

Fitting to Differential Equations in *Igor Pro*

Marc Faggin and Alexander Hoepker

Department of Chemistry and Chemical Biology,
Cornell University, Ithaca, NY 14853

Igor Pro is commercially available at wavemetrics.com. The program described is available by request from either Marc Faggin (mff9@cornell.edu) or Alexander Hoepker (ach42@cornell.edu).

This manual is intended to guide you through the process of fitting data to a set of differential equations. The procedure is designed to be user-friendly and requires no knowledge of programming while still maintaining flexibility in model design, curve fitting strategy and the numerical methods employed.

a) Importing the model

We begin by opening the *master* procedure file ‘*CollumKinetic 5000*’ in *Igor Pro* and compiling the file at the bottom left of the window. Next, we need to import the model described by a set of differential and linear equations. We construct the model in a simple text editor such as *Textpad*. It is imperative to save the file as a plain text file. The file must be in a format as shown in figure 1.

```
2 // number of observables variables
***
2 // number of non-observables variables
***
3 // number of kinetic rate constants
*** // differential equations
dO1dt = -k1*NO2*O1 + k2*O2
dO2dt = k1*NO2*O1 - (k2 + k3)*O2
dNO1dt = k3*O2
*** // linear equations (NO only)
NO2 = NO2_Init - (O2 + NO1)
***
```

Figure 1. Example of a text file format

A set of three asterisks (***) is used to separate the five sections of the model input. Entries after a double-dash are ignored by the program and merely serve as commentary. Spaces between terms are not mandatory. *O1* stands for observable species *one* and *NO1* stands for non-observable species *one*. They must be ordered in increasing numerical sequence with the observables preceding the non-observables (i.e. *O1*, *O2*, ... *NO1*, *NO2*, ...). The differential term in the differential equations are written without the

usual division sign. If there are no differential equations or linear equations, do not reserve a line but instead type ‘***’ directly following the previous three stars. A total number of five three-asterisk sets are required. There is one important limitation in the format of the linear equations; they cannot contain the initial concentration of an observable species as a variable. However, it can be expressed explicitly by a numerical value.

Having saved the text file, it can now be opened in *Igor Pro* by selecting *Import Model* under the *Collum* menu (the *Collum* menu will only appear after compiling). *Igor* will now convert the text file into an Igor procedure. But before doing so, you are first asked to save the new file. Open the newly created file in *Igor Pro* and compile it. This newly created procedure is particular to a model and runs in concert with the *master* procedure.

b) Importing Data

We must next import our data. It is assumed that the data has been preprocessed (e.g. converted to concentrations and corrected for time). Prepare the data in *Excel* in the following format (Figure 2).

	A	B	C	D	
1	t_01	O1_01	O2_01		
2	0	0.005	0		
3	10	0.00497507	2.4306E-05		
4	20	0.00495027	4.7308E-05		
5	30	0.0049256	6.906E-05		
6	40	0.00490106	8.9625E-05		
7	50	0.00487665	0.00010906		
8	60	0.00485238	0.00012742		
9	70	0.00482823	0.00014476		
10	80	0.00480421	0.00016113		
11	90	0.00478031	0.00017658		
12	100	0.00475654	0.00019115		

Figure 2. Example of the *Excel* format for the case of one data set.

The first column contains the x-data (independent variable) while all other columns contain the y-data (dependent variables). The variables must match those defined in the model and are annotated with an underscore and suffix to distinguish between different data sets. This concept will be described later. The suffix is arbitrary but must be consistent among data sets. The *Excel* file can be loaded in *Igor* by selecting *Load Waves* → *Load Excel File* under the *Data* menu. Choose the desired directory, select the *Excel* file and the specific worksheet. Then click *Do It*. *Igor* should now have created a table with your data. Note that *Excel* files cannot be uploaded while the particular file is open in *Excel*. It is not necessary to have an installed version of *Excel* on the platform you are working on. *Igor* will recognize the *xls* extension name and load the file properly. Note that closing the table you just imported does not delete the data. It is still stored in the data browser and ready for recall by the program.

c) Setting the Initial Concentrations

Select *Setup Concentrations* under the *Collum* menu and enter the initial concentration of your non-observable species. *Igor* automatically imports the initial concentrations of the observable species from the data set and therefore must not be entered manually.

d) Setting the Initial Parameters

Select *Solve the Model* under the *Collum* menu, enter initial guesses for the parameters and choose an appropriate integration method from the drop-down menu. *BDF* is the default method and works in many cases. Click *Continue*. Now select *Kinetic Parameter Panel* under the *Collum* menu and adjust the parameters with the up- and down arrows provided. The increment can be adjusted by powers of ten at the top right. The curves will adjust interactively. After having manually adjusted the parameters click *Done*.

