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 1.

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 1 - 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.

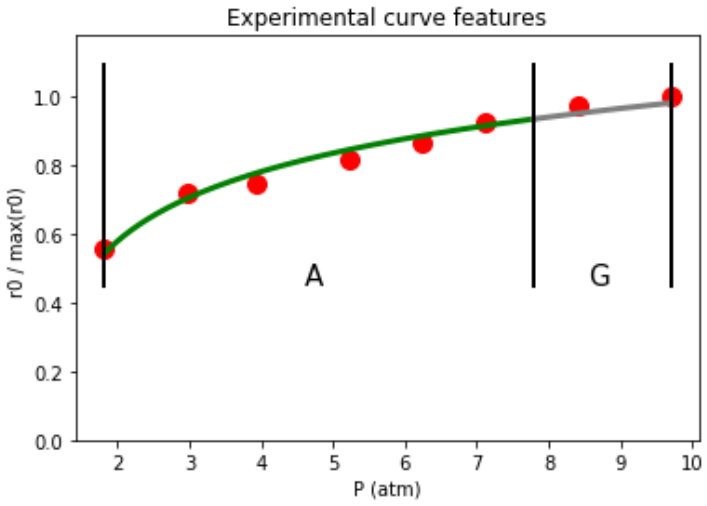
The following prompt asks for the user to write the desired name for the folder in which the output will be stored. Then, the tool presents the mechanisms in the library and asks the user for the mechanisms which are known *a priori* to not be possibilities. The mechanisms selected here will

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 permutated. 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 2.




```
Experimental primitives: ['A', 'G']
Experimental extremes (atm): [1.817     ,  7.7889696969697, 9.7         ]
```

Figure 2 - 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. 1.

$$r_0 = \frac{c_1 \cdot P^{e_1}}{\left(1 + c_2 \cdot \sqrt{P} + c_3 \cdot P + c_4 \cdot P^{e_2} + c_5 \cdot P^{e_3}\right)^n} \qquad \textit{Eq. 1}$$

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. 1 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. 2.

$$\delta = \frac{\left|P_{Theor} - P_{Exp}\right|}{P_{Exp}} \qquad \textit{Eq. 2}$$

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. 3.

$$MSE_{norm} = \frac{\sum_{i=1}^{N} \left(\hat{r}_{0,Theor}(i) - \hat{r}_{0,Exp}(i) \right)^2}{N} \qquad \textit{Eq. 3}$$

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. 4. This was also done for the average of all the determined k' values.

$$MSE = \frac{\sum_{i=1}^{n} \left(r_{0,Theor}(i) - r_{0,Exp}(i) \right)^2}{n} \qquad \textit{Eq. 4}$$

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 3.

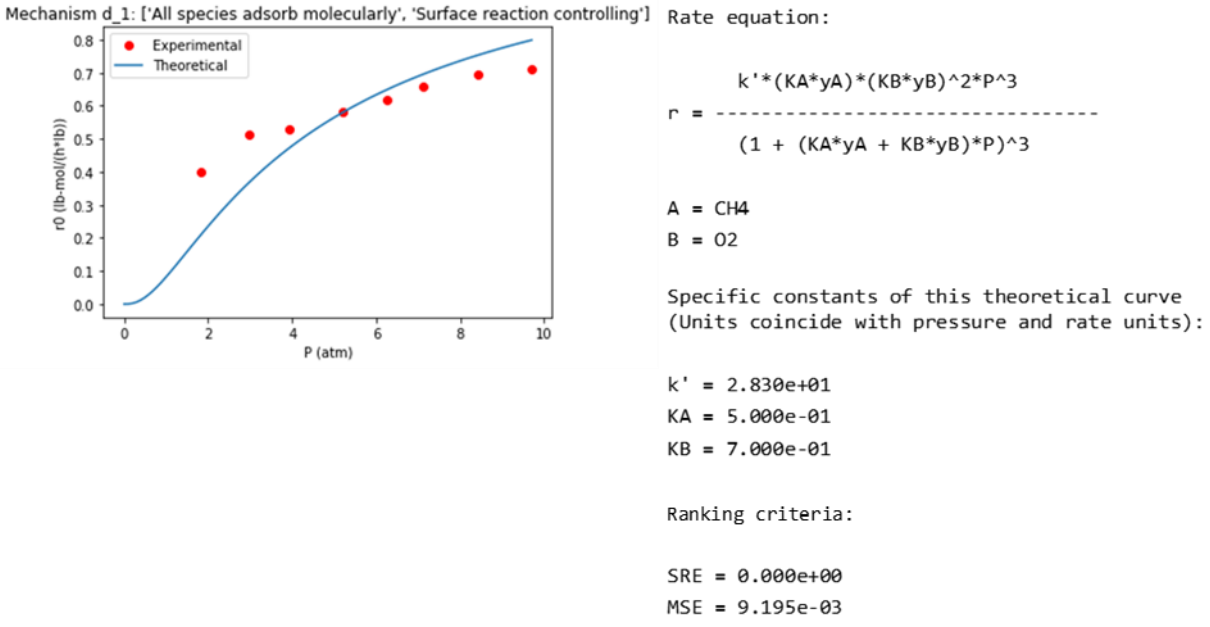


Figure 3 - Example of output graph with best theoretical curve (left) and output text string (right)

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 4.

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 4 - 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 differ 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 2) and the graphs of the best theoretical curves of the proposed initial rate equations (see Figure 3). In this folder will also be an Excel file with all of the text output (see Figure 3), as well as the table with the summary of the output (see Figure 4).