50 Data Fit Module

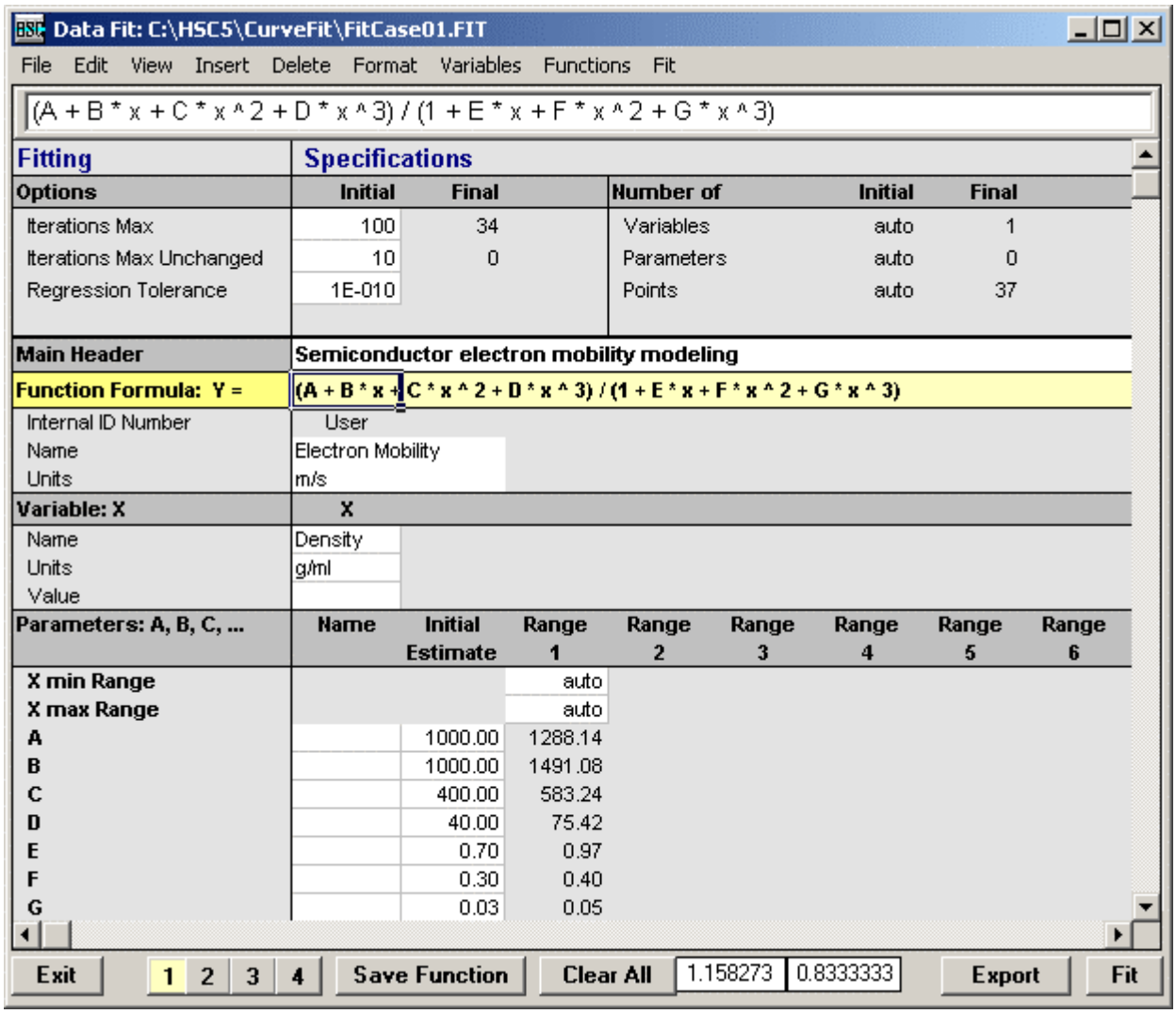


Fig. 1. Settings dialog of the HSC-Sim Curve Fit Module.

The chemical process models which are created by HSC-Sim are often based on experimental data expressed in the function format. The HSC-Sim Data Fit module may be used to convert experimental data points into functions. Of course, these functions may also be created by using some other data fitting applications, however, the built-in HSC-Sim Data Fit module offers more flexible links to the other HSC-Sim modules.

The HSC-Sim Data Fit module may be opened from the Excel Model menu with **"Tools, Data Fit…"** selection. Fig. 1 shows the Settings dialog of the Data Fit module with one simple default sample case.

The HSC-Sim Data Fit module offers versatile multi-regression, curve fitting and statistical properties for linear and nonlinear functions. Up to 20 independent variables (X1, X2, X3, …) and 100 parameters (A, B, C, …) may be used in these functions (models). The fitting may be based on up to 65536 data point sets. The user may specify the function formula freely or use one of the 370 built-in functions. Data Fit utilizes the robust Levenberg-Marquardt method with double precision to perform nonlinear regression.

