



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

Nonlinear growth, decay, and survival models are fitted to data in order to estimate parameters that can be used to compare the effects of treatments and/or different groups. The parameters that are usually estimated are the initial and final sizes and rates of change and meaningful numbers such as the half life and maximum rates of change.

Example 1: Growth data

From the main SIMFIT menu choose [A/Z], open program **gcfi**t, select the option to fit growth curves then browse the default test file **gcfi**t.tf2 containing the following data.

Time	Size	Standard Error
0.0000	0.090501	0.0057406
0.0000	0.085148	0.0057406
0.0000	0.096621	0.0057406
1.0000	0.20400	0.0069302
1.0000	0.21300	0.0069302
1.0000	0.21763	0.0069302
2.0000	0.42858	0.017410
2.0000	0.45530	0.017410
2.0000	0.42261	0.017410
3.0000	0.71832	0.039573
3.0000	0.64283	0.039573
3.0000	0.70118	0.039573
4.0000	0.84408	0.041097
4.0000	0.76262	0.041097
4.0000	0.79382	0.041097
5.0000	0.91559	0.027506
5.0000	0.86060	0.027506
5.0000	0.88937	0.027506
6.0000	0.98545	0.016132
6.0000	0.95853	0.016132
6.0000	0.98738	0.016132
7.0000	1.0552	0.063158
7.0000	0.94115	0.063158
7.0000	1.0452	0.063158
8.0000	1.0433	0.040631
8.0000	0.96285	0.040631
8.0000	1.0130	0.040631
9.0000	0.99185	0.047010
9.0000	1.0452	0.047010
9.0000	1.0856	0.047010
10.000	1.0226	0.018664
10.000	0.98858	0.018664
10.000	0.99220	0.018664

Column 1 contains the time values t which must be nonnegative and in nondecreasing order.

Column 2 contains the size estimates $S(t)$ which must be nonnegative.

Column 3 contains the sample standard deviations for the triplicates to use for weighting, but this column can be set to one or omitted if weighting is not required.

Program **gcfi**t can fit sequences of nonlinear growth, decay, or survival models giving statistics for goodness of fit and model discrimination but, before proceeding further, the definition of $S(t)$ must be explained.

If the data are for longitudinal measurements on the same individual or subjects they will be correlated so that fitting nonlinear models by weighted least squares will generate biased fits instead of maximum likelihood fits. One way to circumvent this is to fit flexible models such as polynomials or splines by techniques that attempt to estimate the autocorrelation. However it is only possible to estimate approximate correlations and polynomials cannot capture the shape of actual growth data or be used to estimate meaningful parameters to characterize growth profiles.

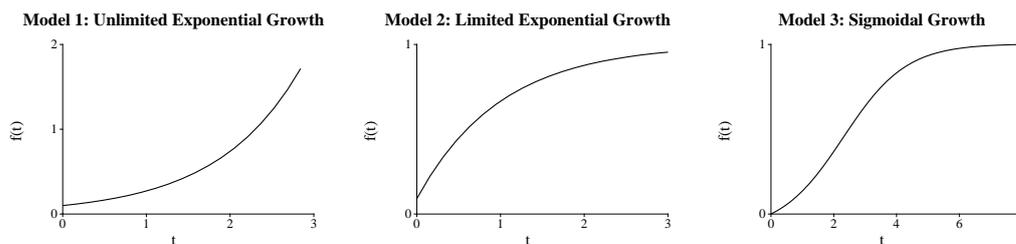
Ideally, **gcfi**t should be used where observed $S(t)$ values are obtained in such a way as to make successive observations independent, e.g. sampling without replacement to estimate growth of bacterial colonies. Users will have to balance the usefulness of growth curve models with possible bias induced by fitting a deterministic model against the model-free data smoothing approach.

First of all note that most simple growth curve models are special cases of differential equations such as the Von Bertalanffy allometric equation

$$dS/dt = AS^\alpha - BS^\beta$$

which can be simulated and fitted using program **deqsol**, and it usual to explore the type of model required by fitting the first three models provided by program **gcfi**t in a preliminary investigation. When a model has been selected there will be no further need to fit sequences of models.

Three typical growth curve shapes are shown in the next figure.



- **Model 1**

This is exponential growth $S_1(t)$ which is only encountered in the early phase of development.

$$S_1(t) = A_1 \exp(k_1 t)$$

- **Model 2**

This is limited exponential growth $S_2(t)$, concave down to an asymptote fitted by the monomolecular model

$$S_2(t) = A_2 [1 - \exp(-k_2 t)]$$

- **Model 3**

This is the logistic equation $S_3(t)$ which can fit sigmoidal profiles.

$$S_3(t) = \frac{A_3}{1 + B \exp(-k_3 t)}$$

Proceeding to fit these three models sequentially leads to the following conclusions and results table for model three, then a plot of data and all three best fit curves.

