tool must contain an algorithm that can extract the features of the curves in a way that they can be compared. The adopted algorithm was the one developed by Siradze[12]. This algorithm automatically performs the qualitative analysis method proposed by Janusz and Venkatasubramanian[14]. First it draws a curve that follows the data trends. Then it splits that curve into several primitives, which are sections of the curve differentiated by the signs of the first and second derivatives. 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[12] (signs of first and second derivatives in brackets, respectively)

The relevant features extracted by this algorithm are the primitives themselves and the x-values at the extremes of the primitives, i.e., the borders of the primitives. These features are graphically represented in Figure 3. After extracting the features, the algorithm checks whether they represent a chemically realistic curve. If not, a new curve is redrawn. This way, the actions that would be performed by an experienced researcher in the field are mimicked.

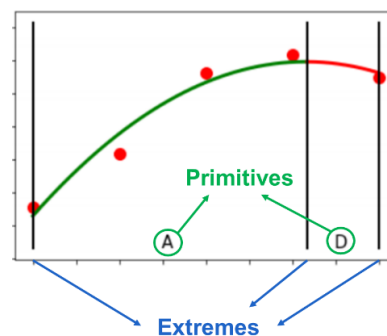


Figure 3 - Graphical representation of curve features

3. Theoretical branch of the tool

3.1. Construction of the library

As mentioned before, the theoretical branch of the tool is responsible for generating the theoretical curves predicted from the rate equations in the library. Therefore, the first step in the development of this branch is the development of a library of rate equations. The library must be applicable to the majority of catalytic reactions and must be complete, i.e., the rate equations in it must consider all possibilities within the assumptions used to deduce them. Another characteristic is that the rate equations should be deduced assuming initial rates of reaction (no products present in the system). This is because the curves resulting from such equations are more easily recognizable.

To ensure that both requirements of the library are met, it is necessary to deduce the rate equations from a complete set of common and reaction mechanisms in heterogeneous catalysis. A good source of these mechanisms is the work of Yang and Hougen[13] where they studied the effects of total pressure on the initial rates of catalysed gaseous reactions with different rate equations. These rate equations and corresponding mechanisms account for all possibilities regarding the adsorption of reactants and RDS. Therefore, this work is a good source for developing the library of the tool. However, the rate equations present in this work

cannot be directly implemented in the tool since some of the assumptions made to deduce them greatly reduce its applicability, such as the assumption that only one molecule of each species is involved in the reaction and that the reactor feed is composed of only the reactants in stoichiometric proportions.

To overcome the limitations of the work of Yang and Hougen, the generic rate equations were deduced from the mechanisms studied in that work for generic stoichiometric coefficients and generic molar fractions of reactants, allowing for the presence of one reactant in excess, as well as the presence of inert species. They were deduced for initial rates through the RDS assumption since it is valid for most catalytic reactions[2]. These generic rate equations are present in Table 1, derived for two reactants (the number of products is irrelevant except when explicitly stated). To arrive to the equations for only 1 reactant (reactant A) all that is necessary is for the parameters relative to reactant B (y_B and b) to be 0 and to take into account that the equations that assume that the adsorption of B is the RDS or that A or B do not adsorb are irrelevant, since it would either be paradoxical or redundant, respectively. 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.

While the deduced generic equations overcome some of the previously mentioned limitations, their applicability is still limited. Dealing only with initial rates, while making the resulting theoretical curves more easily recognizable, reduces the number of datasets that can be analysed by the tool. The generic rate equations also assume that one RDS and only one surface reaction step take place. However, these limitations are still valid for a wide variety of catalytic reactions. Other assumptions are that all products adsorb onto the catalyst, that all species adsorb competitively onto the same type of active site and that the maximum of two reactants and two products allowed.

3.2. Generating theoretical curves

From the generic initial rate equations, the library of the theoretical curves can be generated. For each curve several different values are required, as is schematically represented in Figure 4. These values are 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.

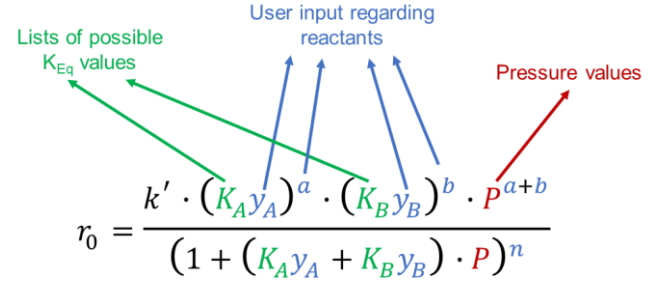


Figure 4 - Origin of values used to generate theoretical curves

All the generic initial rate equations of Table 1 fit the type of Eq. 1.