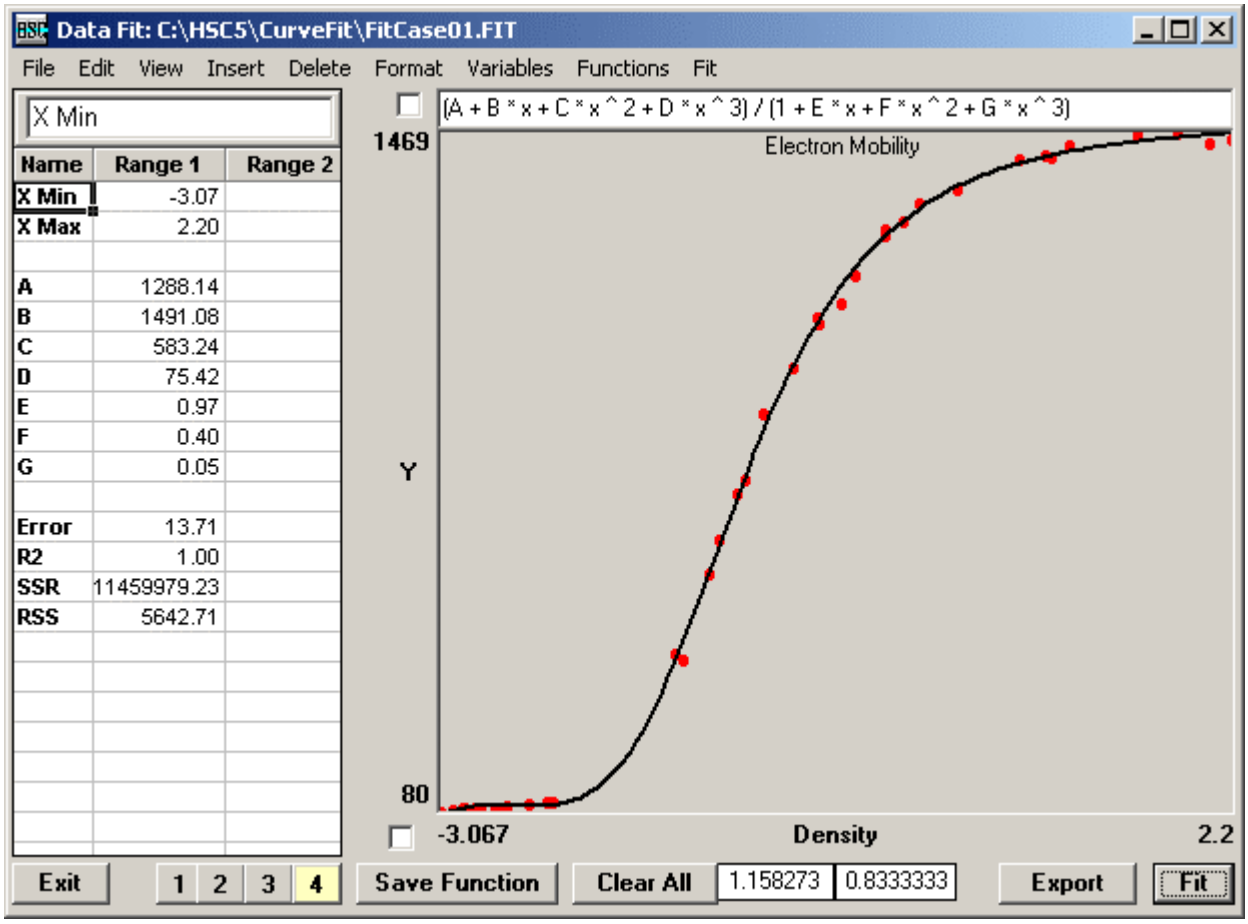


Fig. 2. Data Fit module shows the results after the user presses Fit.

The user may easily test the default sample case by pressing **Fit**. This will run the fitting routine and will release the results in tabular format on the left side of the form and in graphical format on the right side of the form, Fig. 2.

Note that the Tool Tip Text gives brief information about the Data Fit controls. You may see this information by moving the mouse pointer over the control. This manual will give more detailed information.

Some sample cases may be opened by selecting “File, Open, …” from the menu.

Note that you may use Copy – Paste operations from the Edit menu, but also with the right mouse button to copy data and text from other Windows applications to the HSC-Sim Database module. For example, you may easily copy XYdata sets from Excel to the Data Fit module.