Model	$WSSQ/NDOF$	$P(\chi^2 \geq W)$	$P(R \leq r)$	$N > 10\%$	$N > 40\%$	$Av.r\%$	Verdict
1	152	0.000	0.000	29	17	40.03	Very bad
2	18.1	0.000	0.075	20	0	12.05	Very poor
3	1.32	0.113	0.500	0	0	3.83	Incredible

In this table $WSSQ/NDOF$ is the weighted sum of squares divided by degrees of freedom and $P(\chi^2 \geq W)$ is the significance level for this parameter in a chi-square test. $P(R \leq r)$ is the probability of runs less than or equal to the number observed given the number of positive and negative residuals, while $N > 10\%$ and $N > 40\%$ are the number of absolute residuals exceeding the stated percentage, and $Av.r\%$ is the average absolute residual. The conclusions in the last column are based on these results along with several other goodness of fit measures, and clearly model 3 is the preferred model with the estimated parameters shown next.

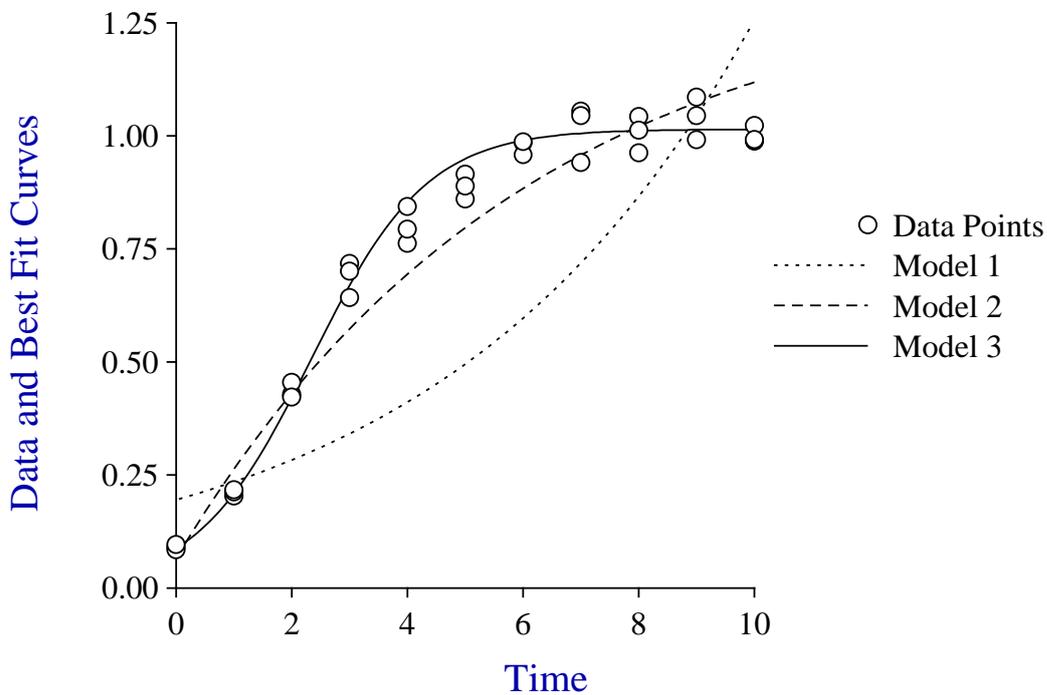
Results for model 3

Parameter	Value	Std.error	Lower95%cl	Upper95%cl	p
A	0.99891	0.0078551	0.9828 7	1.0150	0.0000
B	9.8901	0.33300	9.2100	10.570	0.0000
k	0.98814	0.026785	0.93344	1.0428	0.0000
$t_{1/2}$	2.3190	0.045070	2.2270	2.4111	0.0000

Parameter correlation matrix

1			
-0.0167	1		
-0.4388	0.7192	1	

Fitting Alternative Growth Models



Example 2: Decay data

It is often useful to fit growth models to data that are decreasing as a function of time instead of increasing. For instance, the Gompertz model growth data in the next table are contained in test file `gcfit.tf5` while the same data are arranged into decay form in test file `gcfit.tf6`.