$$r_0 = \frac{c_1 \cdot P^{e_1}}{(1 + c_2 \cdot \sqrt{P} + c_3 \cdot P + c_4 \cdot P^2 + c_5 \cdot P^3)^n} \quad \text{Eq. 1}$$

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. The lumped kinetic constant k' only scales the curve, having no influence on its features. Therefore, its value is not necessary to generate the theoretical curves. The way these parameters are calculated is different for each generic rate equation. The formulas used to calculate them depend on the descriptions of adsorption and RDS present in Table 1 in bold, which are associated to each rate equation in the code implementation of the library. When calculating them,

Table 1 - Library of generic initial rate equations

Adsorption	RDS			
	Adsorption (impact for 'A does not adsorb') of A controlling	Adsorption of B controlling	Desorption of R controlling (for 1 product) <u>2 products: $r_0 = k'/K_R$</u>	Surface reaction controlling
All species adsorb molecularly	$r_0 = \frac{k' \cdot y_A \cdot P}{1 + K_B y_B \cdot P}$	$r_0 = \frac{k' \cdot y_B \cdot P}{1 + K_A y_A \cdot P}$	$r_0 = \frac{k' \cdot r \sqrt{K_G \cdot y_A^a \cdot y_B^b \cdot P^{\frac{a+b}{r}}}}{1 + K_R \cdot r \sqrt{K_G \cdot y_A^a \cdot y_B^b \cdot P^{\frac{a+b}{r}}} + (K_A y_A + K_B y_B) \cdot P}$	$r_0 = \frac{k' \cdot (K_A y_A)^a \cdot (K_B y_B)^b \cdot P^{a+b}}{(1 + (K_A y_A + K_B y_B) \cdot P)^n}$ $n = \max(a + b, r + s)$
A adsorbs dissociatively	$r_0 = \frac{k' \cdot y_A \cdot P}{(1 + K_B y_B \cdot P)^2}$	$r_0 = \frac{k' \cdot y_B \cdot P}{1 + \sqrt{K_A y_A \cdot P}}$	$r_0 = \frac{k' \cdot r \sqrt{K_G \cdot y_A^a \cdot y_B^b \cdot P^{\frac{a+b}{r}}}}{1 + K_R \cdot r \sqrt{K_G \cdot y_A^a \cdot y_B^b \cdot P^{\frac{a+b}{r}}} + \sqrt{K_A y_A \cdot P} + K_B y_B \cdot P}$	$r_0 = \frac{k' \cdot (K_A y_A)^a \cdot (K_B y_B)^b \cdot P^{a+b}}{(1 + \sqrt{K_A y_A \cdot P} + K_B y_B \cdot P)^n}$ $n = \max(2a + b, r + s)$
B does not adsorb	$r_0 = k' \cdot y_A \cdot P$	-	$r_0 = \frac{k' \cdot r \sqrt{K_G \cdot y_A^a \cdot y_B^b \cdot P^{\frac{a+b}{r}}}}{1 + K_R \cdot r \sqrt{K_G \cdot y_A^a \cdot y_B^b \cdot P^{\frac{a+b}{r}}} + K_A y_A \cdot P}$	$r_0 = \frac{k' \cdot (K_A y_A)^a \cdot y_B^b \cdot P^{a+b}}{(1 + K_A y_A \cdot P)^n}$ $n = \max(a, r + s)$
A adsorbs dissociatively, B does not adsorb	$r_0 = k' \cdot y_A \cdot P$	-	$r_0 = \frac{k' \cdot r \sqrt{K_G \cdot y_A^a \cdot y_B^b \cdot P^{\frac{a+b}{r}}}}{1 + K_R \cdot r \sqrt{K_G \cdot y_A^a \cdot y_B^b \cdot P^{\frac{a+b}{r}}} + \sqrt{K_A y_A \cdot P}}$	$r_0 = \frac{k' \cdot (K_A y_A)^a \cdot y_B^b \cdot P^{a+b}}{(1 + \sqrt{K_A y_A \cdot P})^n}$ $n = \max(2a, r + s)$
A does not adsorb	$r_0 = \frac{k' \cdot K_B y_B \cdot y_A^a \cdot P^{1+\frac{a}{b}}}{1 + K_B y_B \cdot P}$	$r_0 = k' \cdot y_B \cdot P$	$r_0 = \frac{k' \cdot r \sqrt{K_G \cdot y_A^a \cdot y_B^b \cdot P^{\frac{a+b}{r}}}}{1 + K_A K_B y_B \cdot y_A^a \cdot P^{1+\frac{a}{b}} + K_R \cdot r \sqrt{K_G \cdot y_A^a \cdot y_B^b \cdot P^{\frac{a+b}{r}}} + K_B y_B \cdot P}$	-
Uncatalyzed reaction	$r_0 = k' \cdot y_A^a \cdot y_B^b \cdot P^{a+b}$			

any equation with a n value larger than 3 is automatically eliminated. This is because the value of n is the number of active sites in the RDS, and having it larger than 3 is unrealistic[15]. The pressure values used to generate the curves are the same as the ones present in the experimental data. The reasons for this are related to the comparison between experimental and theoretical features and will be discussed further ahead.

