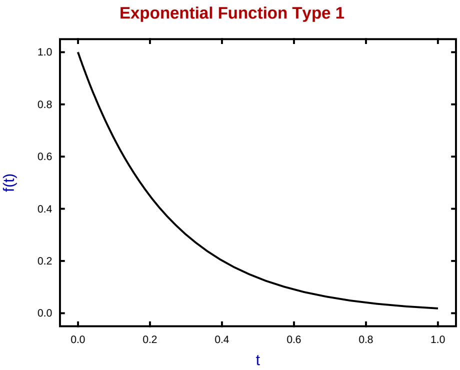
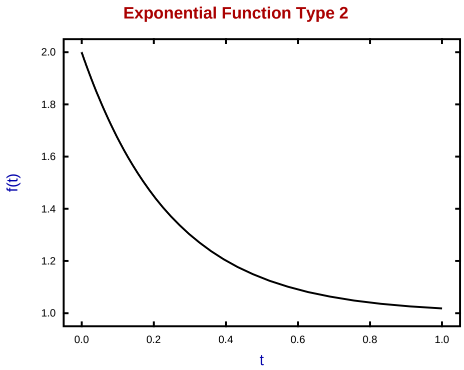
Tutorials and worked examples for simulation,
curve fitting, statistical analysis, and plotting.
http://www.simfit.org.uk

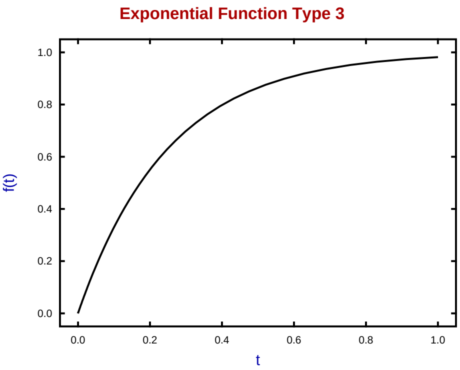
Exponential functions have wide applications in data analysis and the SimFIT package has a dedicated utility to fit six main categories of multi-exponential functions as illustrated below.

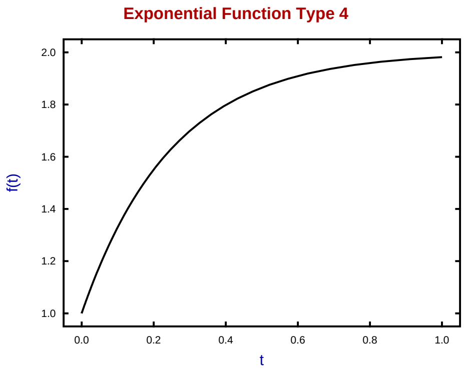
Simple exponential decline to zero (Type 1) or to a baseline (Type 2).



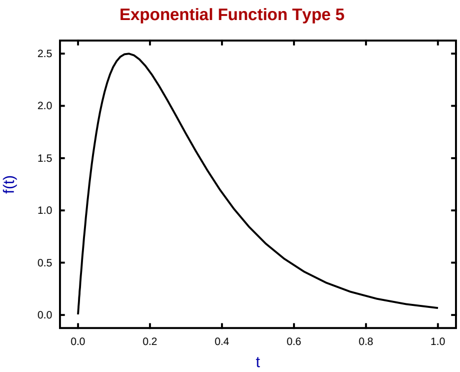


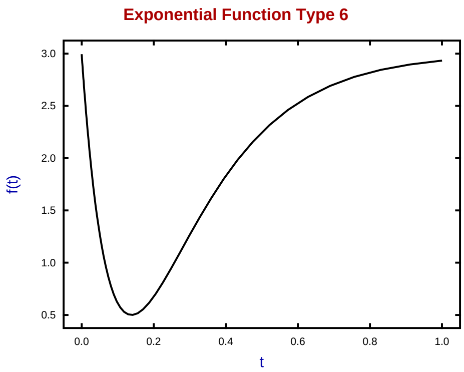
Monomolecular rise to a horizontal asymptote from zero (Type 3) or a baseline (Type 4).





Up-down (Type 5) or Down-up (Type 6) with at least two exponential terms.





