Simfit

Tutorials and worked examples for simulation, curve fitting, statistical analysis, and plotting.
http://www.simfit.org.uk

After fitting a function of two or more variables to a data set, the techniques for visualizing the best-fit model are more restricted than for cases with a single independent variable. To illustrate some possibilities, examples from fitting functions of two variables will be explained.

The first data set

For the first example, the data set contained in the SimFit test file `inhibit.tf1` will be used. This is for the case of enzyme kinetic data for a range of substrate concentrations at each of several fixed inhibitor concentrations. From the format of this data set it will be noted that the data are arranged for increasing sequences of S at each fixed inhibitor concentration. namely $I = 0, 1, 2, 3, 4$ and data should be arranged in such an ordered sequence to facilitate subsequent analysis.

The first model

This is taken from the library of enzyme kinetic models for two variables, namely the generalized inhibition model

$$f(x, y) = \frac{p_1 x}{p_2(1 + y/p_3) + x(1 + y/p_4)}$$

equivalent to the more familiar form

$$v(S, I) = \frac{VmaxS}{Km(1 + I/K_{is}) + S(1 + I/K_{ii})}.$$

The first technique: Plotting sections across a best-fit surface

Before the days of computers, nonlinear regression, and statistical analysis, this was cast in the form

$$\frac{1}{v} = \frac{Km(1 + I/K_{is})}{Vmax} \left(\frac{1}{S}\right) + \frac{(1 + I/K_{ii})}{Vmax}$$

and analyzed by plotting double reciprocal plots then re–plotting the slopes and intercepts to estimate the inhibition constants from the observed slope and intercept effects.

To see the modern way to do this type of analysis, open program **qnfit** from the [A/Z] option on the main SimFit menu, choose to fit a function of two variables, read in the test file `inhibit.tf1` followed by the model just discussed from the library of statistical models for functions of two variables, then fit in the EXPERT mode. This reads in the following starting estimates and limits appended to the data file
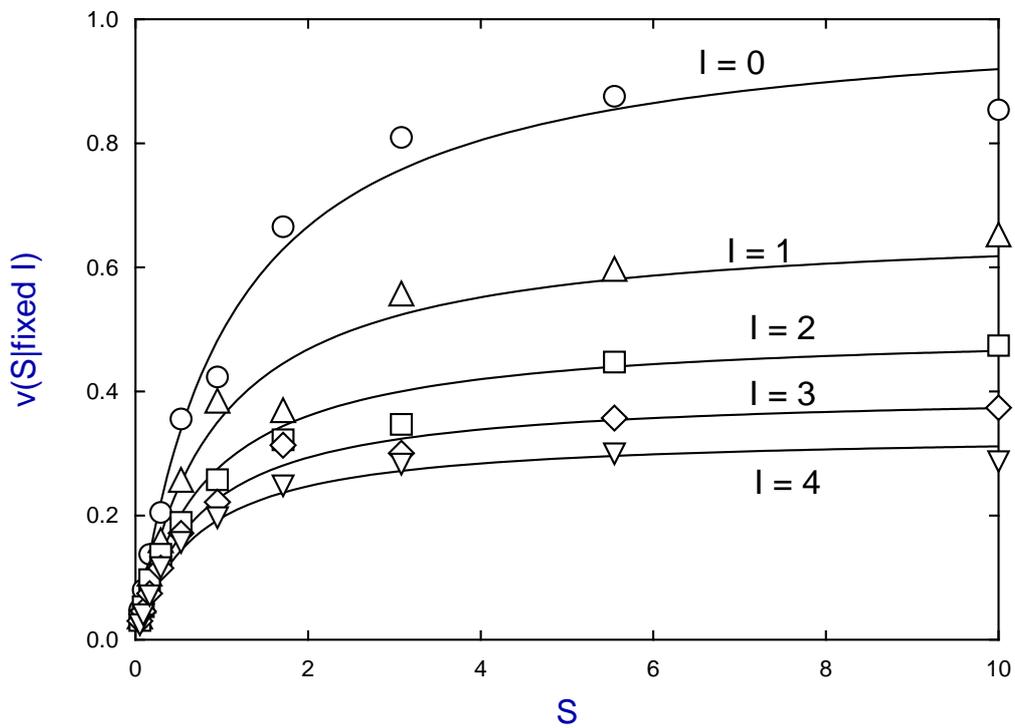
```
begin{limits}
0.0   0.5   5.0
0.0   1.5   5.0
0.0   5.0   9.0
0.0   1.0   5.0
end{limits}
```

and provides the options to display the best-fit surface and the sections across the best-fit surface for functions of S at fixed I, which simply requires accumulating the sections into your project archive in order to plot retrospectively or create a library file referencing the data and best-fit curves contributing to the five cross–sections.

