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

Cubic splines can be used to create standard curves for calibration when the data are nonlinear and more complex than a simple quadratic or cubic. Under some circumstances it is also possible to generate approximate 95% confidence limits for plotting and predicting x from y, or y from x.

From the main SIMFIT menu choose [A/Z], open program **calcurve** and scan the default test file `calcurve.tf2` containing the following data.

x	y
0.50	0.26885
0.50	0.30026
0.50	0.27048
0.50	0.28205
0.50	0.27368
1.00	0.81924
1.00	0.79264
1.00	0.80419
1.00	0.86795
1.00	0.80573
1.50	1.5215
1.50	1.6423
1.50	1.6953
1.50	1.7242
1.50	1.4159
2.00	2.5742
2.00	2.3165
2.00	2.5198
2.00	2.6637
2.00	2.5150
4.00	6.7101
4.00	6.4626
4.00	6.4935
4.00	5.9591
4.00	6.3087
6.00	7.9744
6.00	7.9506
6.00	8.2786
6.00	8.4860
6.00	8.1895
8.00	9.6610
8.00	9.6185
8.00	9.5581
8.00	9.3566
8.00	8.0443
10.0	10.713
10.0	10.619
10.0	9.7203
10.0	9.7127
10.0	9.4258

1. Column 1 contains the fixed independent variable x
2. Column 2 contains the observations y
3. The absence of a third column of weighting factors s is equivalent to a column of $s = 1$ indicating unweighted regression, but if accurate estimates for s, the sample standard deviations of replicates, are available they can be added as a third column.

Example 1

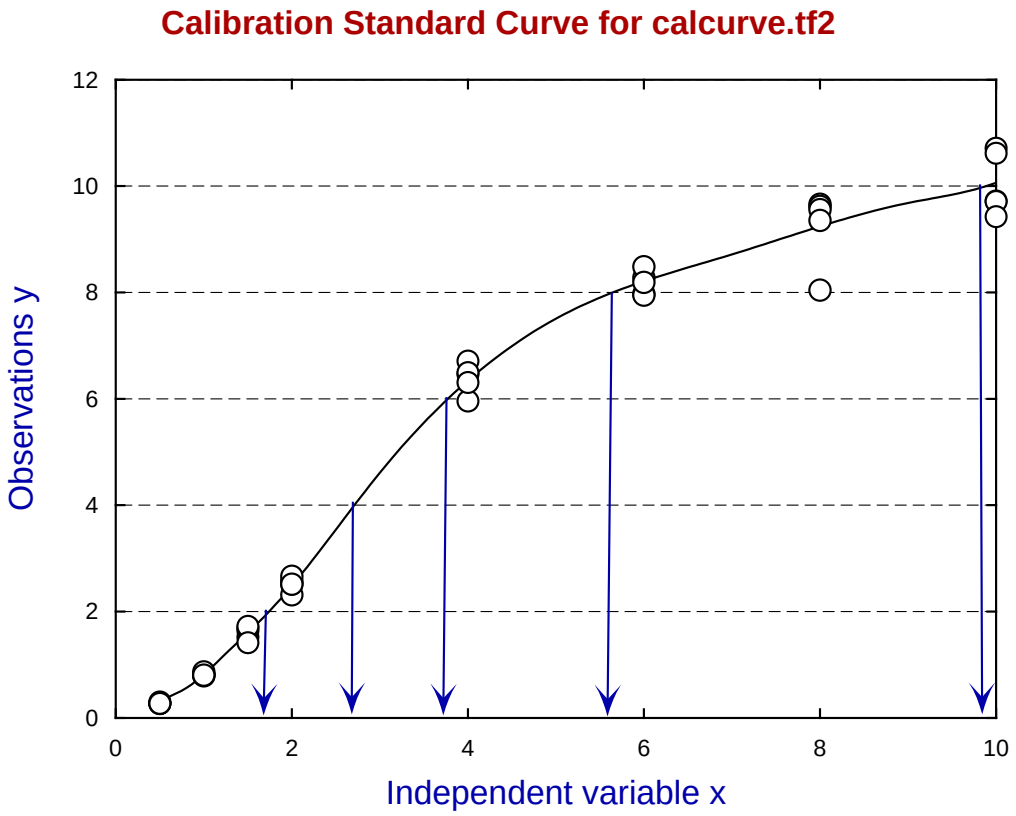
Choosing the option to create a standard calibration curve using the **calcurve** defaults then using the option to predict x given y equal to 2, 4, 6, 8, 10 leads to the following table.

