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

Often observations are made that can be modeled by fitting a differential equation. This can be done using SimFIT program **deqsol** which is very useful for simulating and fitting, but program **qnfit** is recommended for fitting a single differential equations as there are many more options available for altering starting estimates interactively and recording goodness of fit criteria.

It is important to stress several issues that are important when fitting a differential equation.

1. **The Data**

   It is very important to only attempt fitting when data are very extensive and accurate and the model has been simulated to observe the behavior over the data range. Otherwise parameter estimates will be very unreliable.

2. **The initial condition**

   When advancing the solution of a differential equation from the chosen starting point it is clear that the whole trajectory is governed by this starting point. This raises several points.

   - If the initial condition is known with certainty then it is preferable to fit without the initial condition being estimated. However, an incorrectly specified fixed initial condition can result in a tail–wagging–the–dog situation leading to biased parameters estimates.
   - If the initial condition is estimated then the fitting can be dominated by the inial condition parameter having a different effect on convergence than the other parameters.
   - If the initial condition is to be estimated then this will probably be be known within a fairly narrow range, so the parameter limits on the initial condition could be fairly close. However, it should be pointed out that fixing parameter limits too close can prevent the convergence techniques from operating optimally.

3. **The starting estimates**

   Constrained nonlinear regression works best when the internal parameters are of order unity and, if possible, are parameterized in the formulation of the model so that sensitivity of the model to all the parameters to be estimated is similar. Where possible SimFIT scales parameters before optimization commences using the starting estimates. Evidently this cannot be done for very small starting estimates.

4. **The parameter limits**

   Models for differential equation must have n parameters involved in the model plus a further parameter that is the value of the initial condition. That is, for a differential equation of the form

   $$dy/dx = f(x, y, \Theta), y(0) = y_0$$

   involving parameters $\Theta = \theta_1, \theta_2, \dots, \theta_{n+1}$, then parameter $\theta_{n+1} = y_0$. The parameter starting estimates and limits should be such that parameters to be varied should have a small range of variation (but not too small) and parameters to be fixed can be indicated by setting the limits equal to the starting estimates. In this way the inititial condition can be fixed.

5. **The simulation**

   A differential equation can easily be simulated using program **makdat**, or for more comprehensive options **deqsol** where, to initialize the simulation, the range of integration can be appended to the end of the model file.

6. **The methods**

   To allow for stiff equations a Jacobian can be added to the model file, but should also compared with simulation without an explicit Jacobian to ensure that the Jacobian has been coded correctly.

Case 1: Irreversible substrate depletion

This fits the irreversible Michaelis-Menten substrate depletion scheme

$$
\begin{aligned}
\frac{dS}{dt} &= -\frac{V_{max}S}{K_m + S} = f \\
\frac{df}{dS} &= -\frac{V_{max}K_m}{(K_m + S)^2} = J
\end{aligned}
$$

which is also provided in the SIMFIT model library both as a formal integrated equation as well as a built in differential equation model.

The data file deqn_data.tf1 ends with the following limits

```
begin{limits}
0.5 0.75 1.5
0.5 1.25 2.5
0.5 0.8  1.5
end{limits}
```

so **qnfit** can be used in the EXPERT mode. The corresponding model file deqn_model.tf1 is

```
%
  model: irreversible Michaelis-Menten substrate depletion curve
 differential equation: f(1) = dy(1)/dx
                             = -p(2)*y(1)/[p(1) + y(1)]
              jacobian: j(1) = df(1)/dy(1)
                             = -p(1)p(2)/[p(1) + y(1)]^2
    initial condition: y0(1) = p(3)
 Note: the last parameter must be y0(1) in a differential equation
       y(1) = S, p(1) = Km, p(2) = Vmax, y0(1) = S(0)
 %
 1 equation
 differential equation
 3 parameters
 %
 begin{expression}
 D = p(1) + y(1)
 f(1) = -p(2)y(1)/D
 end{expression}
 %
 begin{expression}
 j(1) = -p(1)p(2)/D^2
 end{expression}
  %
 begin{limits} ... low-limits, starting estimates, upper-limits
 0 1 3
 0 1 3
 0 1 3
 end{limits}
 begin{range} ... number of points, x_start and x_end for integration
 121
 0
 10
 end{range}
```