Growth data		Decay data	
t	$S(t)$	t	$S(t)$
0.0000	0.0048	0.0000	97.3685
1.1110	0.3044	1.1110	96.4062
2.2220	3.4696	2.2220	82.1162
3.3330	14.5225	3.3330	74.1991
4.4440	40.8277	4.4440	52.6928
5.5560	52.6928	5.5560	40.8277
6.6670	74.1991	6.6670	14.5225
7.7780	82.1162	7.7780	3.4696
8.8890	96.4062	8.8890	0.3044
10.0000	97.3685	10.0000	0.0048

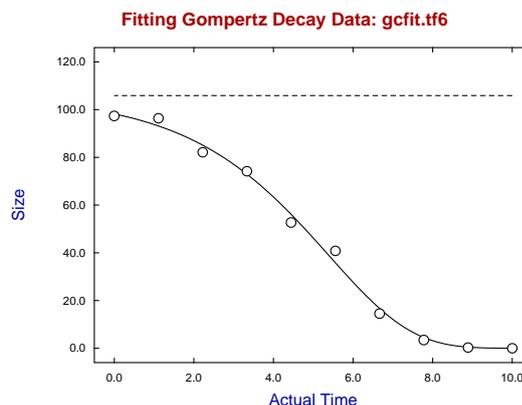
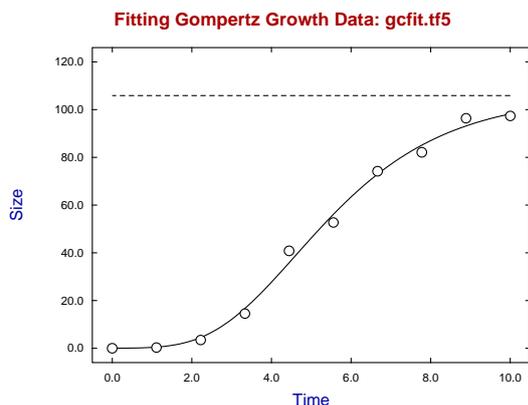
What happens in program `gcfit` when data in `gcfit.tf6` are analyzed is that the data are rearranged into the order of `gcfit.tf5` and then fitted by growth models as normal, except that some of the results and graphs are displayed in the original decay order with the original time scale.

First consider the parameters estimated for data in test file `gcfit.tf5`.

Parameter	Value	Std.error	Lower95%cl	Upper95%cl	p
A	105.87	4.6415	94.898	116.85	0.0000
B	9.1665	1.9138	4.6412	13.692	0.0020
k	0.48054	0.052253	0.35698	0.60410	0.0000
$t_{1/2}$	5.3733	0.20578	4.8867	5.8599	0.0000

Now consider the parameters estimated for data in test file `gcfit.tf6` and the best fit curves.

Parameter	Value	Std.error	Lower95%cl	Upper95%cl	p
A	105.87	4.6415	94.898	116.85	0.0000
B	9.1665	1.9138	4.6412	13.692	0.0020
k	0.48054	0.052253	0.35698	0.60410	0.0000
$t_{1/2}$	4.6267	0.20578	4.1401	5.1133	0.0000



What has happened is that the following model was fitted to both of these data sets

$$S(t) = A \exp[-B \exp(-kt)]$$

but the only difference in the parameter estimates and graphs is that the data are presented in the original time scale for $t_{1/2}$ for the decay data and not using the transformed time $T = t_{max} + t_{min} - t$.

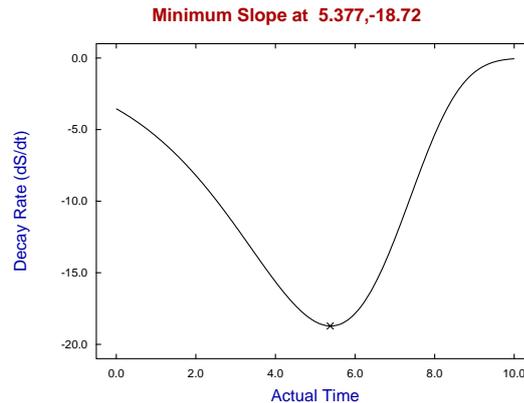
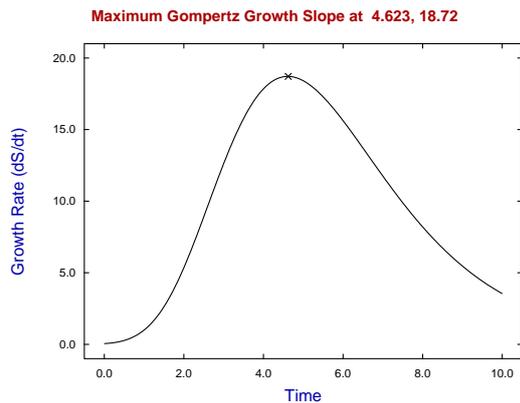
A similar situation is encountered when comparing the maximum and minimum slopes evaluated at the data points with the maximum and minimum values evaluated along the coordinates of the best fit curves. This is illustrated by the results displayed for extreme gradients in the extracts from the results log files and further clarified by the gradient plots.