y-measured	x-predicted
2.00	1.7410
4.00	2.7105
6.00	3.7716
8.00	5.6437
10.0	9.8853

To appreciate how this table results, consider the following graph where it will be clear that, for any given value of $y = y_0$, program **calcurve** solves the equation

$$y_0 - f(x) = 0$$

involving the best-fit spline function $f(x)$, using numerical techniques to locate the x-values located at the arrow heads.



Example 2

The two most often required changes to the default configurations in program **calcurve** are

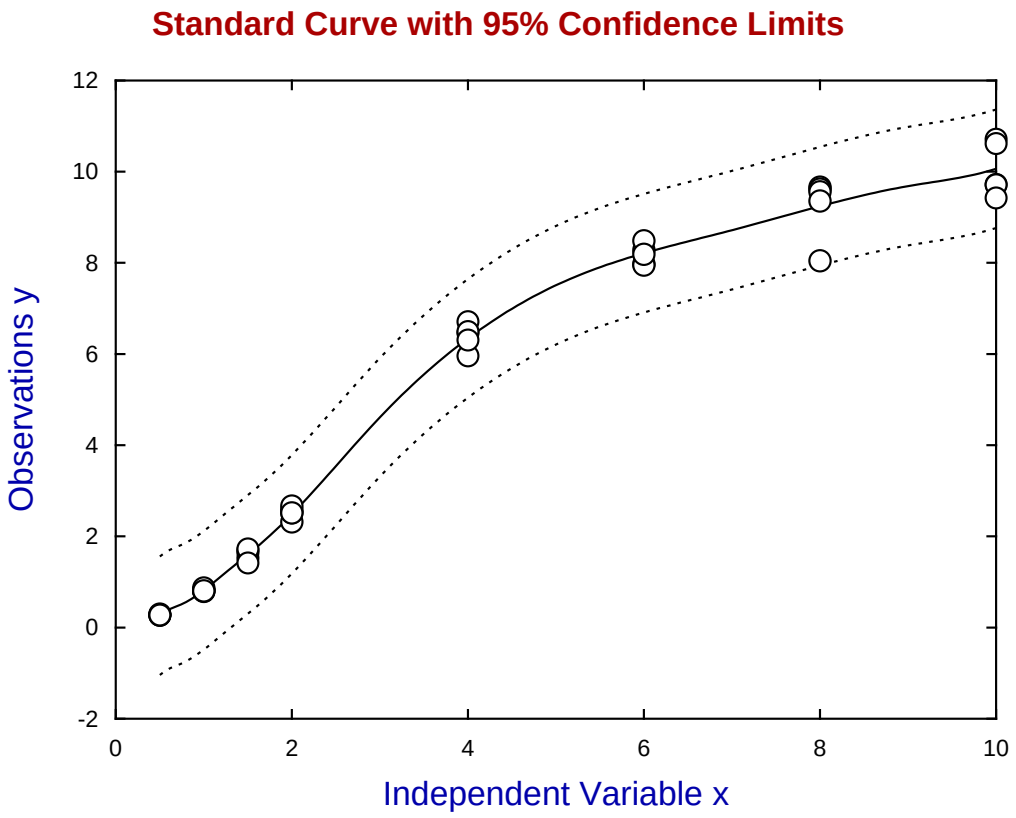
1. swap from using data-files/clipboard for standard curve data to typing in data interactively; and
2. estimate 95% confidence limits for plotting or prediction.

Typing in data interactively instead of using files or the clipboard is merely a convenience choice, but estimating confidence limits is more controversial. Unlike the situation when an appropriate deterministic equation is fitted and the best-fit parameters can be used for this purpose, the cubic spline is an empirical model which can be made to fit arbitrarily closely to data by choosing the knot placements and tension settings. Hence the weighted sum of squares at the solution point cannot be used to create a variance estimate if the default choice of unweighted fitting with cross-validation splines is employed.

Nevertheless **calcurve** will attempt to generate confidence limits in this default case leading to the following outcome when attempting to predict x from y.