50.1 Step by Step Instructions (Quick Start)

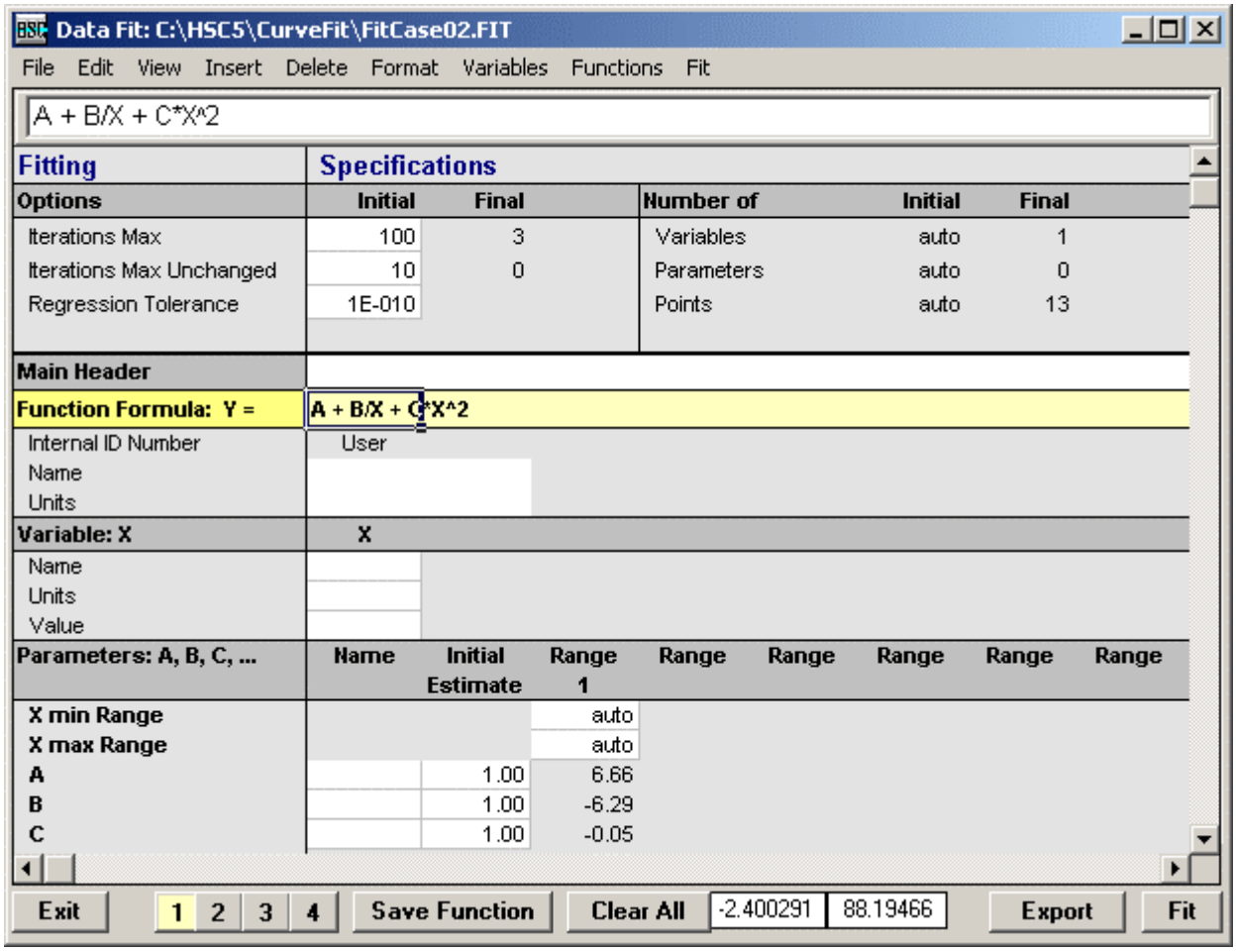


Fig. 3. The user may type a new function after pressing **Clear All**.

Basically, only the function formula and experimental data points are required to carry out the data fit. However, it is also useful to take some other aspects into account:

1. **Start new fit project**

   Press **Clear All**, Fig. 3.

2. **Settings Sheet**

   Activate Settings sheet by pressing "**1**" in the bottom left-hand corner.

3. **Number of Variables X**

   Specify the number of variables using the "**Variable**, …" menu selection, Fig. 3. If one variable is selected then the "X" character must be used in the function, but if more variables are selected then the X1, X2, X3, etc. names must be used. You may use either upper or lower case characters (X or x).

4. **Function Specification (Yellow background cell in Fig. 3)**

   This is the most important step. It is recommended to use X as the variable name and A, B, C, etc. characters as parameter names. The same function formula syntax may be used as in Excel, however, do not use the "=" character at the beginning.