Results for growth data

Maximum observed growth rate	18.655	Best fit curve maximum	18.716
Time when max. rate observed	4.4440	Best fit curve time	4.6107
Minimum observed growth rate	0.048725	Best fit curve minimum	0.048725
Time when min. rate observed	0.0000	Best fit curve time	0.0000

Results for decay data

Minimum observed growth rate	-18.655	Best fit curve minimum	-18.716
Time when min. rate observed	5.5560	Best fit curve time	5.3893
Maximum observed growth rate	-0.048725	Best fit curve maximum	-0.048725
Time when max. rate observed	10.000	Best fit curve time	10.000



The conclusion is simply that, if SimFIT program **gcfi**t is supplied with decay data, the data will be reversed and fitted by the growth models, but the output tables and graphs will use the original decay coordinates.

Example 3: Nonlinear survival curves

In mode 2, **gcfi**t fits a sequence of survival curves for data smoothing where it is assumed that the data are uncorrelated estimates of fractions surviving $0 \leq S(t) \leq 1$ as a function of time $t \geq 0$, e.g. such as would result from using independent samples for each time point. However, as normalizing data to $S(0) = 1$ can introduce bias, mode 2 allows an amplitude factor to be estimated.

It is important to realize that, if any censoring has taken place, the estimated fraction should be corrected for this. In other words, you start with a population of known size and, as time elapses, you estimate the fraction surviving by any sampling technique that gives estimates corrected to the original population at time zero.

The test files **weibull.tf1** and **gompertz.tf1** contain some exact data, which you can fit to see how mode 2 works. Then you can add error to simulate reality using program **adderr**. Note that you prepare your own data files for mode 2 using the same format as for program **makfil**, making sure that the fractions are between zero and one, and that only nonnegative times are allowed. It is probably best to do unweighted regression with this sort of data unless the variance of the sampling technique has been investigated independently.

In survival mode the time to half maximum response is estimated with 95% confidence limits and this can be used to estimate *LD50*. The survivor function is $S(t) = 1 - F(t)$, the *pdf* is $f(t)$, i.e. $f(t) = -dS/dt$, the hazard function is $h(t) = f(t)/S(t)$, and the cumulative hazard is $H(t) = -\log(S(t))$. Plots are provided for $S(t), f(t), h(t), \log[h(t)]$ and, as in mode 1, a summary is given to help choose the best fit model from the models provided, all of which decrease monotonically from $S(0) = 1$ to $S(\infty) = 0$ with increasing time.

The test file `weibull.tf1` has the following data

Time	Fraction	s.e.
0.000	1.000	1
1.000	0.9048	1
2.000	0.6703	1
3.000	0.4066	1
4.000	0.2019	1
5.000	0.08208	1