e) Setting the Epsilon Parameter

Epsilon values are used to calculate partial derivatives with respect to the fit coefficients which in turn are used to determine the search direction for coefficients that give the smallest chi-square. Although it is not necessary to set epsilon values, it is highly suggested as it may greatly improve the quality and efficiency of the fit. Select *Set Epsilon for Curve Fit* under the *Collum* menu. By default the epsilon values are one tenth of the initial parameter values (each parameter receives its own epsilon value). This default scaling factor of *0.1* may be modified in the top box. Alternatively, the epsilon values may be adjusted individually. Choosing epsilon values is an empirical process and should be thought of as a separate adjustable parameter for curve fitting. Begin by sequentially setting the scaling factor to 1, 0.1 and 0.01. Click *Set* when done entering epsilon values.

f) The Curve Fitting Dialog

Select *Curve Fitting* under the *Analysis* menu. In the *Function and Data* tab select *MyFitFunc* as the user-defined function from the drop-down tab and select *allDataY* under the *Y Data* tab in the folder '*gFolderWaves*'. In the *Coefficients* tab, select *KK* from the drop-down *Coefficient Wave* and select *KK_epsilon* from the drop-down tab just to the right. *KK* and *KK_epsilon* are both also contained in '*gFolderWaves*'. Opening the coefficients tab will produce an error that can be ignored. Click *ok* and proceed. By selecting the *KK_epsilon* wave, the epsilon values entered earlier are loaded. In the window below, the parameters can be fixed by checking the mark under '*Hold?*'. The parameters can be constrained to a range by selecting *From Coefficient List* under the *Constraints* drop-down tab.

Select the *Output Options* tab and choose *fitDataY* under the *Destination* drop-down menu in '*gFolderWaves*'. Check the boxes *Error Analysis* and *Covariance Matrix* if you wish to calculate confidence limits and a correlation matrix. Click *Do It* to start the fit. The Correlation Matrix can be displayed by selecting *Calculate Correlation Matrix* under the *Collum* menu. This will only work if the checkbox referred to earlier was checked. Residuals are plotted by selecting *Display Global Fit* under the *Collum* menu. Open the curve fitting dialog and select *autotrace* under the *Residual* menu in *Output*

Options tab (*Display Global Fit* can be called prior to fitting). The residuals are now plotted against the fitted data. One may also plot observables together with nonobservable species on one common time axis. This is done by selecting *Display stacked graphs* in the *Collum* menu.

Alternatively, one may skip the curve fitting dialog by calling *Quick Fit* in the *Collum* menu. Select from the three options listed: (1) epsilon values, (2) residuals and (3) covariance matrix. Click *Continue* to start the fit. *Quick Fit* facilitates repeated curve fitting but does not allow one to restrict parameters or select parameter ranges.

Note that the parameters are continuously updated from fit to fit. A second fit will therefore use the best estimates of the previous fit. Epsilon values are not automatically scaled to the output parameters after each fit. To obtain properly scaled epsilon values, the epsilon parameter panel must be called.

There are two likely error messages after submitting the curve fitting request:

(1) In the command line: “Error Code 2”. The fit may have converged but the computer encountered accuracy problems during numerical integration. This may be solved by scaling concentrations and/or parameters to values near unity. If model variables (i.e. parameters, concentration) vary by more than roughly five orders of magnitude such errors may occur.

(2) Window popup: “Singular Matrix Error”. This may indicate a whole host of problems. It is most likely that the initial guess is too far removed from the minimum for convergence to occur. Second, there may exist a syntax problem or conflict in the model. This is of course much harder to identify.

g) Exporting data and graphics

Graphics are exported by activating the desired graph window and selecting *Export Graphics* in the *Edit* menu. It is best to export the file as a *bitmap* as it will generate high resolution images that are recognized across operating system platforms. The image is copied to the clipboard and can be pasted into programs such as *Word* or *PowerPoint*. The fitted data can be exported as a tab-delimited text file by selecting *Export Fit Data* in the *Collum* menu.

h) Creating Job Folders

It is possible to import multiple data sets simultaneously. The data is loaded as described earlier. Each data set is distinguished by a user-defined suffix. An example of a multiple data set spreadsheet is shown below in figure 3.