With the molar fractions of reactants and the stoichiometry coming from user input and the pressure values coming from experimental data, the only unknown parameters are the equilibrium constants K_{Eq} . These constants can have any value inside a chemically realistic range. So, the approach chosen was to generate several different curves for each equation by permutating between some possible values for these constants. These values were chosen so they cover an adequate range, determined by analysing some examples[16], and generate as many curve features as can be generated from each generic rate equation. A distinction was made between the adsorption constants K_i and the surface reaction equilibrium constant K_r . While the adsorption constants were considered to only have non-negligible values to represent a meaningful adsorption, K was considered able to have any value, even ones too small to be considered for adsorption constants. Therefore, two lists of K_{Eq} values were put in use, one for each of the previously mentioned cases. In addition to this, a reduced version of each of these two lists was also developed, to be used whenever the number of theoretical curves to be generated is too large (at the request of the user), in an effort to limit the runtime of the tool. The four lists of K_{Eq} values are present in Table 2.

Table 2 - Lists of possible values for the equilibrium constants

List				
	K_i (pressure units ⁻¹)	K_i (pressure units ⁻¹)	K_r	K_r
	<u>Full</u>	<u>Reduced</u>	<u>Full</u>	<u>Reduced</u>
	Values	0.1	0.1	1E-7
0.3		0.5	0.01	0.01
0.5		1	0.1	0.1
0.7		10	0.3	0.5
1		100	0.5	1
10		-	0.7	10
100		-	1	100
-		-	10	-
-		-	100	-

It was decided that the full lists of K_{Eq} values must have more values in the range of 0.1 to 1 than from 1 to 100, as it can be seen in Table 2. This is because the features of the curve are affected by the denominator parameters of the rate equation (c_2 , c_3 , c_4 and c_5 in Eq. 1), calculated using the equilibrium constants, and the features change more drastically for a variation of small parameter values than for the same variation on larger parameter values. Even if it is theoretically possible, due to the limited K_{Eq} values, for some possible features of a rate equation to not be represented by any generated theoretical curve, this approach of using more small values than larger ones turns that possibility into an edge case. All the algorithms of the theoretical branch described in this chapter are summarized in Figure 5.

4. Feature comparison and proposal algorithm

The theoretical branch of the tool generates theoretical curves from each generic rate equation of the library. These curves can then be compared with the experimental data to screen the rate equations and eliminate those that cannot be a possible model that describes the data. More precisely, the features of the theoretical curves are compared with the features of the experimental data. Therefore, the sets of features must be extracted in a way that guarantees that the results of the comparison are trustworthy.

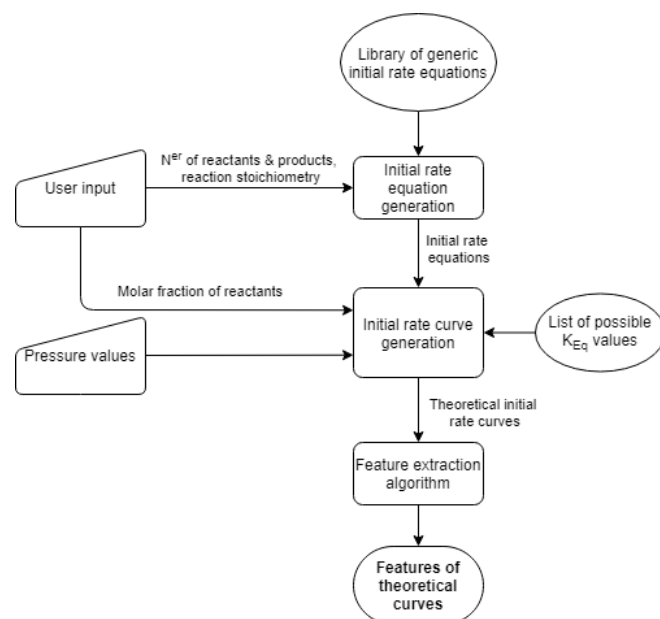


Figure 5 - Flowchart of theoretical branch of the tool (trapezoid – user input; rectangle – process; oval – information included in tool; rounded rectangle – output of theoretical branch)

4.1. Validity of feature comparison

As the feature extraction algorithm was mainly developed for other type of data[12], preliminary testing showed that some implementations on both branches were needed to ensure the comparison between the features in experimental data and in theoretical curves is valid. The first issue is related to the influence of the x -values of the dataset on the recognized features. An example of this is given in Figure 6, where two datasets and their recognized features are represented. Both datasets follow Eq. 2, but the x -values of their points are different (all between 1 and 10, however).