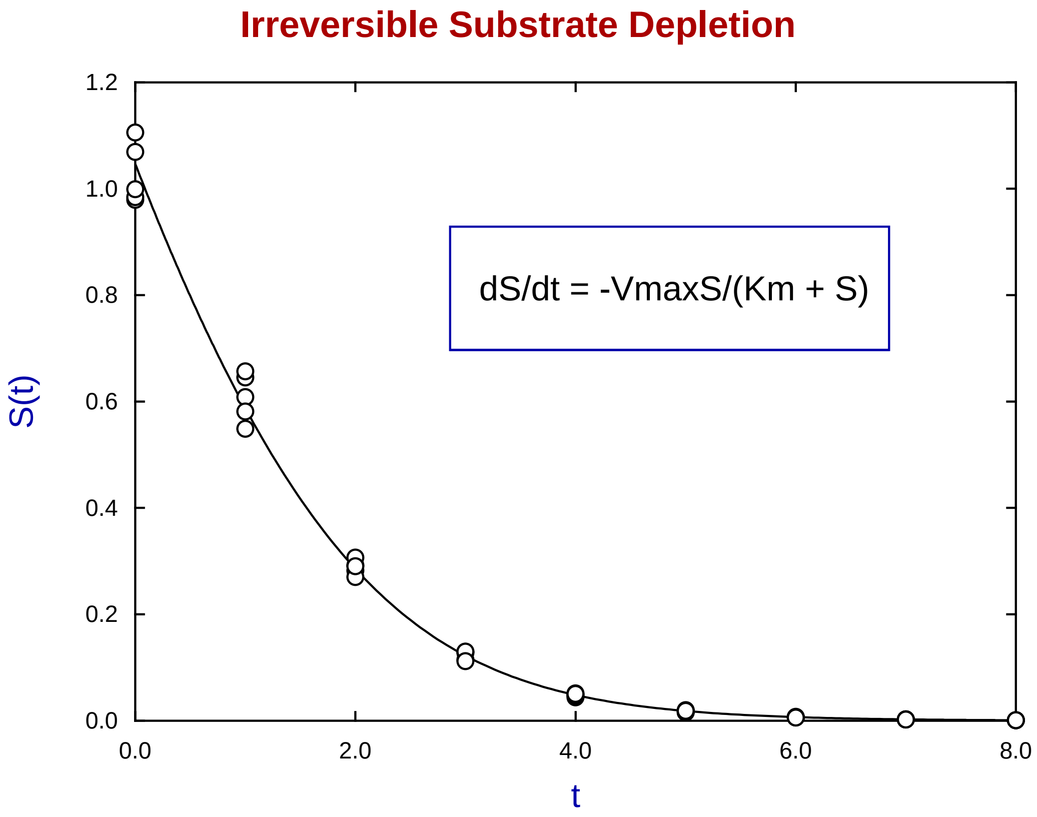
To appreciate how to supply your own data eventually it is suggested you read in the data file `deqn_data.tf1`, then the model file `qnfit_model.tf1`, and study the structure of the model file to make sure you understand the format.

SIMFIT program **usermod** can be used to create your own model if the model you require is not provided by the pre-compiled library of models for simulating and fitting.

A number of details need mentioning.

1. The `begin{limits} ... end{limits}` section appended to the data file is only provided to allow program **qnfit** to be used in the EXPERT mode. Program **qnfit** is opened, fitting a differential equation is selected, and `deqn_data.tf1` is read in using the [Demo] button on the file–open dialogue. Typically these limits would be edited interactively if a good fit cannot be obtained. After reading in the data, the option to open an ASCII text model file would be selected and `deqn_model.tf1` should be read in using the [Demo] button on the file–open dialogue.
2. The `begin{limits} ... end{limits}` section appended to the model file is only needed if fitting or simulation is to be undertaken using program **deqsol**.
3. The `begin{range} ... end{range}` section appended to the model file is only needed to fix the number of points and range of simulation if simulation is to be undertaken using program **deqsol**.

Here is the best–fit curve obtained using program **qnfit**.



