

5. Fitting/Optimization

This section will explain how one can fit almost any numerical value (rate constants, initial concentrations, Troe factors, third body enhancements, energy of activation, starting temperature, etc.) against an experimental or “fabricated” dataset. One might wish to use fabricated datasets to optimize numerical values such as initial concentrations of species to values which would minimize the presence of some harmful intermediate species, maximize certain products, reduce/increase temperature and so on. Note that Kintecus will actually fit the parameters at EXACTLY the time your data was measured. Unlike other programs, Kintecus DOES NOT interpolate a function against your data and then fit the values against this interpolation. There is absolutely no need to “clean” your data, suggest interpolation methods nor specify timing meshes against your experimental data since Kintecus calculates values at exactly the times you specify in your experimental datafile.

The Fitting Procedure

Fitting Options

The fitting procedure in Kintecus can be specified with the inclusion of the –FIT switch on the command line. The –FIT switch has many options:

-FIT:a:b:c:d:e:f [:g:h:i]

FIT Switch Option	Description	Possible Values	Default Value
a	Fitting Algorithm	1, 2, 3	1
b	Comparison Operator	1, 2, 3	1
c	User Dataset Filename	Any allowed text filename	FITDATA.TXT
d	Tolerance	$1 - 10^{-14}$	1×10^{-7} for Fitting Algorithms 1 and 3 1×10^{-5} for Fitting Algorithm 2
e	Maximum Iterations Allowed	1 - 32767	9000
f	Starting Vectors	$1 \times 10^{-100} - 1 \times 10^{10}$	1×10^{-35}
g	Starting “Temperature”	$1 \times 10^{-100} - 1 \times 10^{10}$	1×10^6
h	Number of Cycles to Stay At Current Temperature Before Reducing Temperature	1 - 32767	25
i	Percent Temperature Reduction	0 - 0.99999	0.20 (20%)

Table 6. Options for the –FIT switch. Specifying a “D” on any option field will force Kintecus to use the Default Value.

The first three options (fitting algorithm, comparison operator, dataset filename) are the most important. Options d, e, and f can be used with all three optimization algorithms. Options g, h, and i are only for the Simulated Annealing fitting algorithm method 3. Specifying a “D” on any option field will force Kintecus to use the Default Value.

There are three fitting/optimization algorithms: 1=Meade & Nelder [23], 2=Powell [1] and 3=Simulated Annealing [24]. With each optimization algorithm, a user can choose from three comparison operators: 1=relative least squares, 2=standard least squares and 3=a proprietary function the author has devised. The comparison operators compare the experimental/fabricated data and the simulated data and respond back to the optimizer with a number showing this difference. Once the user chooses a fitting/optimization algorithm and comparison operator, he must then supply the experimental or fabricated data in a text file named FITDATA.TXT. The default filename, FITDATA.TXT, can be change to a user-defined filename. The only

requirements for the datafile is that “Time(s)” must be in the first column, first row, “END” must be in the first column last row, the measured/fabricated species names/Temperature(K) must be in the first row. The species names/Temperature(K) can be in any order on the first row. You can insert comments in this file or comment out lines by placing a “#” or a quote ‘ ’ as the first character on a line. If there is missing data point(s) for a species then the cell must have an UPPERCASE “N” as the first letter, so “NaN” or “None” or “Nothing” or just “N” are all allowed to represent no data at that time point. An example datafile (created in Excel, then saved as a TAB delimited text file) is shown below (this datafile is used in enzyme fitting test #4, FITTEST4.BAT) :