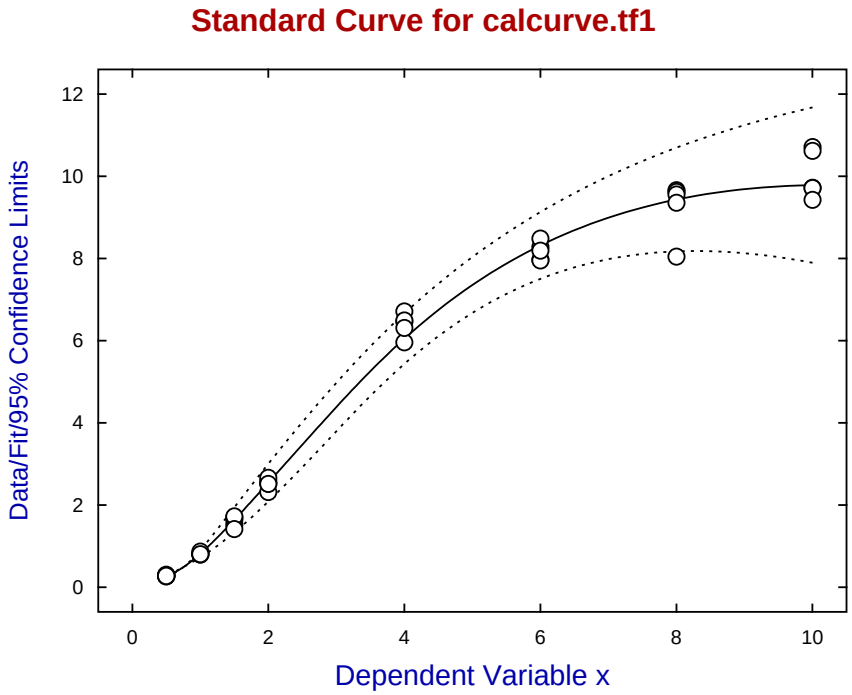
y-measured	x-predicted	Lower95%cl	Upper95%cl	
2.00	1.7410	0.78716	2.4583	
4.00	2.7105	1.9565	3.5280	
6.00	3.7716	2.8267	5.2894	
8.00	5.6437	3.8358	9.6669	
10.0	9.8853	5.5667	10.000	** Discard

The intersection of horizontal lines displayed in the previous graph is used to find the intersections with the confidence limit curves and, as will be obvious from the next graph, this fails to locate the intersection of $y = 10$ with the lower confidence limit curve. A fuller discussion of this subject will be presented later.

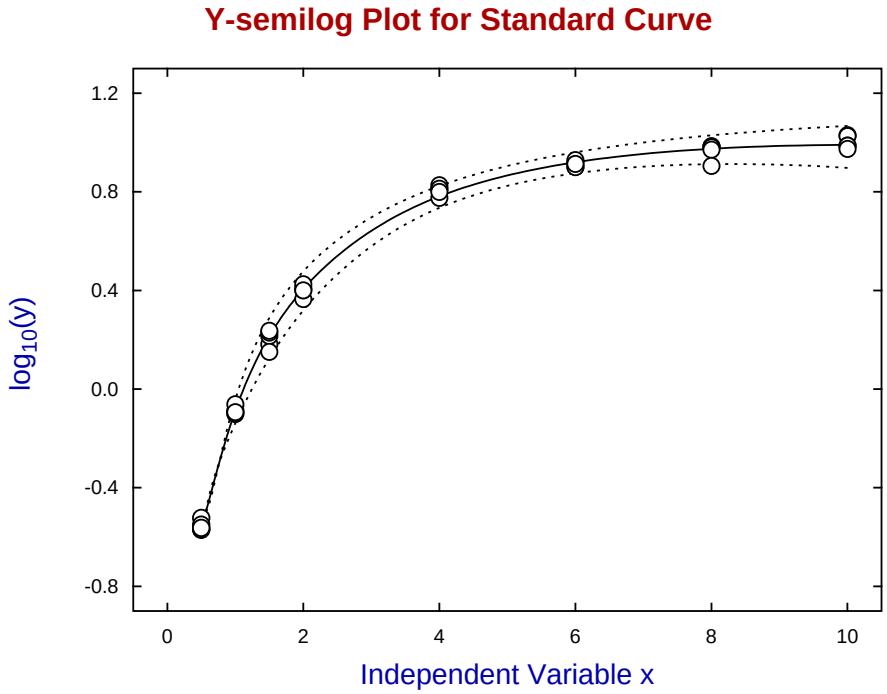


Example 3

Configuration options can be added to the data set to over-ride defaults, as illustrated by fitting `calcurve.tf1` with standard deviations from replicates added as a third column.