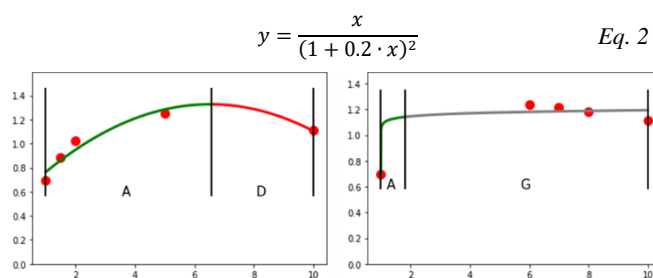


Figure 6 - Comparison of features extracted from two datasets that follow Eq. 2 for different x -values

As can be seen in Eq. 2, the recognized features of the datasets are different. This is a result of the different distribution of points along the x -axis. Therefore, to ensure that the experimental data and theoretical curves with the similar trends have similar recognized features, the pressure values used to generate the latter are the same present in the experimental data, as mentioned before.

Another issue is that, on some occasions, the scaling of the dataset can affect its recognized features. This is due to the numerical methods on which this algorithm is based. An example of this is present in Figure 7, where two datasets and their respective features are represented. Both datasets follow Eq. 3 for the same x -values, varying only k , which does not affect the shape of the curve, scaling it only.

$$y = \frac{k \cdot x}{(1 + 2.75 \cdot x)^2} \quad \text{Eq. 3}$$

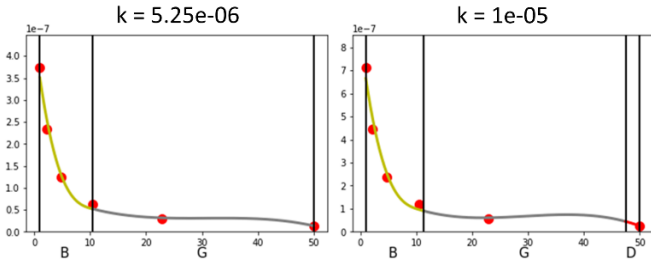


Figure 7 - Extracted features of the same curve at two different scales

As it can be seen in Figure 7, the recognized features are not the same, even though the datasets have the same trends, only different scales. To prevent this from happening between the experimental data and a theoretical curve, it was decided to normalize all datasets (from both experimental data and theoretical curves) prior to the feature extraction, dividing all rate values of the dataset by its maximum, according to Eq. 4.

$$\hat{r}_0(i) = \frac{r_0(i)}{\max(r_0)} \quad \text{Eq. 4}$$

This way, the normalized values will be within the range from 0 to 1, avoiding this issue all together.

4.2. Comparison section

With these issues resolved, the theoretical curves can now be reliably compared to the experimental data based on their features. This is done through two tests, represented in Figure 8, which are done to every theoretical curve generated by every rate equation in the library.

As part of the first test the primitives are compared since they represent the trends of their respective datasets. If a theoretical curve does not have the same features as the experimental data, then it does not have similar features at all. If it does have the same primitives, then this curve moves on to the second test, where the tool verifies if the pressure of each extremes of the theoretical curve is within a tolerated range around the pressure of the corresponding experimental extreme. This is done by calculating the relative error of the pressure of the theoretical extreme relatively to the experimental one, according to Eq. 5, and then checking if this error is smaller than the tolerance value, which is given by the user.

$$\delta = \frac{|P_{Theor} - P_{Exp}|}{P_{Exp}} \quad \text{Eq. 5}$$

This comparison is done for all extremes of the theoretical curve (except for the first and last ones, whose pressure values are those of the first and last points of the curve, respectively, and will always be the same as the corresponding experimental extremes by definition). If this condition is verified for all extremes, then it is considered that the theoretical curve has similar features to the experimental data.

These comparison tests are performed for every generated theoretical curve. In the end, every rate equation that generates at least one curve that passes the tests is proposed as a possible model.