Time(s)	E	S	ES	EIS
3.55E+00	1.52E-09	5.90E-02	2.48E-09	1.38E-13
2.38E+01	1.62E-09	5.34E-02	2.38E-09	1.33E-13
5.33E+01	1.77E-09	4.56E-02	2.23E-09	1.24E-13
8.28E+01	1.94E-09	3.84E-02	2.06E-09	1.14E-13
1.12E+02	2.13E-09	N	1.87E-09	1.04E-13
1.42E+02	2.33E-09	N	1.67E-09	9.24E-14
1.71E+02	2.55E-09	N	1.45E-09	8.05E-14
2.01E+02	N	N	1.23E-09	6.83E-14
2.30E+02	N	N	1.01E-09	N
2.60E+02	N	N	8.13E-10	N
2.89E+02	N	N	6.33E-10	N
.
.
.
1.65E+03	4.00E-09	N	4.39E-14	N
1.68E+03	4.00E-09	N	4.39E-14	N
1.71E+03	4.00E-09	N	4.39E-14	N
1.74E+03	4.00E-09	N	4.39E-14	7.31E-16
1.76E+03	4.00E-09	N	4.39E-14	7.31E-16
1.80E+03	4.00E-09	N	4.39E-14	7.31E-16
END				

Table 7. Sample user supplied spreadsheet containing experimentally obtained data ready for fitting.

The Fit Switch Option d is the optimization tolerance. This tolerance sets the minimum value between successive fits that must be met before the data is considered optimized. Decreasing this value will cause more optimizations (longer execution time but numbers that are more accurate) to be performed. Conversely, increasing this value will cause less optimizations (shorter execution time but numbers that are less accurate) to be performed. It is very important to note that the tolerance is a minimum threshold needed to stop optimization when successive fits (compares between the data and simulated data) falls at or below this value. If the data and simulated data falls at or below the tolerance, the optimization will halt even if the data-simulated comparison is **terrible** (very large difference). In other words, just because Kintecus states that optimization is finished, there is a good chance it has not. Kintecus had to stop the optimization because the program was going in circles (trapped in a local minimum) and

that the user should try another optimization method, comparison operator or a different starting guess(es). For more help, see the Optimizing Tips section below.

Optimization algorithm #2 (Powell method) is fairly sensitive to the tolerance. It is suggested that you do not set a value above 1×10^{-5} . Setting smaller values (under 1×10^{-8}) can result in a substantial longer optimization run in some cases. Optimization algorithms #1 and #3 are more insensitive to the tolerance than the Powell method.

The Fit Switch Option e set the maximum iterations allowed. If the number of optimizations reaches this value, the program will quit. Please note that the average amount of optimization iterations for just two values is between 400-1200 optimizations at a tolerance of 1×10^{-6} .

The Fit Switch Option f sets the starting vectors. The optimization of values, N, is actually an optimization of starting vectors of size, N, in vector space N (or N+1 for the fitting algorithms #1 and #3). Consult references for a full explanation.

The Fit Switch Options g, h and i are for the Simulated Annealing Optimization method #3. The g option is the starting “Temperature”, this is NOT the physical, real thermodynamic temperature of the system, but an analogy to a starting data “cooking” temperature. Refer to references for a full explanation. Option h is the actual “annealing” or “cooking” time in cycles. It is also the number of simulations to be performed before the data “temperature” is reduced by a percent represent by Fit Switch Option i.

There are many examples (examine the various Kintecus command line options in the .BAT files) in the FITTESTS sub-directory demonstrating various parameters and models that Kintecus can fit data to.

What and How Can I Fit/Optimize ?

Selection of numeric parameters for fitting/optimization can easily be done simply by appending the number (actually your guess) with a question mark, “?”. The sample model spreadsheet below (See [Enzyme_sheet.xls](#) in the FITTESTS sub-directory, this model is actually run with FITTEST4.BAT) shows three rate constants that have been selected for fitting. These models can also be executed by loading the Excel or Star-Office workbook: [Enzyme_Regression_Fitting.xls](#) and then clicking the **RUN** button located on the **CONTROL** worksheet. (note the extremely poor guesses! But Kintecus actually obtains the exact answer that is **orders of magnitude away!**):