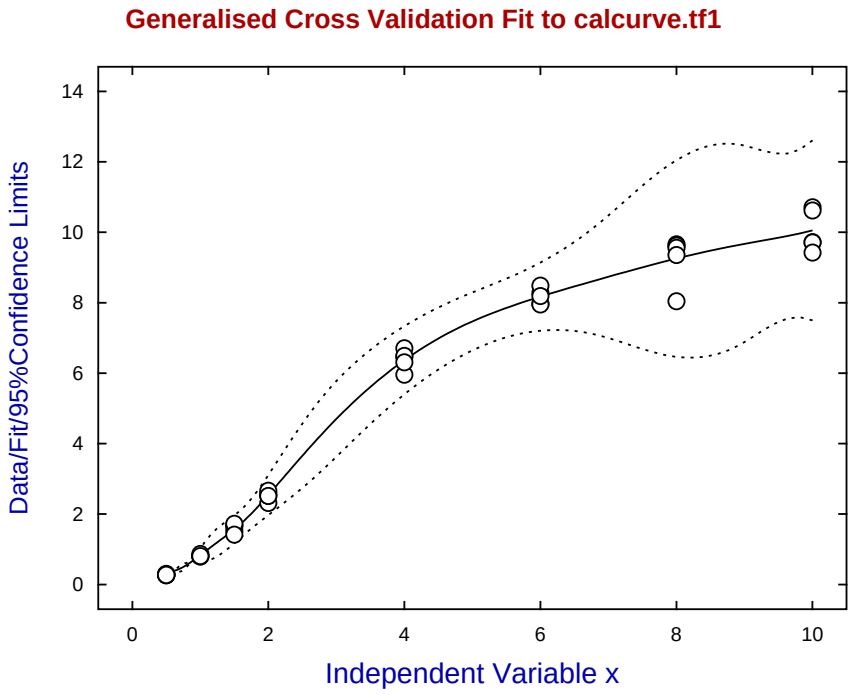


The next graph illustrates how the actual situation of constant relative error appears like constant variance under *y*-semilog transformation.



The best way to use program **calcurve** is to experiment with the configuration options until a satisfactory standard curve is obtained. The details of such a configuration can then be added to the data file to temporarily over-ride the defaults. The great advantage of this expert mode is that data sets for creating standard curves will always give rise to the same standard curve. For instance, fitting `calcurve.tf1` in expert mode reads 8 integers followed by a floating point number that informs **calcurve** to use weights $w = 1/s^2$, $\log x$ instead of x as independent variable, and weighted least squares splines instead of cross-validation splines.

To illustrate this point, the next graph shows the fit obtained for `calcurve.tf1` using the expert mode parameters except that a cross-validation spline curve was used. The previous graphs show how weighted least squares fitting smooths out noisy data, while the next graph illustrates how using too many knots, as with cross validation, can lead to excessive undulations in response to local variations in noise. A full explanation of such configuration issues will now be given.



Expert mode configuration options

It will be seen on inspection that `calcurve.tf1` has an integer after the data indicating the number of additional lines of text appended to the data. The first line appended has the following eight integers followed by a floating point number.

2, 2, 2, 1, 2, 3, 2, 2, 5.0

Program **calcurve** will recognize that this second line after the end of the data is a list of instructions indicating that the default configuration options are temporarily to be replaced by these new options as long as the data file is the current data file.

When a new data file is opened, the last set of active settings derived from the defaults will be re-instated unless the new data set has another set of such temporary re-configuration instructions.