	A	B	C	D	E	F	G	H	I	J
1	t_005	O1_005	O2_005	O3_005	O4_005	t_0075	O1_0075	O2_0075	O3_0075	O4_0075
2	0	0	0	0	0.0074	0	0	0	0	0.0074
3	60	0.00017	-1.4E-05	4.02E-05	0.007203	60	0.000162	4.46E-06	-5E-05	0.007284
4	90	0.000163	7.17E-05	-2.5E-05	0.00719	90	0.000279	2.18E-05	6.35E-05	0.007036
5	120	0.000205	1.78E-05	-9.1E-05	0.007268	120	0.000408	6.88E-06	-5.6E-05	0.007041
6	150	0.000364	-5.6E-05	-0.0001	0.007196	150	0.000441	0.000141	8.07E-05	0.006737
7	180	0.00053	1.79E-05	-5.9E-05	0.006911	180	0.000508	-1.8E-05	-4.8E-05	0.006959
8	210	0.000483	-6.7E-05	-9.4E-05	0.007078	210	0.000647	0.000101	2.24E-05	0.00663
9	240	0.000543	3.66E-05	1.66E-05	0.006804	240	0.000739	1.06E-05	1.48E-05	0.006636
10	270	0.000646	-1.6E-05	-9.3E-05	0.006864	270	0.000897	1.33E-05	-4.8E-05	0.006538
11	300	0.000607	-3.4E-05	3.73E-05	0.00679	300	0.00097	3.88E-05	-1.6E-05	0.006407
12	330	0.000697	-2.2E-05	-2.5E-05	0.00675	330	0.001128	0.000123	-6.5E-05	0.006214
13	360	0.000833	-5.1E-05	-2.2E-05	0.00664	360	0.00117	8.55E-05	-3.2E-05	0.006176
14	390	0.000864	0.000105	4.9E-05	0.006381	390	0.001333	3.18E-05	-5.2E-05	0.006087
15	420	0.000881	-8E-05	-0.00011	0.006706	420	0.001477	1.29E-05	-5E-05	0.00596
16	450	0.001	-3.2E-05	-6.8E-07	0.006432	450	0.001632	7.24E-07	8.69E-06	0.005758
17	480	0.001232	9.82E-06	-1.7E-05	0.006175	480	0.001755	1.5E-05	2.15E-06	0.005628
18	510	0.001247	4.57E-05	4.9E-05	0.006058	510	0.001841	0.000111	-3.7E-05	0.005485
19	540	0.001255	-9.5E-06	6.78E-06	0.006148	540	0.001959	1.8E-05	9.37E-06	0.005413

Figure 3. Example of the *Excel* format for the case of two data sets.

The example illustrates two data sets with four observables each. The data sets differ in the amount of the initial concentration of one reagent (0.005 M for one and 0.0075 M for the other). Again, the choice of suffix is arbitrary but must be consistent within a data set.

Igor places the imported data sets into the *root* folder in the data browser. The data browser can be opened under the *Data* menu. Make sure that the *red arrow* in the data browser points to the root folder. Now select *Create Fit Folder* in the *Collum* menu. Enter a folder name such as '*double_fit*' for the fit you wish to perform. Spaces and symbols are allowed. Also, the name can contain no more than 20 characters and must start with a letter. Click *Continue* and select the data set(s) (either one data set or multiple sets in the case of a global fit) you wish to fit to. *Igor* will generate a subfolder inside of the *root* folder and copy the selected data into this subfolder. By selecting the check box, the *red arrow* mentioned earlier will automatically move away from the *root* to this newly created *subfolder* (the significance of this is explained later). Now click *Continue*.

The process of creating subfolders may be continued. The *red arrow* designates the parent folder in which the subfolder is created. One may therefore either create multiple subfolders in the *root* folder or create folders within subfolders. The purpose of this folder system is to quickly switch between jobs of varying data sets while fitting them to the same model. Each job (subfolder) is independent from the other and therefore contains initial concentrations, parameter values, epsilon values and more that are particular to this job folder. One moves among folders by moving the *red arrow* to the according folder in the data browser. When calling the curve fitting dialog or the *Quick Fit* menu, the data to be fit is automatically the one contained in the folder selected by the *red arrow*. Subfolders are optional but greatly improve work flow and efficiency - even if only one data set is being analyzed. The disadvantage to simply fitting data in the root folder is that the program will automatically fit the entire data set contained in the *root* folder. Individual data sets cannot be selected. Also, new data cannot be imported without deleting the original data from the data browser.

The subfolder lends itself naturally to *global fitting*. By importing multiple data sets into a job folder, the program will fit the model to all data contained in this folder. The menu item *Setup Concentrations* will now ask for the initial concentration of all non-observable species for each data set.