



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

Cubic splines that have been fitted to a data set can be used to calculate parameters that characterize a data profile, or to act as a standard curve for calibration, e.g. LD50 determination. Once a spline has been fitted to a data set it generates a set of knots and coefficients that can be save to a file in order to re-use the spline retrospectively.

From the main SIMFIT menu choose [A/Z] and open program **spline** which has a default data set in test file `compare.tf1`, and a corresponding default spline file `spline.tf2`. Choosing to calculate using the current spline file offers the following choices.

1. X-Range: select $A = X_{min}$, $B = X_{max}$
2. Evaluate: spline and derivatives
3. Evaluate: Area-Length-Curvature
4. Calibrate: predict X given Y
5. Plot: spline
6. Plot: derivative
7. View: knots and coefficients

Choosing to view the knots and coefficients displays the following values.

Spline knots

0.25000
0.25000
0.25000
0.25000
0.67860
0.94662
1.32050
1.84200
5.00000
5.00000
5.00000
5.00000

Spline coefficients

0.01531456
0.05658395
0.20421635
0.71563087
0.82384236
0.63789726
0.50485453
0.46681088

There are always four knots at the lowest X value, and four at the highest X value, then between these are the locations of the interior knots. The coefficients are multipliers for the B splines. It should be noted that, after every fit of a piecewise spline to a data set has concluded, the splines from this fit become the current spline and can be saved to a file formatted like `spline.tf2`.

Evaluate a spline and derivatives

Supplying $x = 1$ for this calculation leads to these results.

X-value	Spline	First derivative	Second derivative	Third derivative
1	0.6678	0.7695	-3.216	7.129

Evaluate Area-Length-Curvature

Choosing this option leads to the following results for the range $A = X_{min}$, $B = X_{max}$.

A	B	Area	L = Arc-length	Integral $ \kappa ds$	In degrees
0.25	5	2.720	5.053	1.738	99.61

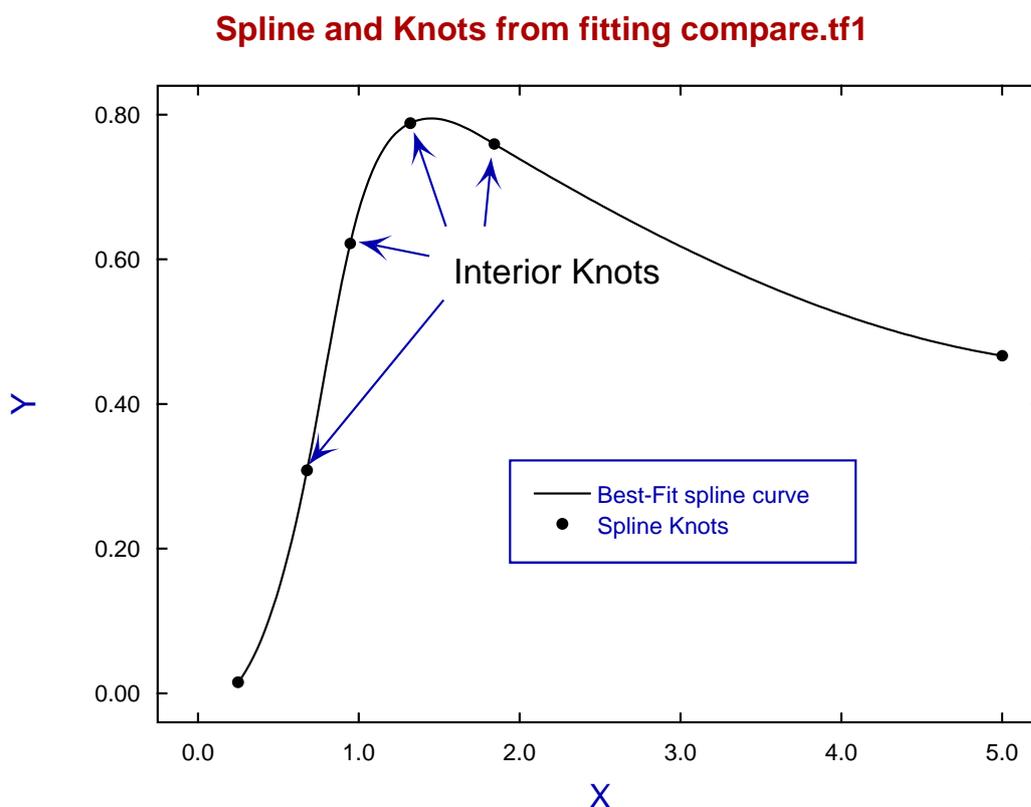
The area under the curve estimate is very valuable for calculating AUC in those pharmacology studies where exponentials cannot be fitted by program **exfit** to quantify drug availability.