1

Best-Fit Surface and Contours for: v = f(S,I)

0.91974

$\uparrow$

v = f(S, I)

0.022408

0.0

0.05

I

S

4.0   10.0
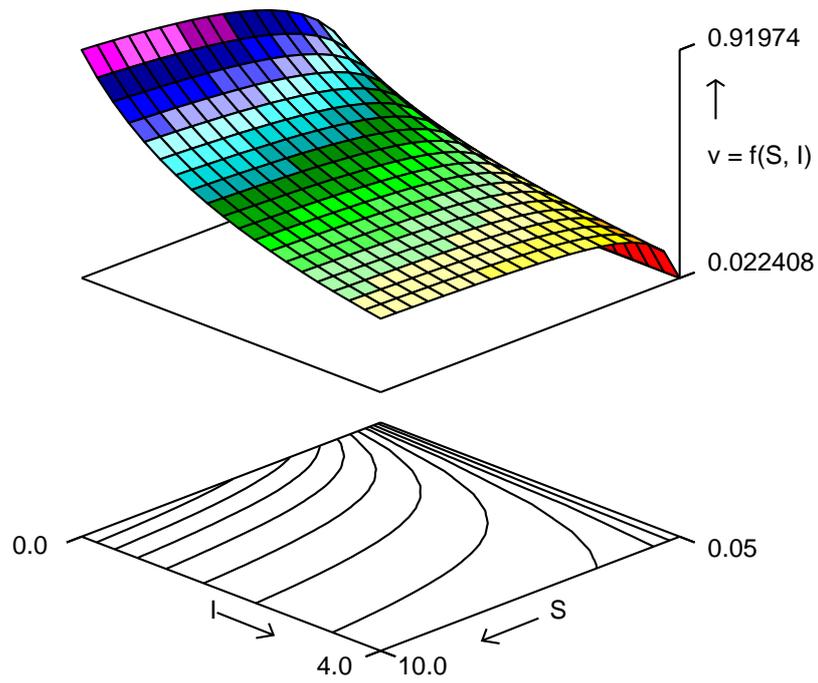
Cross-Sections of the Surface v = f(S,I)

I = 0

I = 1

I = 2

I = 3

I = 4

v(S|fixed I)

S

The second technique: Plotting the sign of residuals on contours

It is not always possible to create a good display of the data plotted with the corresponding best–fit surface because some data may lie below the surface and some above no matter how the surface is rotated.

To illustrate this phenomenon and present an alternative way to view the residuals consider the second example using the following statistical model.

The bivariate normal pdf: $p_1 = \mu_x, p_2 = \sigma_x, p_3 = \mu_y, p_4 = \sigma_y, p_5 = \rho$

$$\frac{p_6}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} + p_7$$

This can be fitted as follows.

1. Open SimFiT program **qnfit** then select to fit a function of two variables.

2. Read in the test file `bivariate.tf1`, which contains data generated using program **makdat** followed by adding random error using program **rannum**.

3. Fit in the EXPERT mode which reads the starting estimates from the data file.

4. Observe the goodness of fit results, view the residuals, and then plot the best–fit surface, as shown next.

# Best-Fit Scaled Bivariate Normal Distribution
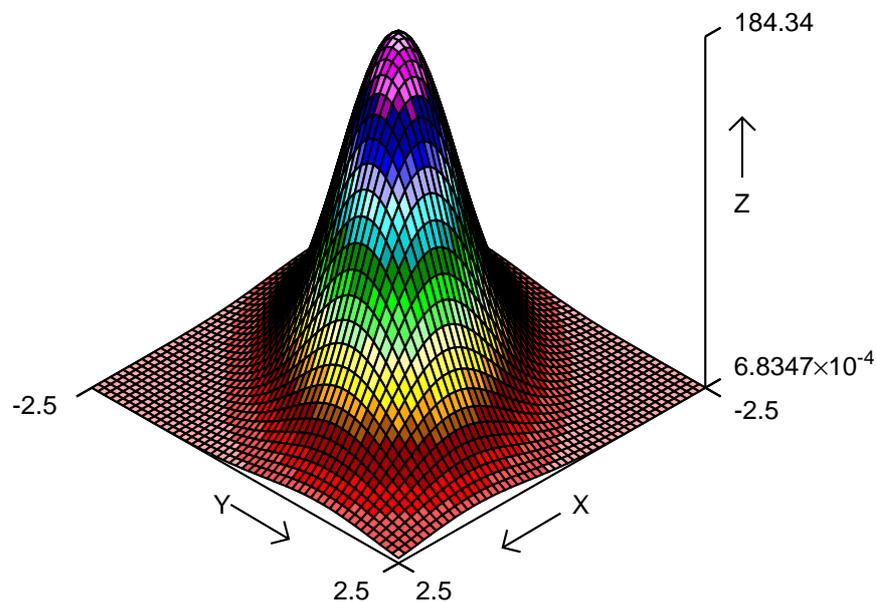
It is obvious from the table of residuals that there is a scatter of data points both above the best-fit surface (positive residuals) and below the best-fit surface (negative residuals) and a technique to display this scatter is presented next.