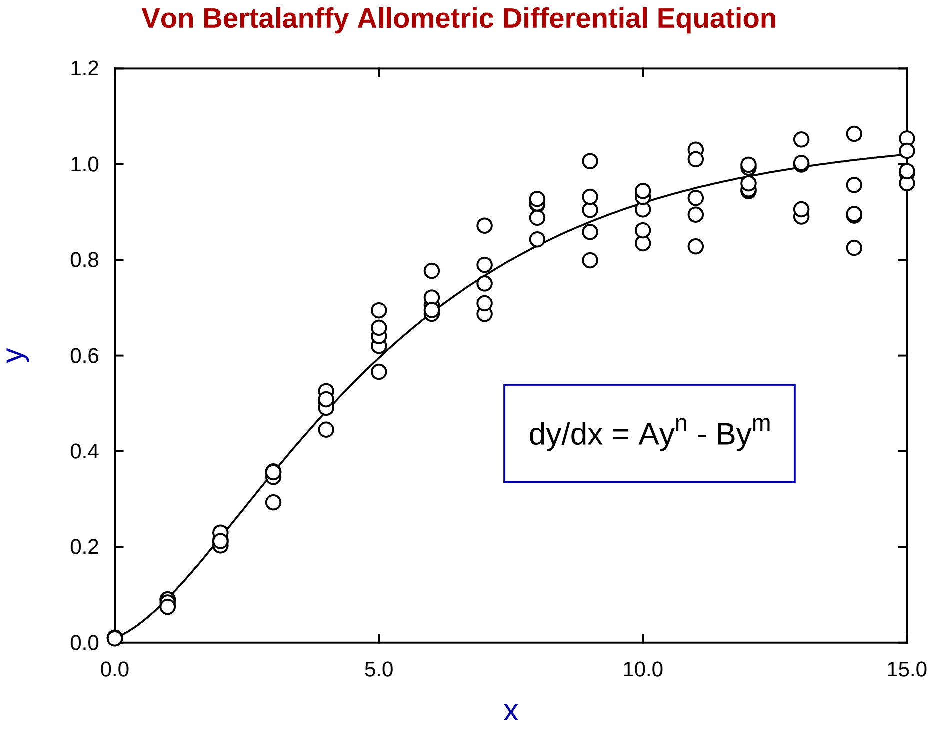
Case 2: The Von Bertalanffy Allometric Differential Equation

The appropriate default data file is `deqn_data.tf2` and the model file is `deqn_model.tf2` for

$$dy/dx = Ay^n - By^m$$

which can be integrated for some n and m to give the family of growth equations fitted by program **gcfit**.

```
%
 f(1) = p(1)*y(1)^p(2) - p(3)*y(1)^p(4), y(0) = p(5)
 j(1) = p(1)*p(2)*y(1)^(p(2) - 1.0) - p(3)*p(4) y(1)^(p(4) - 1.0)
 %
 1 equation
 differential equation
 5 parameters
 %
 begin{expression}
 f(1) = p(1)y(1)^p(2) - p(3)y(1)^p(4)
 end{expression}
 %
 begin{expression}
 A = p(2) - 1.0
 B = p(4) - 1.0
 j(1) = p(1)p(2)y(1)^A - p(3)p(4)y(1)^B
 end{expression}
 %
 begin{limits}
 0  1.0       3
 0  0.6666667 3
 0  1.0       3
 0  1.0       3
 0  0.01      1
 end{limits}
```



Case 3: Modified von Bertalanffy equation

The model dealt with in Case 1 is very easy to fit, but the model described in Case 2 is much more difficult to fit as it involves exponents. This means that the model fitting must be constrained to avoid raising negative numbers to non-integer powers, and to restrict the range within which exponents can vary.

However, the example to be described in Case 3 is chosen because it is an extreme example of a simple looking model that proves very difficult to fit, requiring both high quality data and careful choice of parameters and limits.

The motivation for this model should be explained first. The von Bertalanffy differential equation had a fanciful origin in proposing to balance the opposing effects of anabolism and catabolism by an appeal to the contrast between effects due to surface area and those due to volume. As it is an autonomous differential equation the derivative is defined unambiguously by the value of its argument and hence the integral can have no turning points. Now, as many growth situations have a phase of decline following a rise to maximum size, then clearly such a monotonic growth equation will not give a satisfactory fit, and parameter estimates will not be meaningful. To remedy this situation we can introduce a process of time–dependent deterioration affecting the surface area term more significantly than the volume term as follows, where C is a further parameter to be estimated.

$$dy/dx = \exp(-Cx)Ay^n - By^m$$
$$J = \exp(-Cx)Any^{n-1} - Bmy^{m-1}$$

The data set `deqn_data.tf3` was generated using program **makdat**, then pseudo–random error was added in triplicate using program **adderr** with constant relative error. The following model file (`deqn_data.tf3`) was composed using program **usermod**, noting than the last parameter is for the initial condition.

```
%
  model: modified von Bertalanffy growth model
 differential equation: f(1) = dy(1)/dx
                             = exp(-p(5)x)*p(1)*y(1)^p(2) - p(3)*y(1)^p(4)
              jacobian: j(1) = df(1)/dy(1)
                             = exp(-p(5)x)*p(1)*p(2)*y(1)^(p(2) - 1.0)
                               -p(3)*p(4)*y(1)^(p(4) - 1.0)
    initial condition: y0(1) = p(5)
 %
 1 equation
 differential equation
 6 parameters
 %
 begin{expression}
 C = p(1)*exp(-p(5)x)
 f(1) = C*y(1)^p(2) - p(3)y(1)^p(4)
 end{expression}
 %
 begin{expression}
 A = p(2) - 1.0
 B = p(4) - 1.0
 j(1) = C*p(2)y(1)^A - p(3)p(4)y(1)^B
 end{expression}
%
```

Note that this model has no protection against underflow or overflow, but the SimFIT library versions of this and the previous model are protected against such numerical instability.