The arc length estimate can be used in studies where response along a path is more meaningful than simply over the range of dependent or independent variable traversed.

The estimated absolute curvature integral is given in radial measure and degrees, and summarizes the amount of oscillation in the best-fit curve.

Plot a spline

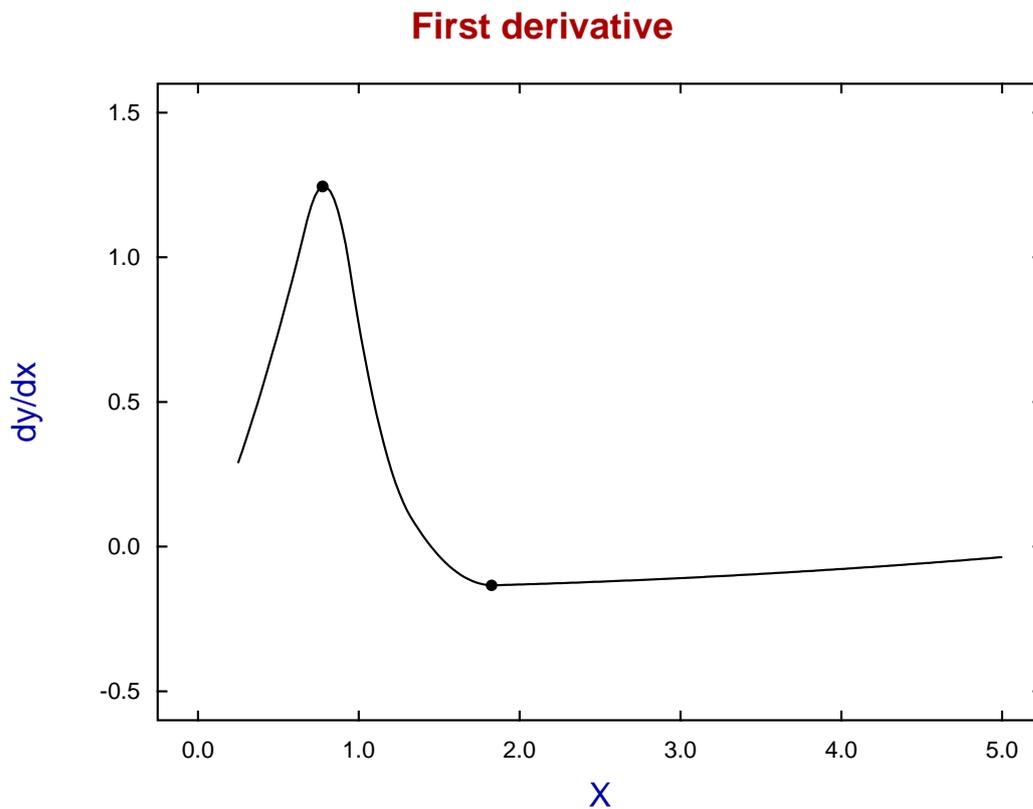
The best-fit spline curve and the position of the interior knots can be plotted as follows.



Plot the derivative

This estimates extreme values of dy/dx and corresponding x values as in the next table and graph.