Note that each time a request is made to view the best-fit surface after fitting a function of two variables to a data set, program **qnfit** archives two temporary files containing the *x*, *y* coordinates of the data partitioned as follows.

- `qnfit_positive_residuals.txt` (contains coordinates for the positive residuals).

- `qnfit_negative_resdiduals.txt` (contains coordinates for the negative residuals).

These two temporary files are written to your `...\Documents\Simfit\res` folder, and it only remains to explain how these are to be overlayed as shown in the following contour diagram.
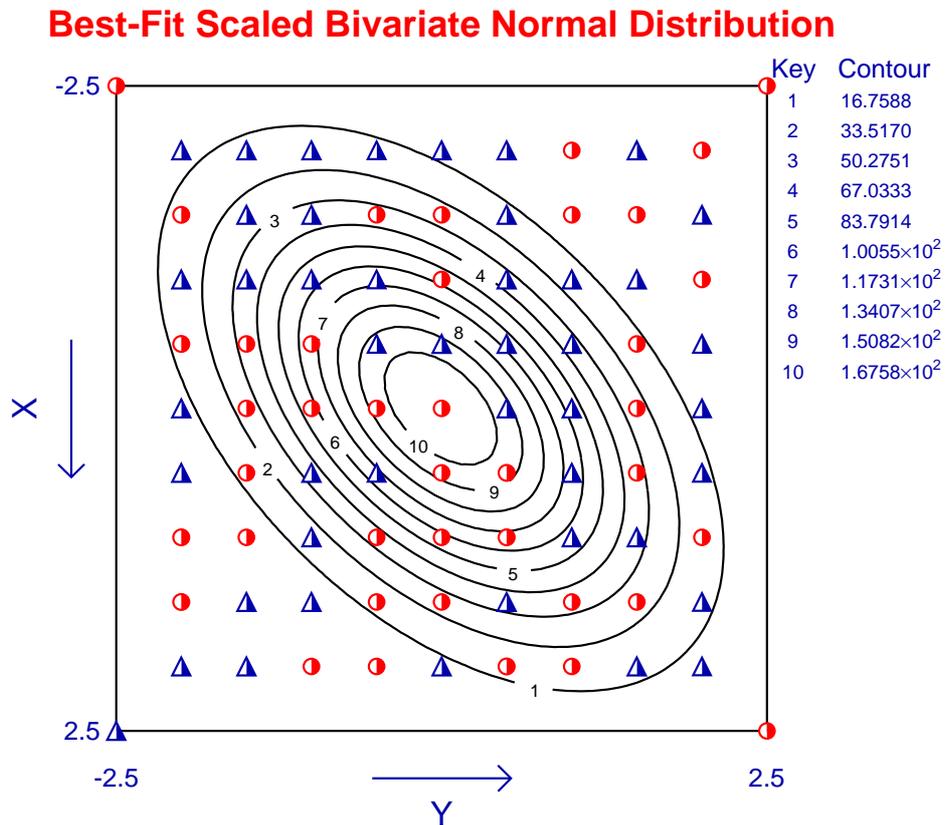
Each time the choice to display a best-fit surface is made there are several alternatives. In particular, it will be seen that on the main interface to the displays is an option to add overlays to the contour diagrams. This allows users to add arbitrary coordinate files to be plotted as lines or symbols in chosen colors as overlays on the contours.

For instance, in the above example the positive residuals are displayed as half-filled circles in red, while the negative residuals are shown as blue half-filled triangles. It is clear from the scatter of color over the evenly spaced grid of data points that there is no evidence for systematic lack of fit, and this can be corroborated from inspection of the best-fit parameters and tables of results together with using several other alternative methods for statistical analysis.