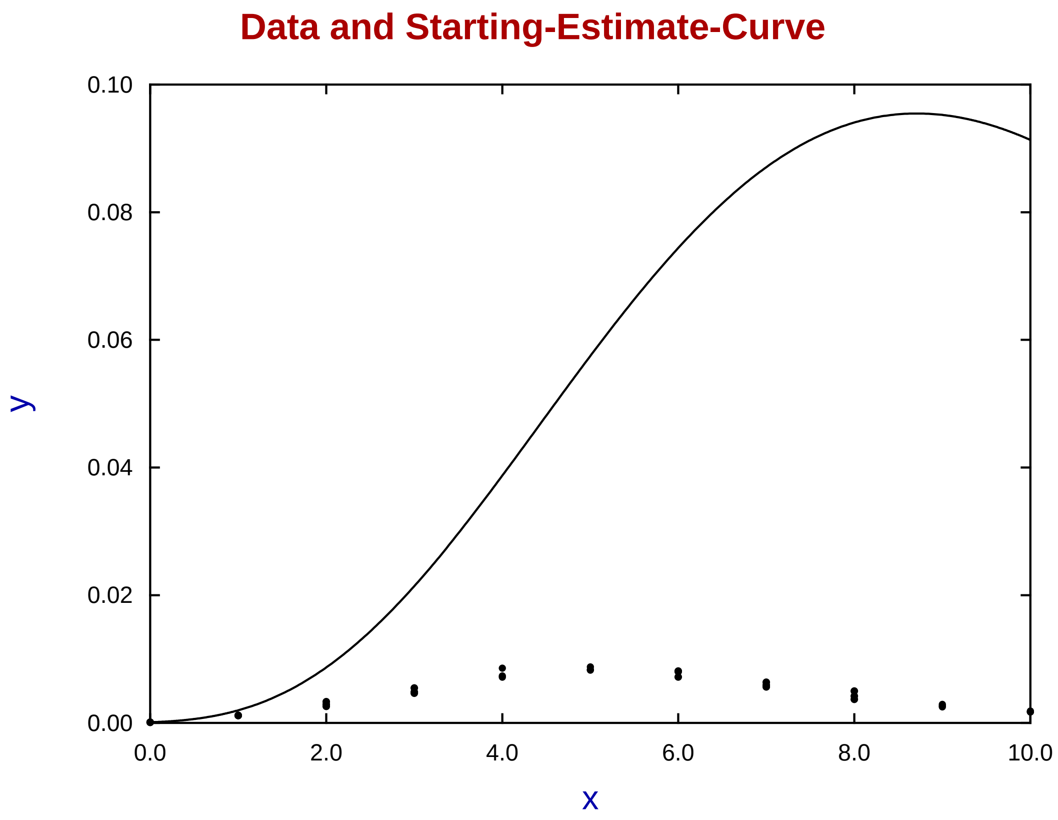
The parameter starting estimates and limits appended to the data set in `deqn_data.tf3` are shown next

```
begin{limits}
0.5     1.1    2.0
0.75    0.8    0.85
0.5     0.9    2.0
0.85    0.9    0.95
0.001   0.05    0.2
1.0e-5  1.0e-4 1.0e-3
end{limits}
```

which should be compared with the parameters used to generate the data.

```
  p(1) = 1.00000E+00
  p(2) = 8.00000E-01
  p(3) = 1.00000E+00
  p(4) = 9.00000E-01
  p(5) = 1.00000E-01
  p(6) = 1.00000E-04
```

After reading the data set `deqn_data.tf3` into program **qnfit** followed by the assumed model contained in the file `deqn_model.tf3`, the EXPERT mode was selected and, before commencing to fit, the starting-estimate best–fit curve was overlayed on the data as shown next.



This illustrates how, with differential equations containing parameters as exponents, a small variation in parameters can lead to large and somewhat unexpected changes in the integral, and at first sight it would seem improbable that this model could be made to fit the data.