4.3. Ranking of models

Each proposed rate equation can generate multiple curves that pass the comparison tests. Therefore, it is necessary to choose the best curve generated by each equation. This requires some ranking criteria that reflect how similar the features of a curve are to the experimental features. The chosen criterion was the sum of relative errors (SRE) of all the extremes of the theoretical curve calculated through Eq. 5. For differentiate between two theoretical curves with the same SRE, the mean squared error (MSE) was chosen as a secondary criterion. Since at this point the theoretical curves are still normalized, this criterion is calculated by Eq. 6.

$$MSE_{norm} = \frac{\sum_{i=1}^N (\hat{r}_{0,Theor}(i) - \hat{r}_{0,Exp}(i))^2}{N} \quad \text{Eq. 6}$$

With these criteria, the best curve generated by each proposed rate equation is chosen.

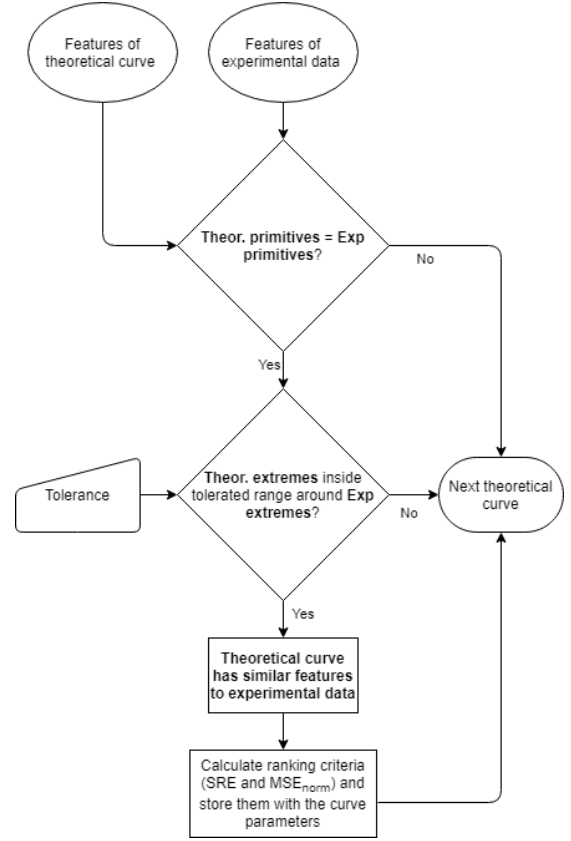


Figure 8 – Flowchart of the comparison algorithm (trapezoid – user input; rectangle – process; oval – output of previous branches; rhombus – decision; rounded rectangle – terminator)

4.4. Estimating k'

The objective of this tool is not only to propose rate equations for experimental data but also to propose some initial values for their parameters for an eventual regression. While these initial values were already determined for the equilibrium constants (from the lists of K_{Eq} values in Table 2), no initial value was proposed for the lumped kinetic constant k' . Therefore, this value is estimated for the best curves of each proposed rate equation. This is done by calculating the k' values necessary for the curve to intersect each point of the experimental data and the MSE of each resulting curve, according to Eq. 7. The average of these k' values is also calculated, as is the MSE of the curve with that value for k' . The chosen value for the initial guess for k' is the one that generates the curve with the lowest MSE. While this method is rudimentary, no fitting of the equilibrium constants was performed, so fitting this value to the experimental data would be of no use.

$$MSE = \frac{\sum_{i=1}^N (r_{0,Theor}(i) - r_{0,Exp}(i))^2}{N} \quad \text{Eq. 7}$$

With the initial rate equations to be proposed determined, as well as the initial guesses for their parameters, it is necessary to rank them. The ranking criterion to do so is once again the SRE. The secondary criterion to differentiate between rate equations with the same SRE is the MSE, more specifically the one associated to the best initial guess for k' , calculated through Eq. 7.

4.5. Output of tool

Once the proposed initial rate equations are ranked, the output of the tool is completed. First, a graphical representation of the experimental data and its recognized features appears, similar to that of Figure 3. Then, for each proposed initial rate equation and in ranking order, a graphical representation of its best curve and of the experimental data appears (Figure 9), accompanied by some text output containing the rate equation itself and the parameter values of

its best curve (Figure 10). In the end, a table with a summary of the output appears.

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

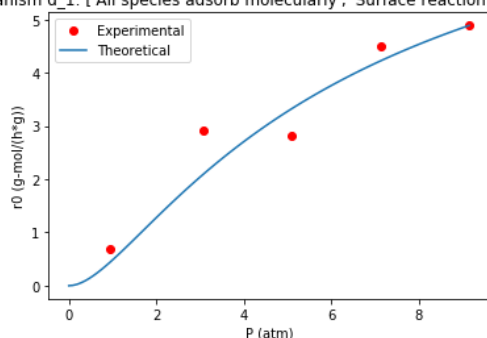


Figure 9 - Example of output graph with theoretical curve of proposed rate equation

Rate equation:

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

A = C3H6

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

$$k' = 9.102e+00$$

$$K_A = 3.000e-01$$

Ranking criteria:

$$SRE = 2.140e-02$$

$$MSE = 2.089e-01$$

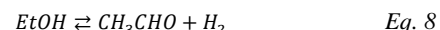
Figure 10 - Example of output text string

Once the output is presented, the user can choose to run the comparison and proposal algorithm again for a different tolerance value.