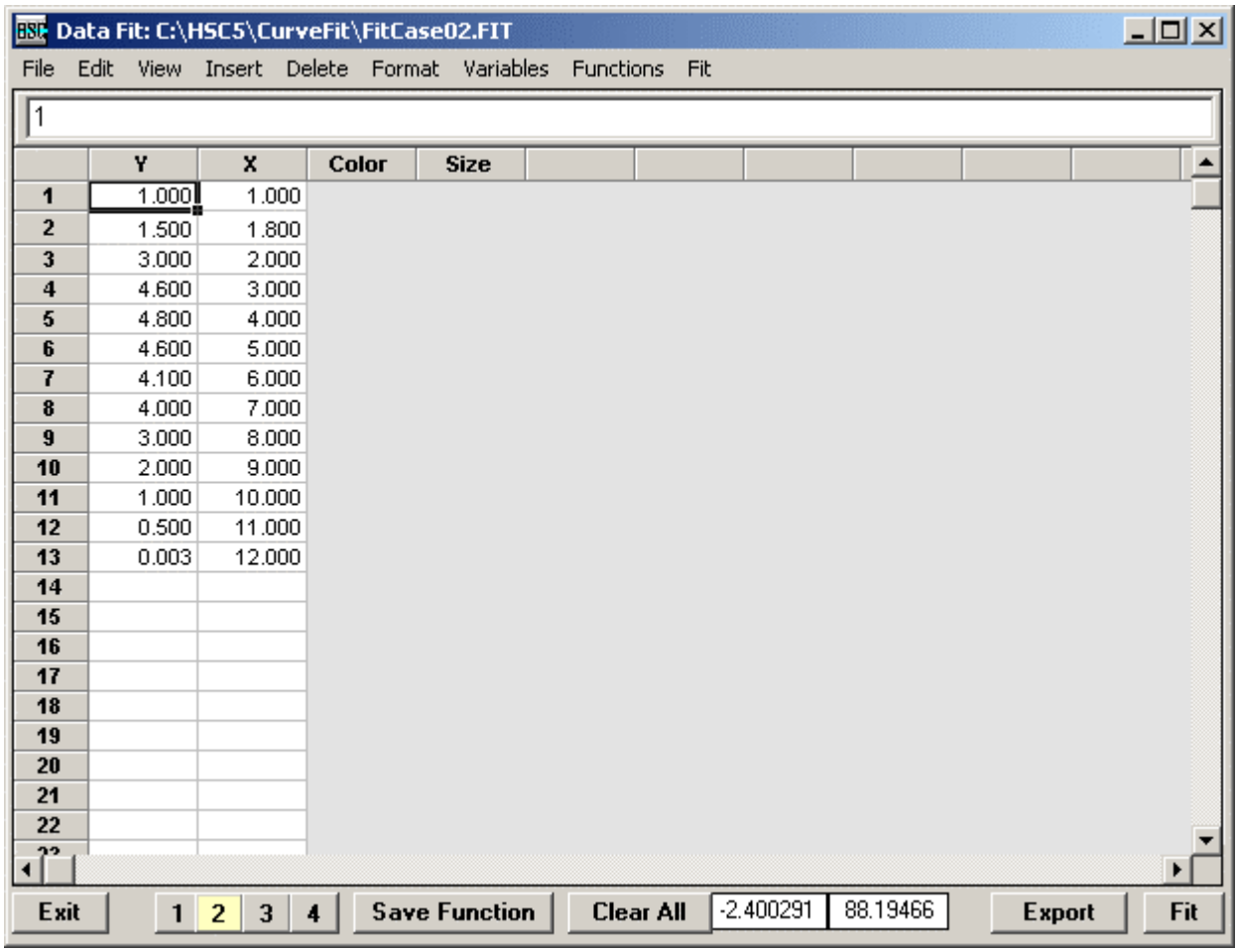


Fig. 4. The X – Y data is always needed for data fitting. Note that Ydata is always specified in the first column.

5. Other Specifications

Usually no other specification information is needed apart from the function formula because this is created automatically, Fig. 3. However, sometimes the ***initial estimate*** is needed for the parameters. The default initial estimate is always 1. The user may specify all the information given in cells with a white background to describe the case in more detail. Please do not edit the other cells.

6. Parameter Range

This specification will make it possible to divide the X-number scale into ranges. Parameters will be fitted separately for each range. However, this property is not available in Data Fit version 1.0. It is only included to make the file format compatible with future versions, Fig. 3.

7. Data

Function formula and data are always needed. Select "**View, Data…**" from the menu or press "**2**" to see the data sheet, Fig. 4. Type Y and X data onto the sheet, and you may also use Copy-Paste to import data from Excel. Y is always given in the first column and X or X1, X2, X3, … in the following columns. Note that you may use also text rows within the data.

Color and size refer to the color and size of the points in the diagram. The user does not need to specify these values. They may be needed only to change the default settings (red and green).

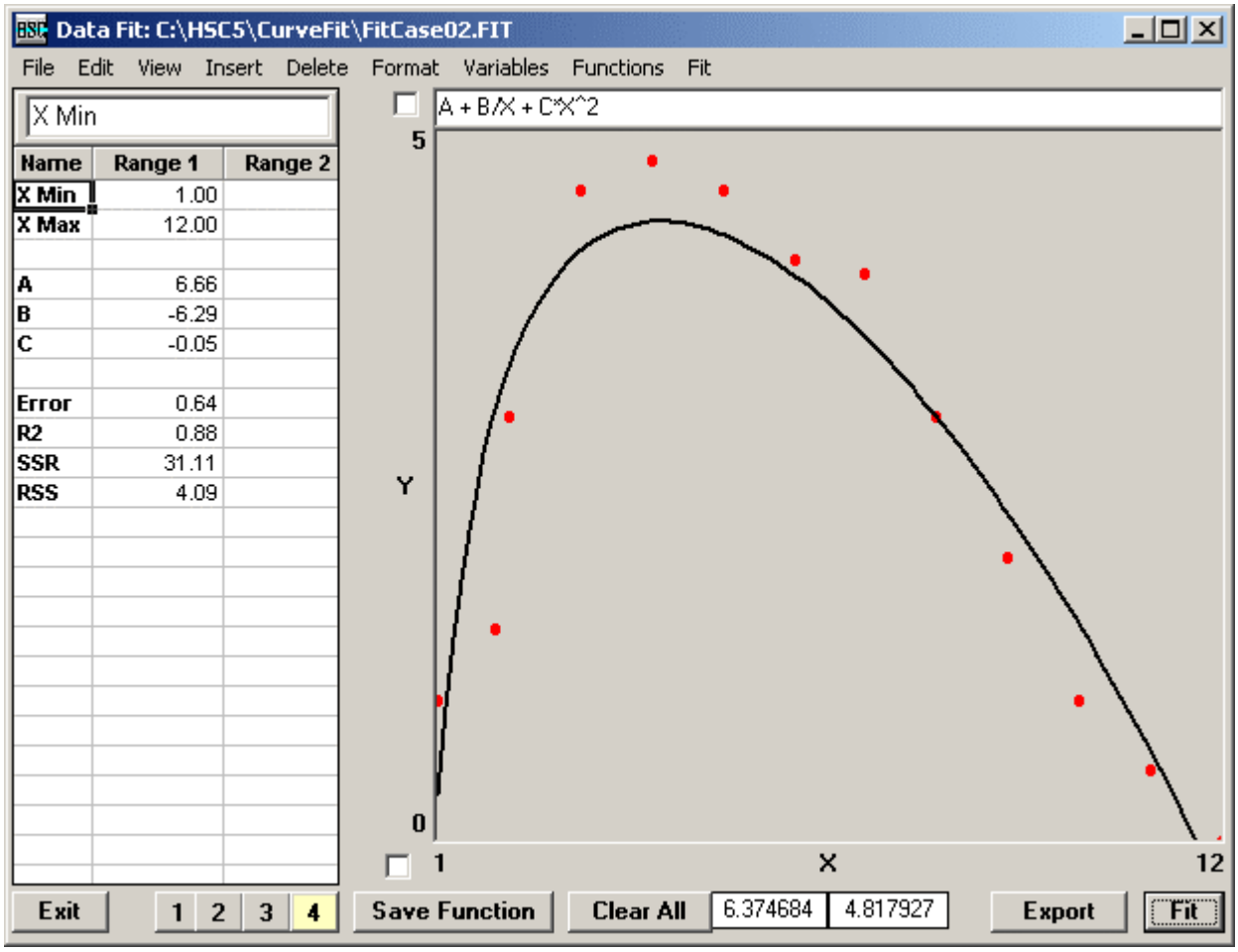


Fig. 5. Results of Data Fit in tabular and graphical format.