In fact all of the exponential types are special cases of the following model

$$f(t) = A_1 \exp(-k_1 t) + A_2 \exp(-k_2 t) + \cdots + A_n \exp(-k_n t) + C$$

where the decisions facing the user are to define the category, i.e. the Type, and the number n of exponentials required.

There are three distinct SIMFIT programs provided for fitting exponentials to data.

1. **exfit**
   This is a simple user-friendly interface to automatically scale the data, locate sensible starting estimates, and perform unconstrained weighted least squares fitting with goodness of fit analysis and model discrimination. It requires the data to be nonnegative, i.e. $f(t) \ge 0$ for $t \ge 0$, and also the time to start must be zero, i.e. $t = 0$, and t must be in nondecreasing order, as this is assumed when scaling and finding starting estimates. Additional parameters like the area under the curve (AUC) are calculated.

   Normally you should fit one then two exponentials, but in the case of Type 5 and Type 6 the lowest order to fit is two exponentials. After fitting Type 5 or Type 6 it is recommended to do a further relaxation fit to allow for variations in the starting and asymptotic values in order to fit data such as pharmacokinetics after a bolus ingestion. It is possible to fit models with more than two exponentials but this may require several random starts or user-defined starting values for success. In any case, model discrimination with more than two exponentials is somewhat unpredictable as explained in the following publication.

   The F test for model discrimination with exponential functions.
   Bardsley,W.G., McGinlay,P.B. & Wright,A.J. (1986) *Biometrika* **73**, 501-508

2. **qnfit**
   This is a quasi-Newton constrained nonlinear regression program for more experienced analysts which requires user-defined starting estimates and parameter limits. It has the advantage that fitting does not necessarily require $f(t) \ge 0, t \ge 0$ or t in nondecreasing order and provides many options for setting starting estimates and parameter limits as well as numerous fine tuning possibilities.

3. **deqsol**
   This simulates and fits systems of nonlinear differential equations.

Example 1: Simple exponential decay (Type 1)

This will illustrate fitting the most frequently used exponential model, namely exponential decay from a positive value at $t = 0$ to a final zero asymptote as $t \to \infty$.

From the main SIMFIT menu choose [A/Z] then open program **exfit** and read the default test file provided which is `exfit.tf4`. If you choose to start with a model with one exponential the program will first fit the equation

$$f_1(t) = A_1 \exp(-k_1 t)$$

giving goodness of fit criteria, and if you select to fit up to order 2 the program will then fit the model

$$f_2(t) = A_1 \exp(-k_1 t) + A_2 \exp(-k_2 t)$$

giving information to compare these two fits, followed by the option for graphical deconvolution of $f_2(t)$ to illustrate the relative contribution of the two exponential terms.

The data, results tables, and graphs follow but note that the subscripts in equation $f_2(t)$ are arbitrary so, to preserve uniformity, parameters are rearranged if necessary after fitting so that the amplitudes A_i are in nondecreasing order.

t	$f(t)$	se
0.035983	1.7440	0.048730
0.035983	1.8367	0.048730
0.035983	1.8164	0.048730
0.054896	1.7028	0.033089
0.054896	1.6480	0.033089
0.054896	1.7075	0.033089
0.083750	1.6290	0.060314
0.083750	1.5359	0.060314
0.083750	1.6490	0.060314
0.12777	1.3919	0.013361
0.12777	1.3676	0.013361
0.12777	1.3702	0.013361
0.19493	1.1454	0.089321
0.19493	1.2240	0.089321
0.19493	1.3237	0.089321
0.29739	0.99897	0.043211
0.29739	0.94038	0.043211
0.29739	0.91466	0.043211
0.45370	0.80103	0.047423
0.45370	0.70902	0.047423
0.45370	0.73507	0.047423
0.69217	0.53660	0.041121
0.69217	0.50323	0.041121
0.69217	0.58501	0.041121
1.0560	3.8157	0.023248
1.0560	3.4769	0.023248
1.0560	3.9221	0.023248
1.6110	1.8573	0.011054
1.6110	1.9103	0.011054
1.6110	2.0697	0.011054

The columns contain data in the following format.

1. **Column 1**: the non–negative time t which must be in non-decreasing order.
2. **Column 2**: the non–negative response $f(t)$ presumed to be dependent on time in column 1.
3. **Column 3**: the positive sample standard deviation of the replicate response measurements. This column can be omitted or set to 1 if unweighted regression is required.