# Non-competitive Inhibition # of an enzymatic reaction		Equilibrium Constants
# Example spreadsheet. You can see comments by moving pointer above the red triangles		
# *Note, if you wish to use this model with KINTECUS, be sure to save		
# it as a Text Tab Delimited sheet, and don't forget to also save it		
# as a regular Excel spreadsheet or you lose all your formatting and notes!!		
1? E+S==>ES		75
1? ES ==> E+S		
1? E+I==>EI		0.07543
112687259.7 EI==>E+I		
9.57E+06 ES+I==>EIS		2.53
3782608.696 EIS==>ES+I		
5.53E+07 EI+S==>EIS		17.87
3.09E+06 EIS==>EI+S		
7.13E+06 EI+P==>EIS		35.5
2.01E+05 EIS==>EI+P		
1.14E+05 ES==>E+P		

END

One can fit initial temperature, initial concentration, external concentration, flux, rate constant, Arrhenius factor, energy of activation, and any parameters for Troe, Lindemann, SRI, Landau-Teller and third body enhancement factors. [There are sample batch files, FITTESTx.BAT, located in the FITTESTS subdirectory.](#) These sample batch files when run will fit various parameters to various data. Some use different –FIT options, such as larger tolerance settings (-FIT switch option ‘d’ in Table 5 above) which cause the optimization to finish much quicker. You can examine the one line Kintecus call in those files and modify them.

Once Kintecus is finished optimizing your parameters, it will write a file name [optout.txt](#) in the current directory. This file will contain the final optimized results, the final data to simulation difference and the total amount of iterations. These values are also duplicated to the screen. In addition the output concentration file, CONC.TXT will contain the concentrations/temperature for the final optimized model. You should plot this file, CONC.TXT against your experimental data and compare them. Note that the heading for CONC.TXT is not present, but it is stored in the file HEADINGS.TXT, which can be inserted at the top of the CONC.TXT file. Since most, if not, all the time series in CONC.TXT should line up with your experimental values, you should be able to easily plot the residuals (your data minus the simulated data) and see if the residuals are small and “noisy”, this indicates a good fit.

Some Tips To Quickly Start Optimizing

Once you have saved your data in a text file named FITDATA.TXT (this is the default name for your datafile Kintecus looks for, you can name it to another filename by using the `-FIT` switch). You can quickly get started by using the `-FIT:2:3:FITDATA.TXT` switch on the Kintecus Command line. **ALSO** try `-FIT:1:3:FITDATA.TXT` on the Kintecus Command line and compare the results. In addition you should also re-run with the `-FIT:1:1:FITDATA.TXT` and also the `-FIT:2:1:FITDATA.TXT` and compare all four sets. Be prepared, optimizing more than one value can take some time.

Optimizing stiff reactions such as a combustion model can be quite challenging. You will most likely get OVERFLOW ERRORS. This is due to the optimizer picking values that are making the system stiffer than where it started. A quick solution is to reduce the accuracy in the parameter spreadsheet to 1/100, 1/1000 or even 1/10,000 of the current accuracy value. You should end up with an accuracy in the range of 1×10^{-8} to 1×10^{-12} . Another problem with optimizing combustion models is that the optimizer might put the system where the temperature is too high (or low) for one or more species' thermodynamic coefficients' range. An additional option has been provided for the `-THERM` switch: **FORCE**, ie. `-THERM:THERM.DAT:FORCE` or `-THERM:D:FORCE`. This will "force" the optimizer **AND** the integrator to use temperature values that are within all the species' thermodynamic coefficients' range and should alleviate your OVERFLOW problems.

Be Careful!

Just because you managed to get a great fit to your data and it passes every statistical test you can throw at it, the fitted values can be COMPLETELY WRONG! This can be especially true when you are fitting more than one value. For example, if a user tried to optimize all three expanded Arrhenius options at the same time at one temperature. Depending on your starting guess, one user might get $A=9.20E+16$, $m=-0.6$ and an $E_a=1.1841$ KJ, but another user might get $A=1.10E+15$, $m=0.588$, $E_a=2.57$ KJ and at 298 K both sets yield the overall rate constant 4.08×10^{14} ! Who's right? That depends on further molecular detail on the system. Can E_a be estimated and frozen in the optimization? Maybe the Arrhenius factor can be calculated or referenced from literature and frozen during the optimization? Maybe you should just optimize k , and forget about the three expanded Arrhenius parameters? Other examples can be shown! Be careful!