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

Data smoothing is used to fit arbitrary curves to observations when there is no sense in fitting deterministic models to estimate meaningful parameters, because all that is required is a smoothed representation of the data to identify trends, or perform calibration for instance.

SIMFIT provides the following procedures.

1. **Smoothing time series.**

   Given a succession of points measured at a constant interval of space, time, etc. then various filters, such as Hanning or T5432H, can be applied to generate a smoothed representation. This technique, available from program **simstat**, is widely used in time series analysis.

2. **Autocorrelation.**

   Program **simstat** can analyze an arbitrary succession of values for lags and autocorrelations.

3. **Moving average analysis.**

   Program **simstat** can also fit ARIMA models in order to predict future performance.

4. **Fitting a polynomial.**

   This can be done for a succession of polynomials of degree one to six with statistics given to find the highest degree that can be justified. The resulting best-fit polynomial fitted by program **polnom** can be used to estimate deviations from linearity, or to act as as a calibration curve.

5. **Fitting piecewise cubic splines.**

   Several types of splines are available.

   - Program **compare** can be used to fit splines with automatically calculated knots to two profiles in order to compare the similarity and differences between two sets of measurements over a similar range of independent variable. For instance, for nonparametric comparison of two growth profiles.
   - Program **calcurve** fits splines with knots fixed by users in order to generate a calibration curve that can be used for inverse prediction. That is, given measurements of a variable y as a function of some variable x, to predict x given y.
   - Program **spline** can fit splines with knots fixed by users, calculated automatically, or chosen by cross-validation in order to visualize trends.

Items 1, 2, and 3 in the above list require data in the form of a vector V, that is, a succession of n data points

$$V = (x_1, x_2, x_3, \ldots, x_n).$$

However items 4 and 5 require a n by 3 matrix M with independent variable x, observations y and standard error estimates se

$$M = \begin{pmatrix} x_1 & y_1 & se_1 \\ x_2 & y_2 & se_2 \\ \ldots & \ldots & \ldots \\ x_n & y_n & se_n \end{pmatrix}$$

where the third column would be standard deviations determined from replicates for weighting $w = 1/se^2$, or more usually set to 1, or even omitted altogether for unweighted fitting.