8. Fitting

After you have specified the function and data you may try fitting by pressing **Fit**, Fig. 5. The parameter A, B, C, … values are given in the left-hand table and the data points and the function are shown on the diagram.

9. Results

The Data Fit module gives the results in tabular and graphical format. The module gives the following results:

X Min, X Max	X Data Range
A, B, C, ..	Function Parameters
Error	Model Standard Error
R2	Model R2 Value
SSR	Regression Sum of Squares
RSS	Residual Sum of Squares

10. No Solution

If the Data Fit module does not find the solution you may try to:

- Edit the function formula directly in the cell in Fig. 5.
- Edit the initial estimates in Fig. 3.

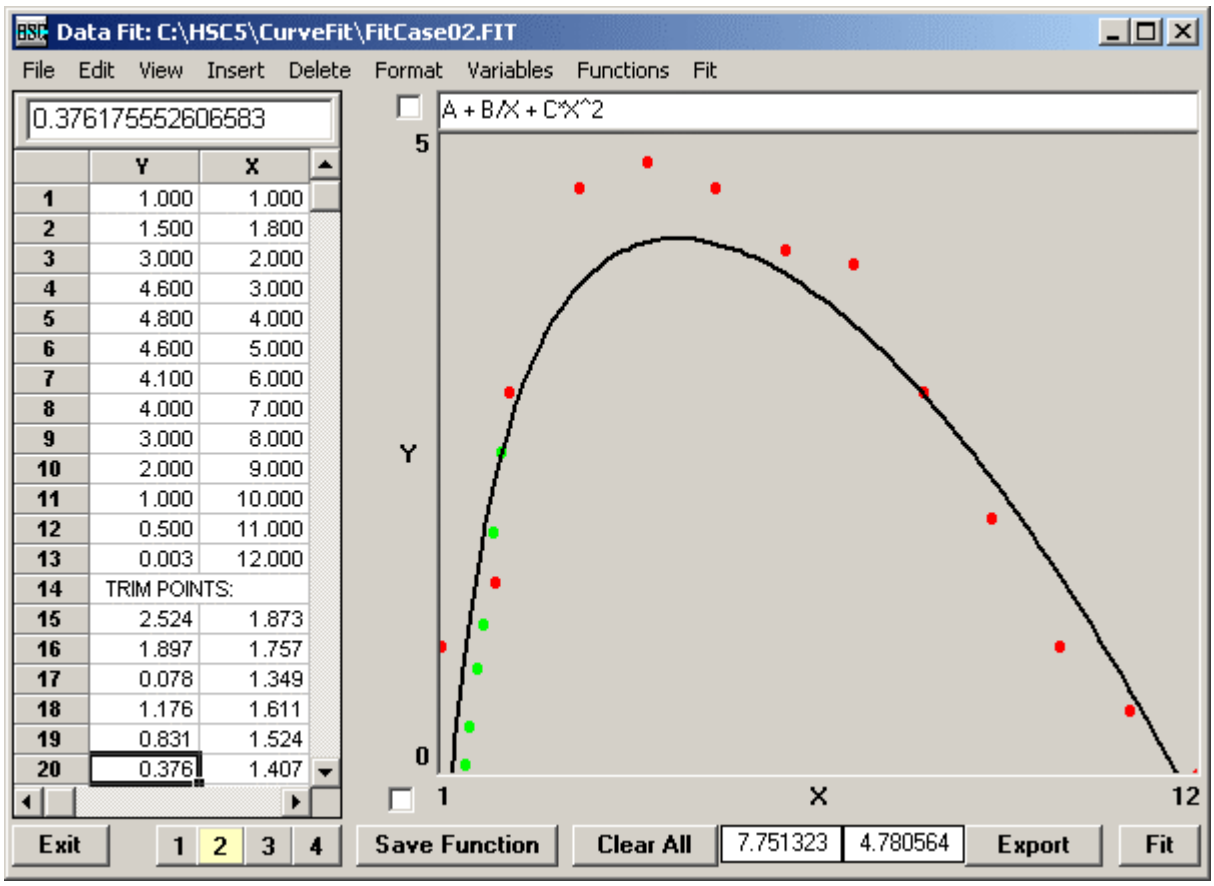


Fig. 6. The user may add new points (green) by clicking the mouse.

11. Manipulating Results

Sometimes experimental points are missing but the user still knows, on the basis of other information sources, the approximate path of the function curve. In these cases the user may easily add new green points to the diagram by clicking with the left mouse button and removing points with the right mouse button, Fig. 6. The user may also manually add and remove trim points to and from the data table, Fig. 6.