A flowchart describing in more detail the steps taken by the tool from start to finish is represented in Figure 11.

5. Case study

After developing the tool, some case studies were performed. One of them was on ethanol dehydrogenation, which has its global reaction present in Eq. 8.



The analysed dataset was obtained from the work of Franckaerts and Froment[17]. While there were six datasets of this reactions, extracted for different temperatures, only the one extracted at 275 °C will be discussed here. A graphical representation of that dataset is present in Figure 12.

This dataset was extracted for a feed composed, in molar fractions, of 0.865 of EtOH and 0.135 of water, which acts as an inert component. The rate equation determined by the researchers is present as Eq. 9. The values determined for k' and K_{EtOH} were $4.03 \text{ mol} \cdot \text{h}^{-1} \cdot \text{g}^{-1}$ and 0.40 atm^{-1} , respectively.

$$r_0 = \frac{k' \cdot K_{EtOH} \cdot P_{EtOH}}{(1 + K_{EtOH} \cdot P_{EtOH})^2} \quad Eq. 9$$

Eq. 9 corresponds to the rate equation present in the library (Table 1) that assumes that all species adsorb molecularly and the RDS is the surface reaction. Therefore, if the tool works properly, this rate equation should be among the proposed ones.

This dataset was analysed by the tool, for a tolerance value of 0.5. The full runtime, including the time necessary for all the user input, was approximately 4 minutes. The features recognized from the experimental data are present in Figure 13.

For this dataset, only one rate equation was proposed, which precisely corresponds to the one determined by the researchers. This is because from all the tested rate equations (taking into account that only one reactant is involved in the reaction), only this one generated a curve with the same primitives as the experimental data. This rate equation is present as Eq. 10.