Minimum independent variable	0.2500	Maximum independent variable	5.000
Minimum selected value (A)	0.2500	Maximum selected value (B)	5.000
Minimum first derivative	-0.1335	in range(A, B) at x -value	1.825
Maximum first derivative	1.245	in range(A, B) at x -value	0.7751



This procedure can be used to estimate the maximum and minimum growth rates with data sets are too noisy, or that cannot be analyzed for some other reason by program **gcfit**.

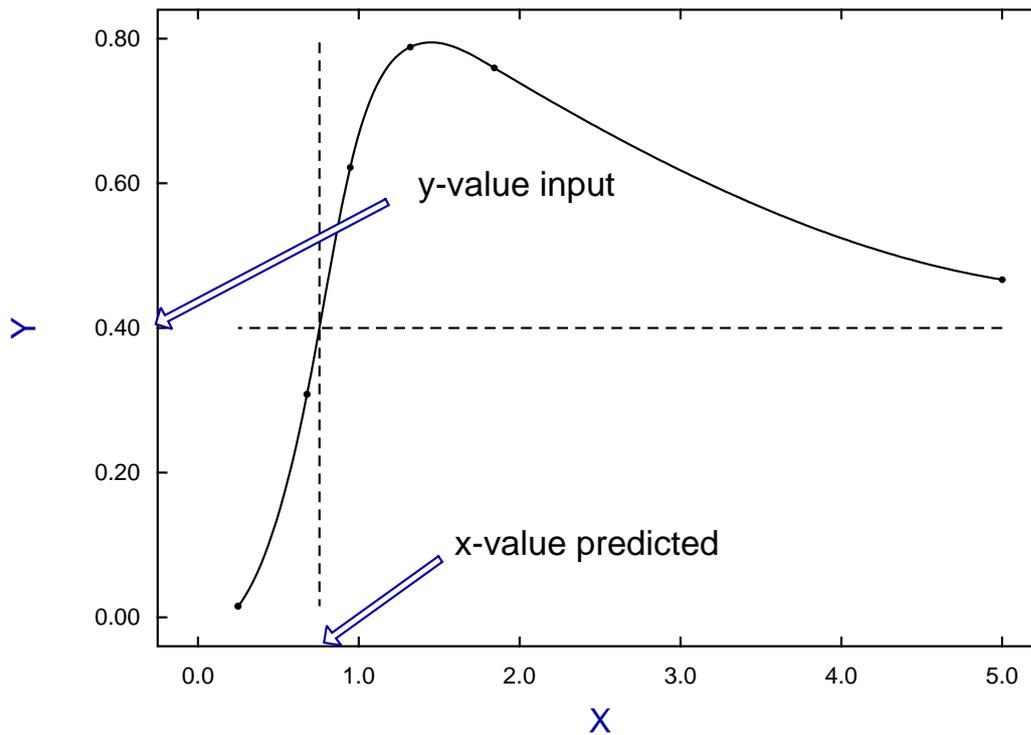
Calibrate by predicting x given y

Spline reference curves are useful with noisy data, or in situations where bias would result from using an ill-fitting deterministic equation, e.g. LD50 determination.

From requesting a value for x predicted given a value of $y = 0.4$ we get the next result and graph illustrating the values input and predicted as dotted lines.

For $Y = 0.4$, X predicted = 0.75545

Spline with Knots and X predicted from Y



X-Range: select $A = X_{min}$, $B = X_{max}$

The ability to vary the range of investigation has many applications because, if the values for A and B are such that A is larger than the lowest point where the spline is defined, or B is less than the highest point where the spline is defined, then calculations for areas, arc length, curvature, derivatives, and calibration will refer to the new restricted range of values $A \leq x \leq B$.

For instance, from the previous graph it will be clear that, because of the turning point, some attempts to predict x_0 given y_0 will fail because of ambiguity leading to two intersections of the line $y = y_0$ with the best-fit spline curve. This is because, for a spline defined as $f(x)$, SIMFIT calculates $y_A = f(A) - y_0$ and $y_B = f(B) - y_0$ in order use these as starting estimates for solving the equation

$$f(x_0) - y_0 = 0$$

numerically, and iteration cannot commence when $y_A y_B > 0$.