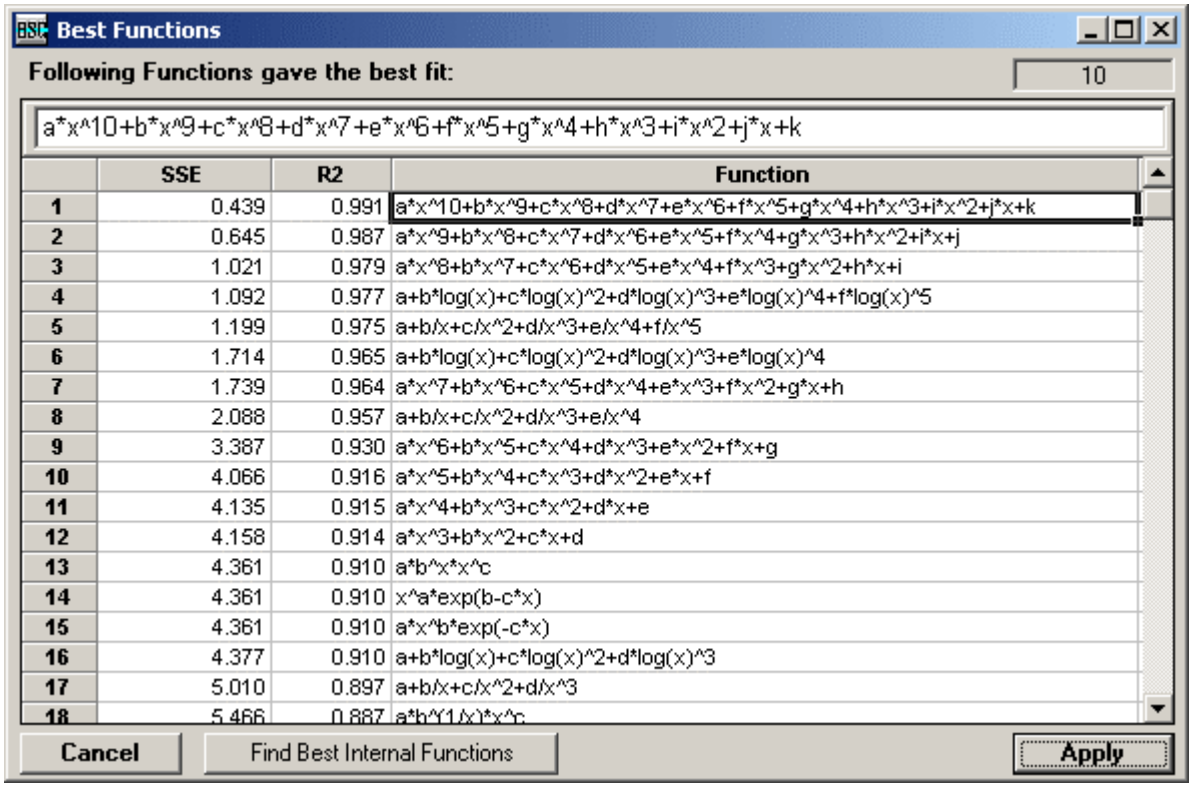
The user may totally remove the green trim points with the right mouse button, but not the original red points. The red points are converted to text values and not taken into account when the user runs the Fit routine.

12. Build in Functions

The menu selection "Fit, Find Best Internal Functions" will test all the internal functions and shows the function which gives the best fit for the given data points, Fig. 7.

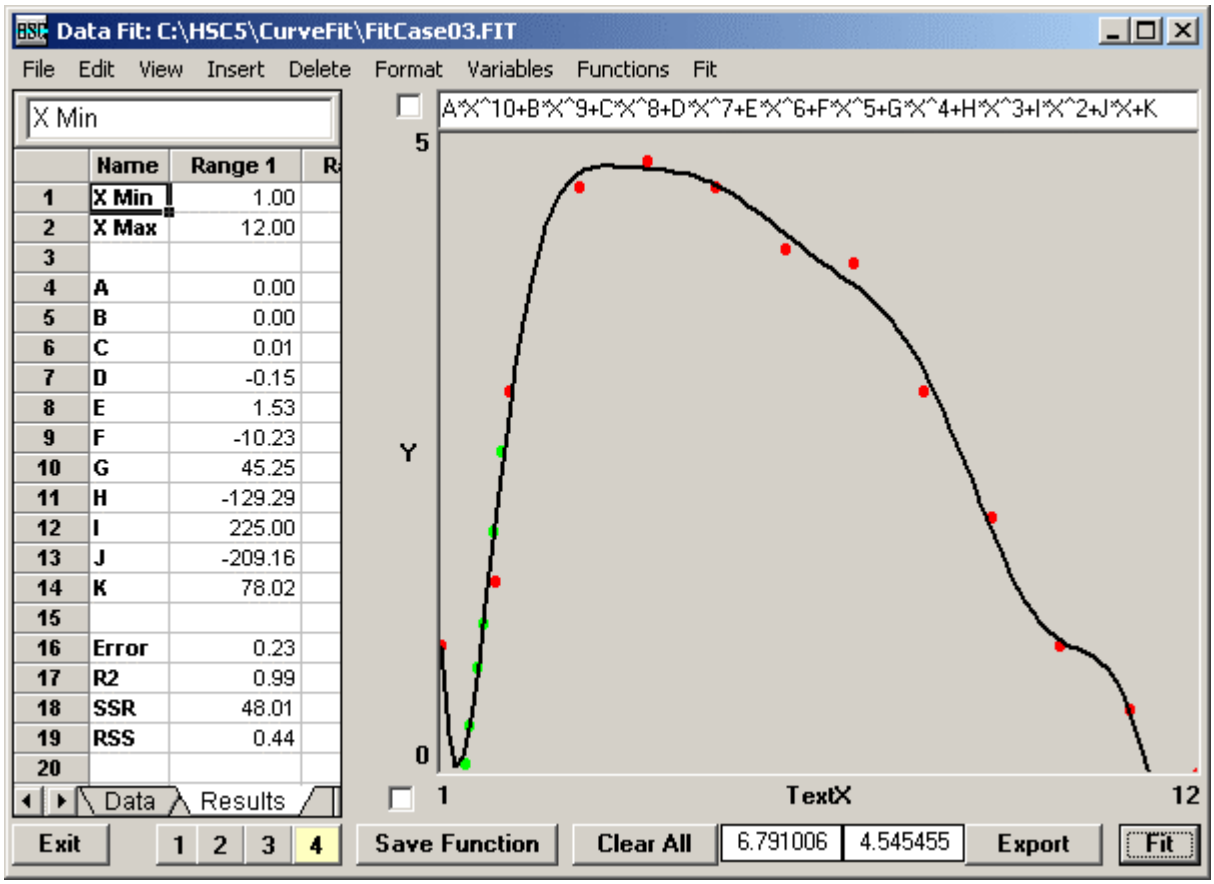
Quite often the most complicated functions give the best fit for a small number of data points, Fig. 8. However, it is very important to realize that this is often purely an illusion caused by the low number of data points. Usually it is recommended to use the simplest functions for the element distribution and content ratio coefficients, Fig. 9.