The results from fitting two exponentials are as follows.

Parameter	Value	Std. error	Lower95%cl	Upper95%cl	p
A_1	0.85255	0.067715	0.71336	0.99174	0.0000
A_2	1.1764	0.074759	1.0228	1.3301	0.0000
k_1	6.7934	0.85439	5.0372	8.5496	0.0000
k_2	1.1121	0.051102	1.0070	1.2171	0.0000
AUC	1.1834	0.014710	1.1532	1.2136	0.0000

AUC is the area under the curve from $t = 0$ to $t = \infty$
Initial time point (A) = 0.035983
Final time point (B) = 1.6110
Area (from $t = A$ to $t = B$) = 0.93832
Average over range (A, B) = 0.59575

Parameter correlation matrix

1			
-0.8757	1		
-0.5963	0.8996	1	
-0.8479	0.9485	0.8199	1

Analysis of residuals: $WSSQ$	24.397
$P(\chi^2 \geq WSSQ)$	0.5533
R^2, cc(theory,data)^2	0.9934
Largest absolute relative residual	11.99%
Smallest absolute relative residual	0.52%
Average absolute relative residual	3.87%
Absolute relative residuals in range 0.1–0.2	3.33%
Absolute relative residuals in range 0.2–0.4	0.00%
Absolute relative residuals in range 0.4–0.8	0.00%
Absolute relative residuals $>$ 0.8	0.00%
Number of negative residuals (n_1)	15
Number of positive residuals (n_2)	15
Number of runs observed (r)	16
P(runs $\leq r$: given n_1 and n_2)	0.5759
5% lower tail point	11
1% lower tail point	9
P(runs $\leq r$: given n_1 plus n_2)	0.6445
P(signs $\leq$ least number observed)	1.000
Durbin-Watson test statistic	1.8061
Shapiro-Wilks W statistic	0.9387
Significance level of W	0.0841
Akaike AIC (Schwarz SC) statistics	1.7979 (7.4027)