There follows a full description of all the **calcurve** options

Configuration options

- **Option 1**

  1. Use a file (or equivalently the clipboard) for all data input
  2. Use a file (or equivalently the clipboard) for observations but type in prediction values
  3. Type in both observations and prediction values from the terminal

  The value of 3 is not valid in expert mode which must use file/clipboard input mode for observations. If sub-option 3 is selected, then data typed in will be written to a temporary file that can be saved from the SIMFIT user results folder for re-use. It is always best to use a file for input of observations, especially if the number of observations is large, and this is also true if a large number of prediction values are required

- **Option 2**

  1. Use the independent variable as supplied
  2. Transform the independent variable internally into the logarithm for fitting

  The second sub-option can only be used if the independent variable supplied is positive, and is valuable for calibration data that approaches a horizontal asymptote in order to prevent an undulating standard curve. All transformation occurs internally, and all communication with the user is in the coordinates supplied for the original observations.
  If either of these sub-options are changed interactively the standard curve will have to be re-fitted.

- **Option 3**

  1. Sparse knots ($K = N/12$)
  2. Medium knots ($K = N/6$ but $K \geq 1$)
  3. Dense knots ($K = N/3$ but $K \geq 2$)
  4. Solid knots ($K = N - 1$ but cross validation)

  Four knots are always placed at the first and last distinct points, but the number of equally spaced interior knots K depends on the number of distinct data points N. Solid knots places a knot between each distinct data point then uses generalized cross validation to estimate a smoothing factor. As explained, the type of knots used will strongly effect the shape of the standard curve. For instance, sparse knots will usually fit a simple cubic and may give rise to turning points, while solid knots may give too much undulation.
  If either of these sub-options are changed interactively the standard curve will have to be re-fitted.

- **Option 4**

  This option is linked to option 6.

  1. Weights given by $w = 1/s^2$ where s values are supplied
  2. Weights given by $w = 1/(cv\%|y|)^2$ where $cv\%$ is the percentage coefficient of variation assumed
  3. Weights given by $w = 1$ i.e., unweighted regression

  It is only sensible to supply weights s as sample standard deviations if the sample sizes are sufficiently large to be meaningful (i.e. ≥ 5), but using unweighted regression assumes constant variance which is usually incorrect. If constant relative error rather than constant variance can be established, it is best to input the estimated percentage coefficient of variation and use $w = 1/(cv\%|y|)^2$ for smoother weighting.
  If either of these sub-options are changed interactively the standard curve will have to be re-fitted.

- **Option 5**

  1. Plot y as a function of x
  2. Also plot approximate 95% confidence limits using the settings of options 4 and 6.

  Confidence limits should only be plotted if the settings of options 4 and 6 are sensible.

- **Option 6**

  This option is linked to option 4.

  1. No confidence limits on prediction
  2. Slack confidence limits on prediction (using $4s$)
  3. Medium confidence limits on prediction (using $3s$)
  4. Tight confidence limits on prediction (using $2s$)

  Confidence limits are calculated as the intersection of either y_0 or x_0 as appropriate with the upper and lower confidence limit envelopes described for option 4. They are very approximate and must be interpreted with restraint. When sub-option 3 of option 4 is selected it is not sensible to use $WSSW/NDOF$ as a variance estimated as the degrees of freedom depend on the knots and, using too many knots with no replicates can lead to $WSSQ = 0$. In this case the number of standard deviations s used to space the confidence limits above and below the best-fit standard curve are calculated using the percentage coefficient of variation assumed and the average absolute value of the observations.
  If either of these sub-options are changed interactively the standard curve will have to be re-fitted.

- **Option 7**

  Reserved for future use.

- **Option 8**

  Reserved for future use.

- **Option 9**

  If sub-options 2 or 3 of option 4 are selected the estimated percentage coefficient of variation must be provided.
  If this option is changed interactively the standard curve will have to be re-fitted.

Some examples for Expert mode settings

So, for instance, the expert line to restore defaults is

2, 1, 4, 3, 1, 1, 2, 2, 7.5

while changing to input of observations and prediction values from the terminal would require

3, 1, 4, 3, 1, 1, 2, 2, 7.5

Also changing from cross-validation splines to medium density weighted least squares would need

3, 1, 2, 3, 1, 1, 2, 2, 7.5

while a further change to invoke weighting $w = 1/s^2$, medium prediction confidence limits and addition of 95% confidence limits for graphs and predictions would result from

3, 1, 2, 1, 2, 3, 2, 2, 7.5