A large number of data points and knowledge of the background may sometimes justify the use of the complicated functions.



	SSE	R2	Function
1	0.439	0.991	a*x^10+b*x^9+c*x^8+d*x^7+e*x^6+f*x^5+g*x^4+h*x^3+i*x^2+j*x+k
2	0.645	0.987	a*x^9+b*x^8+c*x^7+d*x^6+e*x^5+f*x^4+g*x^3+h*x^2+i*x+j
3	1.021	0.979	a*x^8+b*x^7+c*x^6+d*x^5+e*x^4+f*x^3+g*x^2+h*x+i
4	1.092	0.977	a+b*log(x)+c*log(x)^2+d*log(x)^3+e*log(x)^4+f*log(x)^5
5	1.199	0.975	a+b/x+c/x^2+d/x^3+e/x^4+f/x^5
6	1.714	0.965	a+b*log(x)+c*log(x)^2+d*log(x)^3+e*log(x)^4
7	1.739	0.964	a*x^7+b*x^6+c*x^5+d*x^4+e*x^3+f*x^2+g*x+h
8	2.088	0.957	a+b/x+c/x^2+d/x^3+e/x^4
9	3.387	0.930	a*x^6+b*x^5+c*x^4+d*x^3+e*x^2+f*x+g
10	4.066	0.916	a*x^5+b*x^4+c*x^3+d*x^2+e*x+f
11	4.135	0.915	a*x^4+b*x^3+c*x^2+d*x+e
12	4.158	0.914	a*x^3+b*x^2+c*x+d
13	4.361	0.910	a*b^x*x^c
14	4.361	0.910	x^a*exp(b-c*x)
15	4.361	0.910	a*x^b*exp(-c*x)
16	4.377	0.910	a+b*log(x)+c*log(x)^2+d*log(x)^3
17	5.010	0.897	a+b/x+c/x^2+d/x^3
18	5.466	0.887	a*b^(1/x)*x^c

Fig. 7. Functions which give the best fit.



	Name	Range 1
1	X Min	1.00
2	X Max	12.00
3		
4	A	0.00
5	B	0.00
6	C	0.01
7	D	-0.15
8	E	1.53
9	F	-10.23
10	G	45.25
11	H	-129.29
12	I	225.00
13	J	-209.16
14	K	78.02
15		
16	Error	0.23
17	R2	0.99
18	SSR	48.01
19	RSS	0.44
20		

Fig. 8. Results when the first function is applied in Fig. 7.