simulated by program `makdat` using the Weibull model

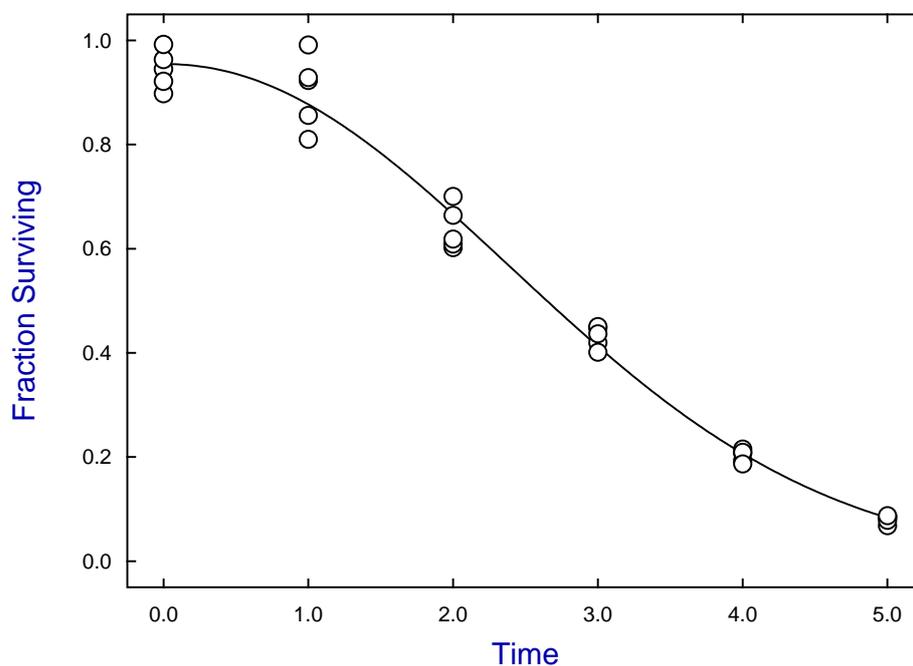
$$S(t) = p_1 \exp[-p_2 t^{p_3}]$$

$$= S_0 \exp[-(At)^B]$$

for $p_1 = 1, p_2 = 0.1, p_3 = 2.0$ which, in the nomenclature used by `gcf` is $S_0 = 1.0, A = 0.362, B = 2.0$. Then 7.5% relative error was added for five replicates using program `adderr` to generate test file `weibull.tf2` which was analyzed using the option to estimate S_0 giving the following table of parameter estimates and the best fit curve plot. Of course, if the starting fraction were known exactly, as in actual survival data, there would be no values for $t = 0$ since it would be assumed that $S_0 = 1$. However, allowing the $t = 0$ value to be estimated should perhaps always be used for data smoothing to avoid bias.

Parameter	Value	Std.error	Lower95%cl	Upper95%cl	<i>p</i>
<i>A</i>	0.30648	0.0056521	0.29489	0.31808	0.0000
<i>B</i>	2.0879	0.11926	1.8432	2.3326	0.0000
S_0	0.95465	0.015961	0.92190	0.98740	0.0000
$t_{1/2}$	2.7375	0.063837	2.6065	2.8685	0.0000

Fitting a Weibull Survival Model



Growth and Survival Models available in program gcfit

1. Exponential model $dS/dt = kS$
 $S(t) = A \exp(kt)$, where $A = S_0$
2. Monomolecular model $dS/dt = k(A - S)$
 $S(t) = A[1 - B \exp(-kt)]$, where $B = 1 - S_0/A$
3. Logistic model $dS/dt = kS(A - S)/A$
 $S(t) = A/[1 + B \exp(-kt)]$, where $B = A/S_0 - 1$
4. Gompertz model $dS/dt = kS[\log(A) - \log(S)]$
 $S(t) = A \exp[-B \exp(-kt)]$, where $B = \log(A/S_0)$
5. Von Bertalanffy 2/3 model $dS/dt = \eta S^{2/3} - \kappa S$
 $S(t) = [A^{1/3} - B \exp(-kt)]^3$
 where $A^{1/3} = \eta/\kappa$, $B = \eta/\kappa - S_0^{1/3}$, $k = \kappa/3$
6. Model 3 with constant $f(t) = S(t) - C$
 $df/dt = dS/dt = kf(t)(A - f(t))/A$
 $S(t) = A/[1 + B \exp(-kt)] + C$
7. Model 4 with constant $f(t) = S(t) - C$
 $df/dt = dS/dt = kf(t)[\log(A) - \log(f(t))]$
 $S(t) = A \exp[-B \exp(-kt)] + C$
8. Model 5 with constant $f(t) = S(t) - C$
 $df/dt = dS/dt = \eta f(t)^{2/3} - \kappa f(t)$
 $S(t) = [A^{1/3} - B \exp(-kt)]^3 + C$
9. Richards model $dS/dt = \eta S^m - \kappa S$
 $S(t) = [A^{1-m} - B \exp(-kt)]^{1/(1-m)}$
 where $A^{1-m} = \eta/\kappa$, $B = \eta/\kappa - S_0^{1-m}$, $k = \kappa(1 - m)$
 if $m < 1$ then η , κ , A and B are > 0
 if $m > 1$ then $A > 0$ but η , κ and B are < 0
10. Preece and Baines model $f(t) = \exp[k_0(t - \theta)] + \exp[k_1(t - \theta)]$
 $S(t) = h_1 - 2(h_1 - h_0)/f(t)$
1. Exponential survival model $S(t) = \exp(-At)$
 $f(t) = AS(t)$
 $h(t) = A$
2. Weibull survival model $S(t) = \exp[-(At)^B]$
 $f(t) = AB[(At)^{B-1}]S(t)$
 $h(t) = AB(At)^{B-1}$
3. Gompertz survival model $S(t) = \exp[-(B/A)\{\exp(At) - 1\}]$
 $f(t) = B \exp(At)S(t)$
 $h(t) = B \exp(At)$
4. Log-logistic survival model $S(t) = 1/[1 + (At)^B]$
 $f(t) = AB(At)^{B-1}/[1 + (At)^B]^2$
 $h(t) = AB(At)^{B-1}/[1 + (At)^B]$