To illustrate how this problem can be circumvented we calculate the two solutions for $y = 0.6$.

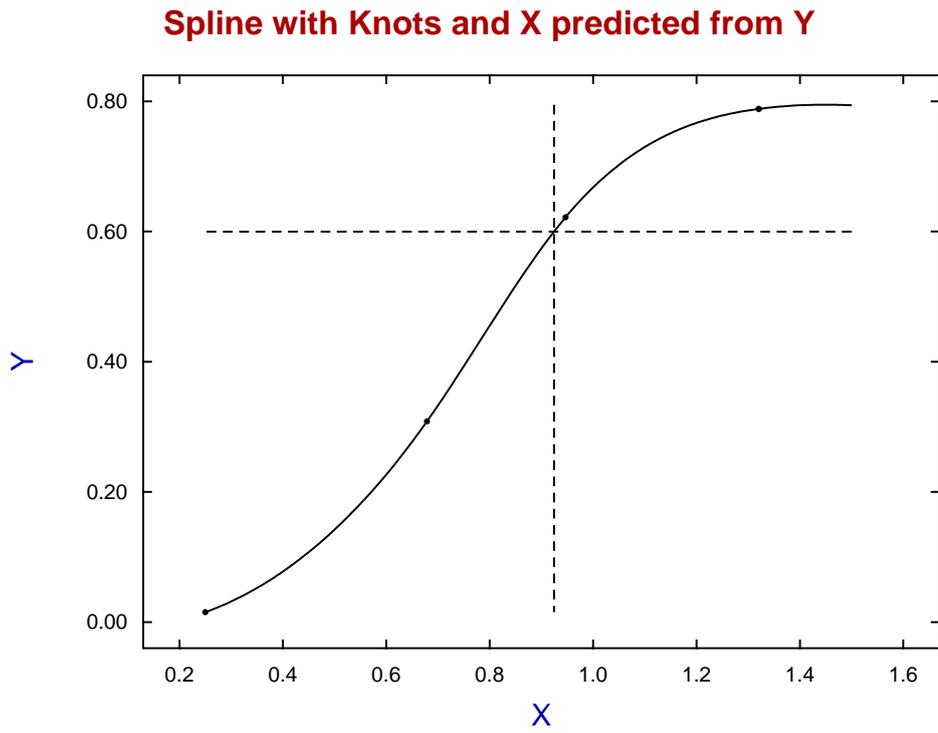
First we select the range $A = 0.25$, $B = 1.5$ which gives the lower solution.

For $Y = 0.6$, X predicted = 0.92438

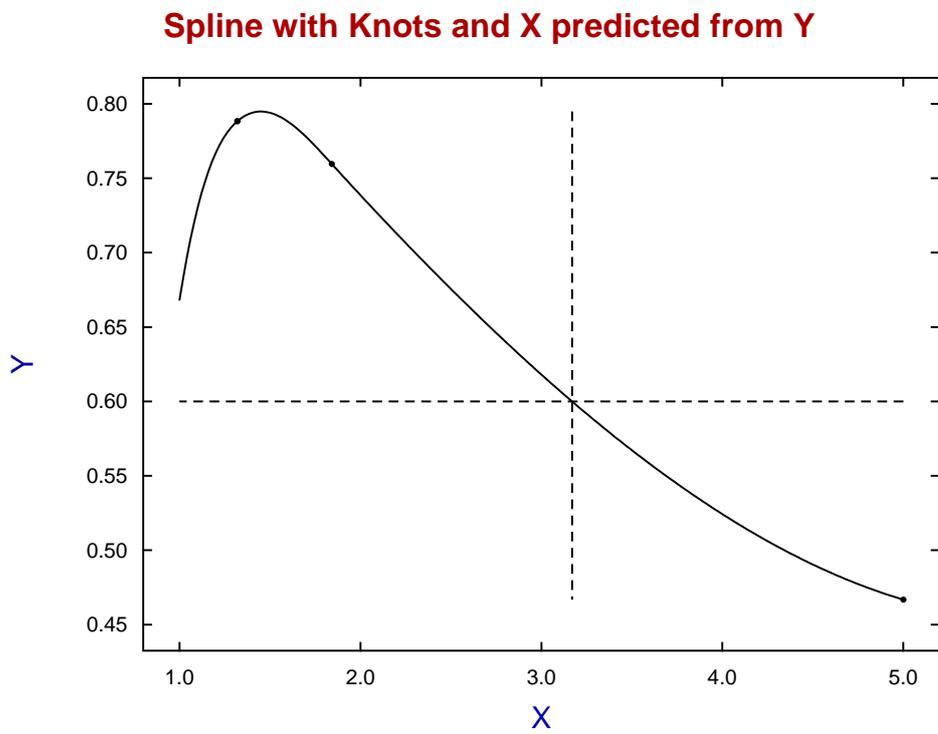
Then we re-define $A = 1.0$, $B = 5$ which gives the upper solution.