However, **qnfit** found a good solution with well–defined parameters and excellent goodness of fit criteria as shown next.

Results from curve-fit number 1

Number of data points = 33

Number of parameters = 6 (0 currently fixed)

Degrees of freedom = 27

WSSQ before entry = 1482310.0

IFAIL from LBFGSB = 0

WSSQ from fitting = 31.383

P($\chi^2 \geq$ WSSQ) = 0.2557

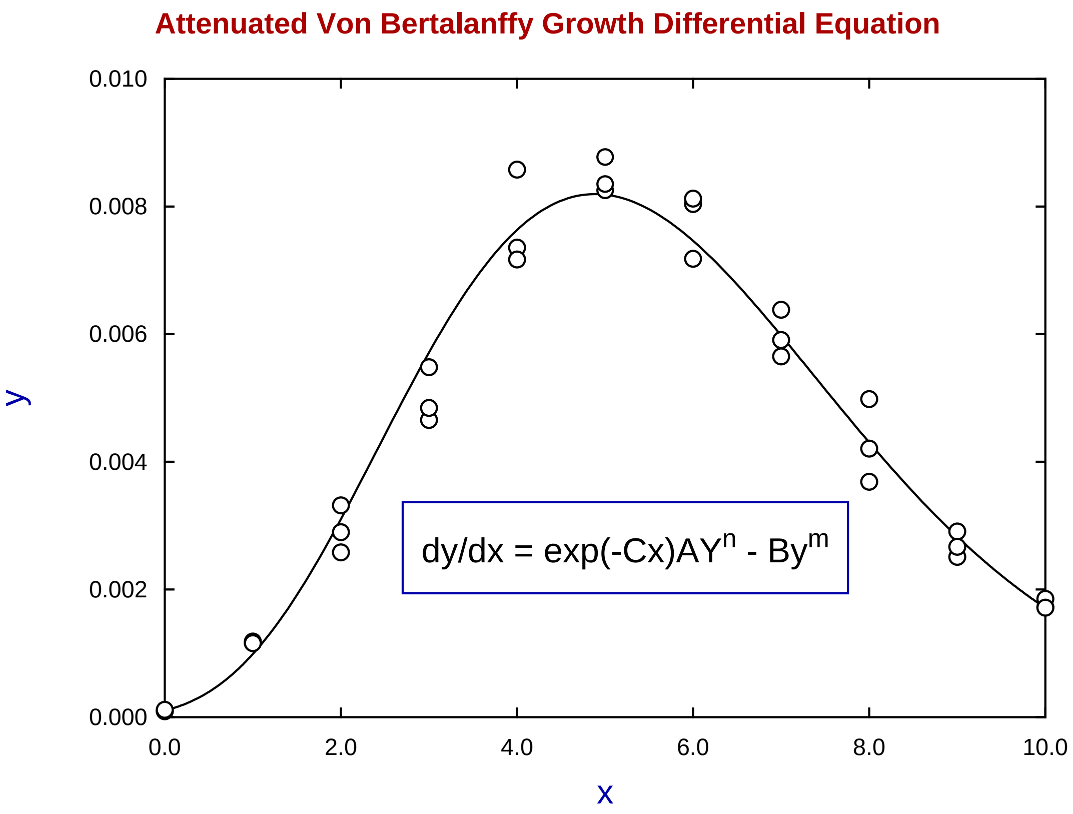
Time taken to fit = 2.5994 (secs cpu time)

Best-fit parameters for curve-fit 1 using LBFGSB/DVODE

Number	Low-Limit	High-Limit	Value	Std.Error	Lower95%cl	Upper95%cl	*p*
1	0.5	2.0	1.68365	0.00188	1.67979	1.68750	0.0000
2	0.75	0.85	0.84996	0.00625	0.83713	0.86279	0.0000 *
3	0.5	2.0	1.78207	0.03956	1.70091	1.86324	0.0000
4	0.85	0.95	0.92920	0.00661	0.91563	0.94278	0.0000
5	0.001	0.2	0.06624	0.00376	0.05853	0.07395	0.0000
6	1.000E-05	0.001	0.00010	4.04070E-06	9.64899E-05	0.00011	0.0000

For 50,90,95,99% confidence limits using [parameter value +/- t(α/2)*std.err.]

t(0.25) = 0.684, t(0.05) = 1.703, t(0.025) = 2.052, t(0.005) = 2.771


Attenuated Von Bertalanffy Growth Differential Equation
y
x
dy/dx = exp(-Cx)AY^n - By^m