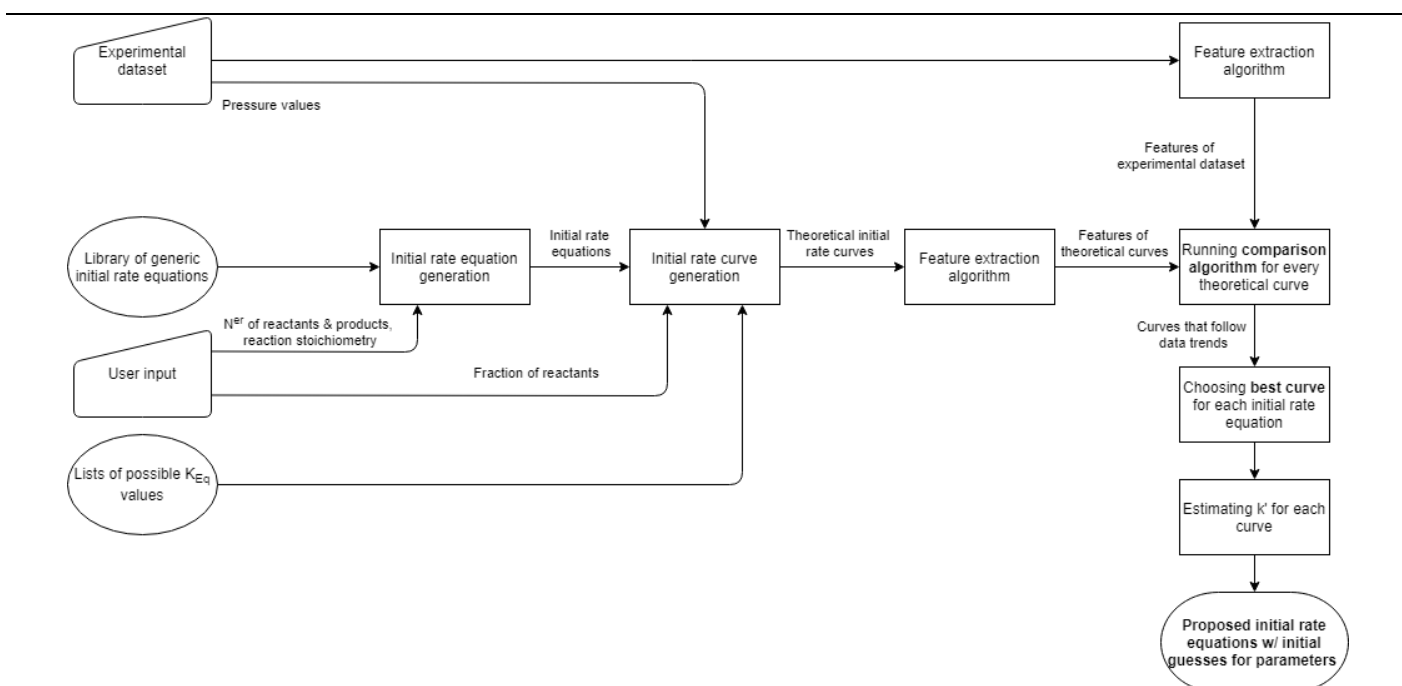


Figure 11 - Full flowchart of the tool (trapezoid – user input; rectangle – process; oval - information included in tool; rounded rectangle – output of tool)

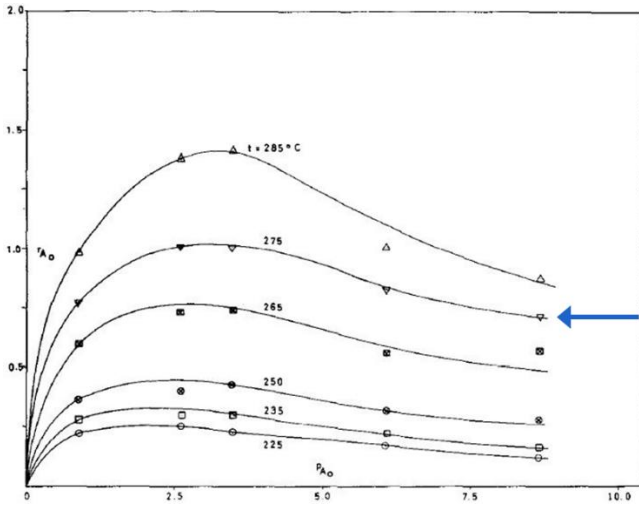


Figure 12 - 275 °C ethanol dehydrogenation experimental dataset [17], indicated by arrow (markers – experimental points; solid line – model fitting). Adapted with permission from [17]. Copyright 1964 Published by Elsevier Ltd.

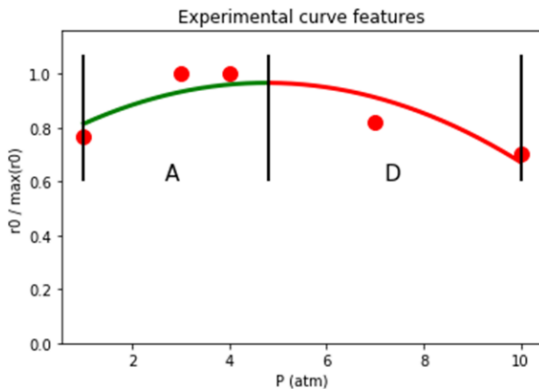


Figure 13 - Recognized features of the 275 °C ethanol dehydrogenation dataset

$$r_0 = \frac{k' \cdot K_A \cdot y_A \cdot P}{(1 + K_A \cdot y_A \cdot P)^2} \quad \text{Eq. 10}$$

The k' and K_A values proposed by the tool for Eq. 10 are $4.2 \text{ mol} \cdot \text{h}^{-1} \cdot \text{g}^{-1}$ and 0.50 atm^{-1} , respectively, while the ranking criteria SRE and MSE had the values 0.312 and $4.31\text{E-}3$, respectively. The constant values proposed by the researchers and the tool are again present in Table 3.

Table 3 - Constant values proposed by the researchers and the tool

	k' ($\text{mol} \cdot \text{h}^{-1} \cdot \text{g}^{-1}$)	K_A (atm^{-1})
Franckaerts and Froment [17]	4.03	0.40
Tool	4.2	0.50

Comparing the values proposed by the tool for k' and K_A ($A = \text{EtOH}$) to the ones proposed by the researchers, it can be observed that they are very similar to each other. This is indication that the tool performed well. The corresponding curve is present in Figure 14.

Analysing the ranking criteria of this rate equation, it can be concluded that this rate equation is only proposed by the tool for this dataset for a tolerance value larger than 0.312 (since the curve only has one extreme from which to calculate the SRE). This is due to the fact that only a limited number of K_{Eq} values can be used to generate theoretical curves and also due to the fact that the K_{EtOH} value associated with the experimental data is small, and as mentioned before, a variation on small K_{Eq} values has a greater effect on the features than the same variation on larger K_{Eq} values. Therefore,

even though the value of K_{EtOH} (K_A) proposed by the tool is similar to the one proposed by the researchers, it is still different enough for the resulting curve to have features (more specifically the extremes) different enough from the experimental ones, preventing Eq. 10 from being proposed for smaller tolerance values.