For $Y = 0.6$, X predicted = 3.1696

This is the graph for the low range solution



while this is the graph for the high range solution.



Theory

As splines are defined by knots and coefficients rather than equations, special techniques are required to re-use best fit spline functions. SIMFIT provides procedures to save spline parameters to a file so that they can be re-used to restore previously fitted spline functions. This is particularly useful when a best-fit spline is to be re-used as reference function or standard curve, as in calibration.

Input a spline file such as `spline.tf1` into program **spline** to appreciate how to re-use a best fit spline stored from **spline**, **calcurve**, or **compare**, to estimate derivatives, areas, curvatures and arc lengths.

SIMFIT spline files of length $k \geq 12$, such as `spline.tf1`, have $(k + 4)/2$ knots, then $(k - 4)/2$ coefficients as follows.

- There must be at least 8 nondecreasing knots
- The first 4 of these knots must all be equal to the lowest x value
- The next $(k - 12)/2$ must be the non-decreasing interior knots
- The next 4 of these knots must all be equal to the highest x value
- Then there must be $(k - 4)/2$ spline coefficients c_i

With \bar{n} spline intervals (i.e. one greater than the number of interior knots), $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are knots corresponding to the lowest x value, $\lambda_5, \lambda_6, \dots, \lambda_{\bar{n}+3}$ are interior knots, while $\lambda_{\bar{n}+4}, \lambda_{\bar{n}+5}, \lambda_{\bar{n}+6}, \lambda_{\bar{n}+7}$ correspond to the largest x value. Then the best-fit spline $f(x)$ is

$$f(x) = \sum_{i=1}^{\bar{n}+3} c_i B_i(x).$$

where the c_i are the spline coefficients, and the $B_i(x)$ are normalized B-splines of degree 3 defined on the knots $\lambda_i, \lambda_{i+1}, \dots, \lambda_{i+4}$.

When the knots and coefficients are defined in this way, the function $y = f(x)$ can be used as a model-free best fit curve to obtain point estimates for the derivatives y', y'', y''' , as well as the area AUC , arc length L , or total absolute curvature κ over a range $\alpha \leq x \leq \beta$, defined as

$$\begin{aligned} AUC &= \int_{\alpha}^{\beta} y \, dx \\ L &= \int_{\alpha}^{\beta} \sqrt{1 + y'^2} \, dx \\ \kappa &= \int_0^L \frac{|y''|}{(1 + y'^2)^{\frac{3}{2}}} \, dl \\ &= \int_{\alpha}^{\beta} \frac{|y''|}{1 + y'^2} \, dx \end{aligned}$$

which are valuable parameters to use when comparing data sets. For instance, the arc length L provides a valuable measure of the length of the fitted curve, while the total absolute curvature indicates the total angle turned by the tangent to the curve and indicates the amount of oscillatory behavior.