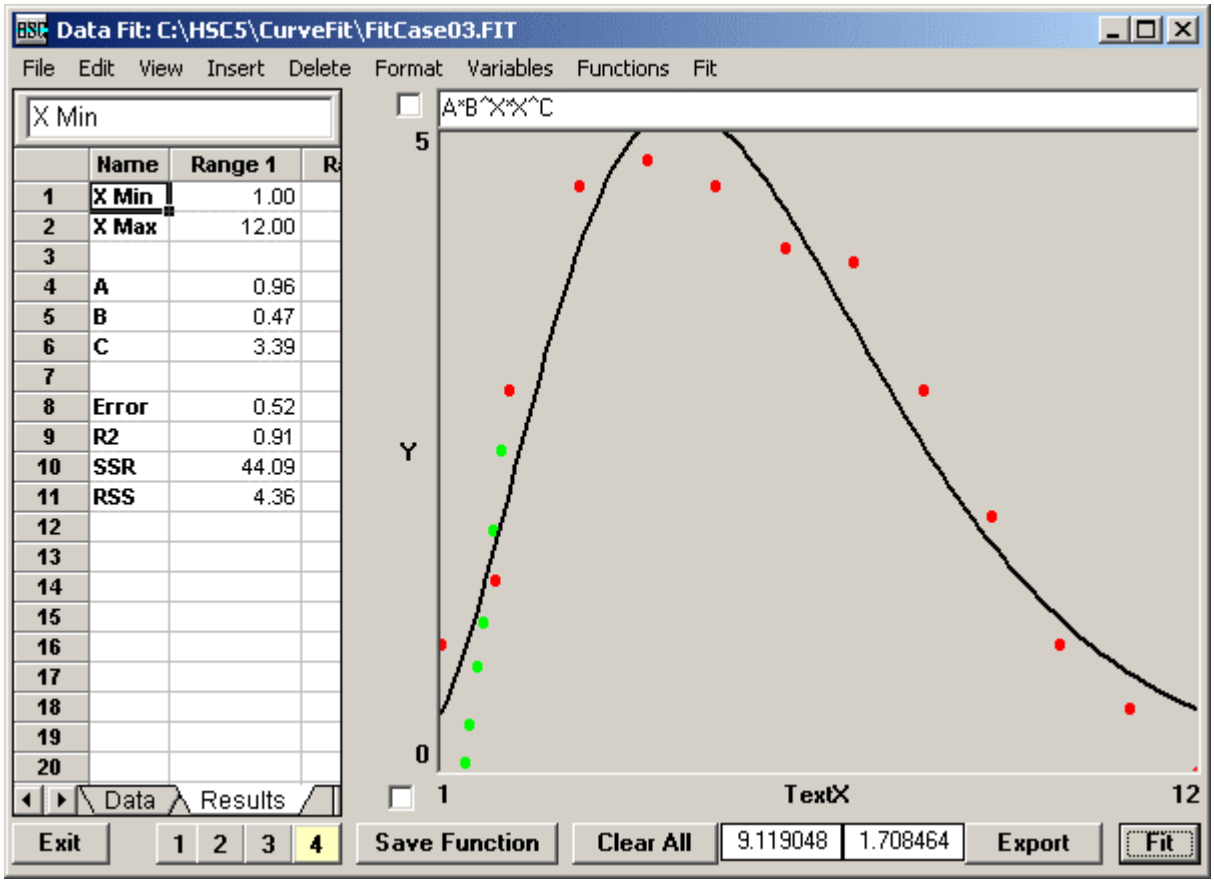


Fig. 9. Results when a more simple function is applied in Fig. 7.

50.2 Built-in Internal Functions

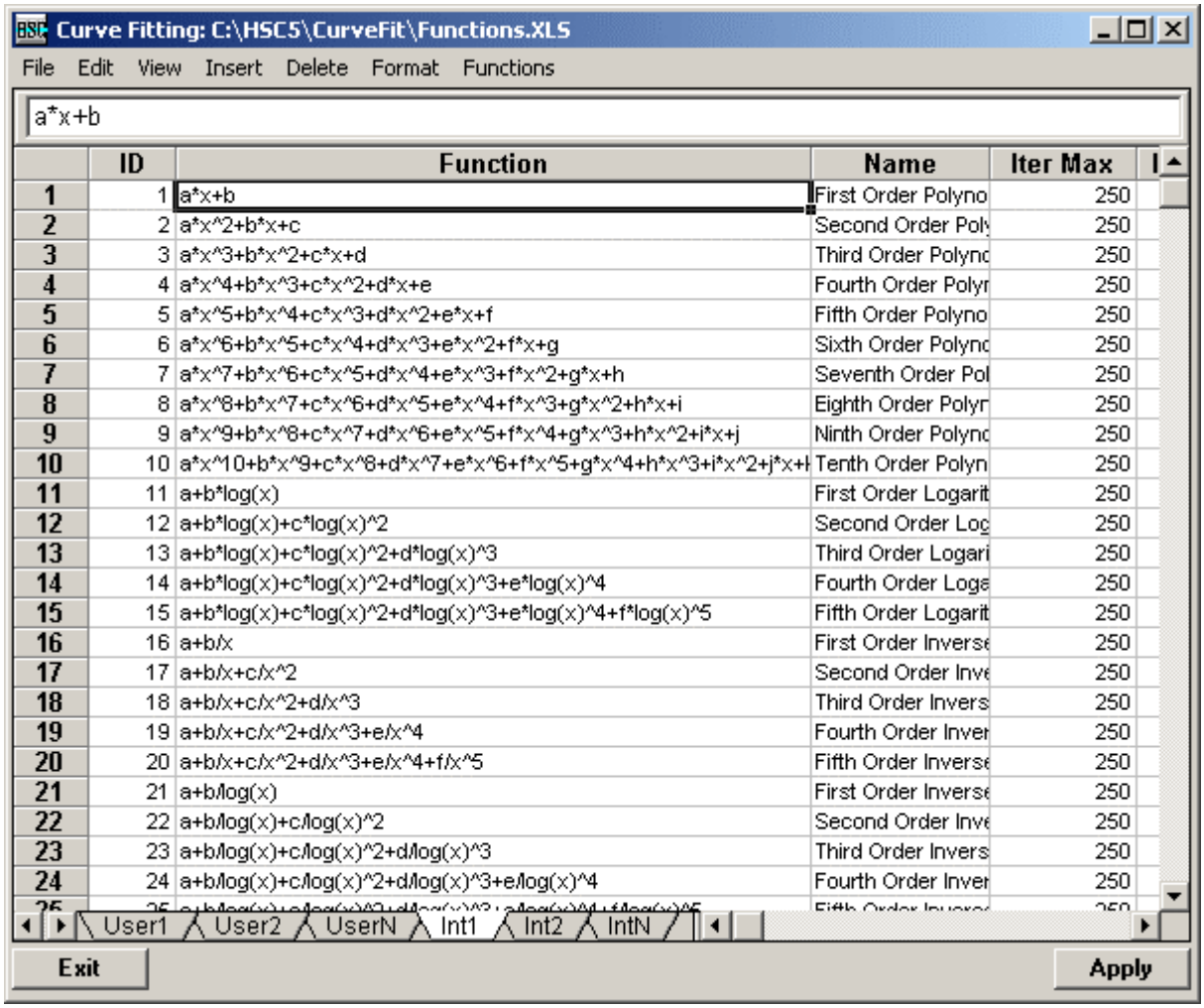


Fig. 10. The user may select one of the 370 built-in functions using the Apply button.

The possibility to specify the user's own custom functions makes the Data Fit module very flexible for many different types of data fit projects. However, sometimes the built-in internal functions offer a faster way to achieve results, see Chapter 50.1.11.

The Data Fit module offers 74 one-variable functions, 242 two-variable functions and 54 N-variable functions. The user may browse these functions from the menu by selecting "Functions", Fig. 9. The user may select any of these functions with the mouse and by pressing "Apply", Fig. 10.

The user may also save his/her own custom functions in this same Functions.XLS file by adding them to User1, User2 and UserN sheets, Fig. 10. When the changes are ready the user may save the new functions by selecting "File, Save, …", Fig. 10.