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

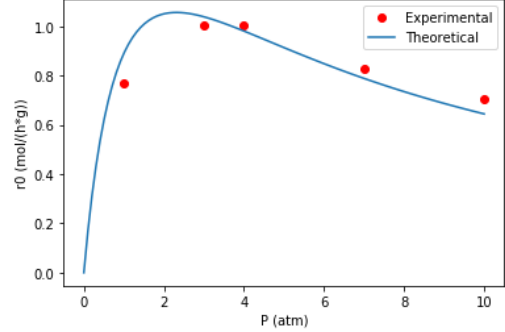


Figure 14 - Best curve of rate equation proposed for the 275 °C ethanol dehydrogenation dataset

As mentioned before, other case studies were performed to test the tool. These served to test several aspects of the tool, such as its generic performance, influence of the data characteristics (distribution of data points along the pressure axis, data variability, experimental noise and double points from repeated experiments) on the performance of the tool and the results of the analysis of dataset for which the determined rate equation is not present in the library of the tool.

From these case studies, it was observed that the tool proposed the same rate equation that the researchers did for most cases. When it failed to do so, it was due to characteristics of the data, such as experimental noise, outliers, the lack of variability of the data and the absence of data points on pressure ranges. The case study of the ethanol dehydrogenation was specifically chosen to be presented in this article because this specific case study is a good example of the qualitative analysis of an experimental dataset serving to screen several different models in search of those that can explain the data, since for this case only one model was proposed. It also shows the influence of the chosen tolerance value on the proposed models, since the SRE value of the proposed rate equation was relatively high, when compared to other case studies, while still being an example of a successful performance of the tool

6. Conclusions

In this work a tool that automatically proposes initial rate equations for catalytic reactions based on experimental kinetic data was developed and tested. This tool contains a pre-defined library of rate equations, from which theoretical curves are generated. The proposal of rate equations results from the screening of the ones in the library, made by comparing the generated curves from each equation with the experimental data. This automatic process reduces the expertise level required from researchers, thus being the main achievement of the tool. Another achievement is that the tool can perform this process much faster than a human researcher would manually. These two achievements mean that the tool fulfils the objective of mitigating the bottleneck present in the kinetic modelling process.

One of the advantages of the tool is the fact that its library of rate equations covers numerous possibilities for heterogeneous catalysis, within certain constraints. This comes from the fact that these equations were deduced from typical mechanisms for catalytic reactions and cover the different possibilities for both reactant adsorption and RDS. This way, the correct rate equation can be proposed to an acceptable number of experimental datasets.

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.

In future developments, the initial rates limitation can be surpassed by updating the library for it to be able to consider the presence of product. This could be done by re-deducing the rate equations without assuming the inexistence of products in the system. The fully developed version of the tool should ideally analyse datasets of rate as a function of variables other than total pressure, such as molar fraction of reactants. The resultant increase of obtained kinetic information 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.

Notations

a, b, r, s – stoichiometric coefficients of species A, B, R and S, respectively

δ – relative error

k' – lumped kinetic constant

K_r – equilibrium constant of surface reaction step

K_{Eq} – equilibrium constant

K_i – adsorption constant of species i (pressure units⁻¹, $i=A, B, R, S$)

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

P – pressure (pressure units)

P_{Exp}, P_{Theor} – pressure values of extremes of experimental data and theoretical curve, respectively (pressure units)

r_0 – initial rate of reaction (rate units)

\hat{r}_0 – normalized initial rate value

SRE – sum of relative errors

y_i – molar fraction of reactant i

References

- [1] K. Kakaei, M. D. Esrafil, and A. Ehsani, "Introduction to Catalysis," in *Interface Science and Technology*, vol. 27, Academic Press, 2019, pp. 1–21.
- [2] O. A. Hougen and K. M. Watson, "Solid catalysts and Reaction rates," *Ind. Eng. Chem.*, vol. 35, no. 5, pp. 529–541, 1943.
- [3] 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.
- [4] K. Toch, J. W. Thybaut, and G. B. Marin, "A Systematic Methodology for Kinetic Modeling of Chemical Reactions Applied to n-Hexane Hydroisomerization," *AIChE J.*, vol. 61, no. 3, pp. 880–892, 2015, doi: 10.1002/aic.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] S. Siradze, "Automated kinetic feature extraction from Open Access data," Universiteit Gent, 2019.
- [13] 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.
- [14] 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.
- [15] M. E. Davis and R. J. Davis, *Fundamentals of Chemical Reaction Engineering*. 2013.
- [16] M. A. Vannice, *Kinetics of catalytic reactions*. Springer Science & Business Media, 2005.
- [17] 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.