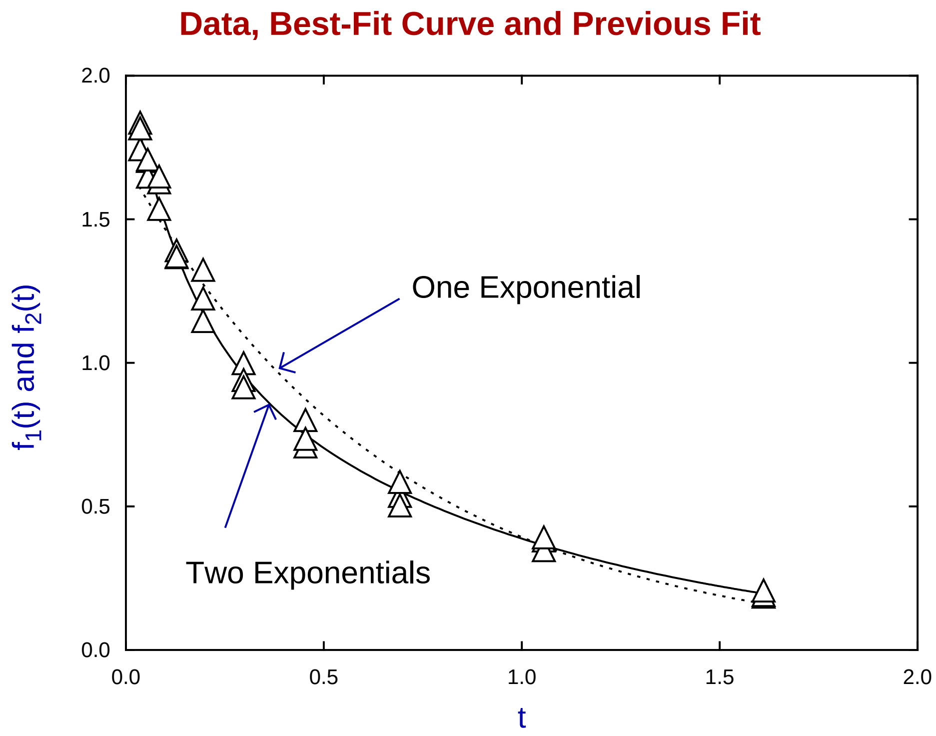
Verdict on goodness of fit: incredible

$WSSQ$-previous	224.9
$WSSQ$-current	24.4
Number of parameters-previous	2
Number of parameters-current	4
Number of x-values	30
Akaike AIC-previous	64.44
Akaike AIC-current	1.798, $ER = 3.998E + 13$
Schwarz SC-previous	67.24
Schwarz SC-current	7.403
Mallows C_p	213.7, $C_p/m_1 = 106.9$
Numerator degrees of freedom	2
Denominator degrees of freedom	26
F test statistic (FS)	106.9
$P(F \geq FS)$	0.0000
$P(F \leq FS)$	1.0000
5% upper tail point	3.369
1% upper tail point	5.526

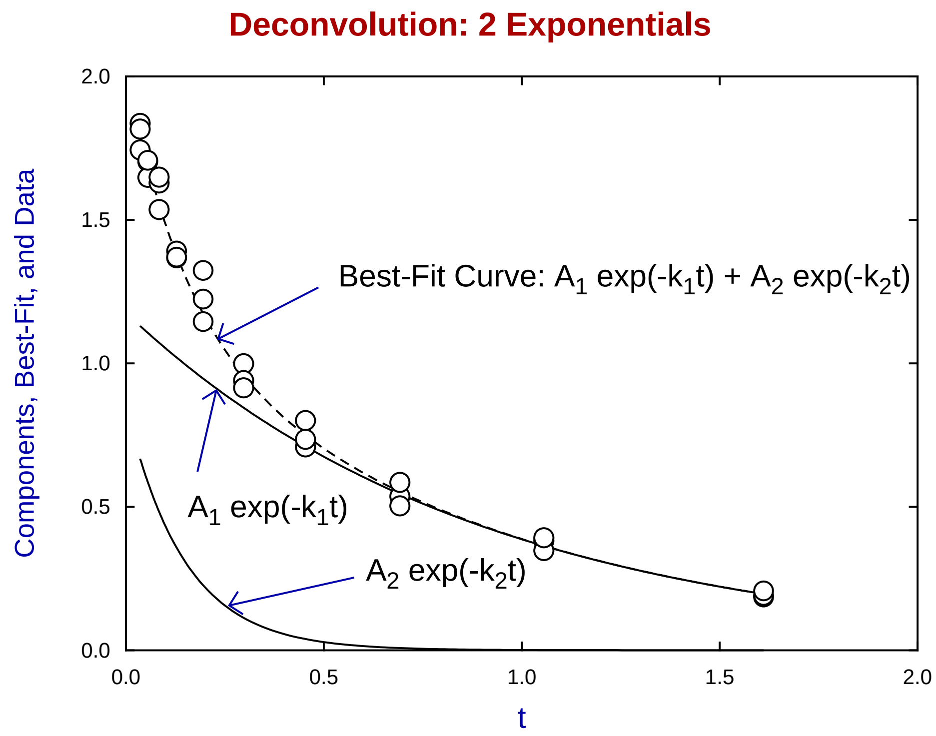
Conclusion based on the F test

Reject the previous model at 1% significance level
There is strong support for the extra parameters
Tentatively accept the current best fit model

The fit with two exponentials is clearly better than the fit with one exponential as show in the next graph.



From the next plot it is evident that both components make a significant contribution to the best fit double exponential curve.



Example 2: sequential kinetics (Type 5)

Another scheme that is frequently encountered is with an irreversible chemical reaction such as

$$X \xrightarrow{k_1} Y \xrightarrow{k_2} Z$$

with rate constants k_1 and k_2 and initial conditions $X(0) = X_0, Y(0) = Y_0, Z(0) = Z_0$ leading to

$$\begin{aligned}
X(t) &= X_0 \exp(-k_1 t)\\
Y(t) &= \frac{X_0 k_1}{k_2 - k_1}\exp(-k_1 t) + \left[Y_0 - \frac{X_0 k_1}{k_2 - k_1}\right]\exp(-k_2 t)\\
Z(t) &= X_0 + Y_0 + Z_0 - \frac{X_0 k_2}{k_2 - k_1}\exp(-k_1 t) - \left[Y_0 - \frac{X_0 k_1}{k_2 - k_1}\right]\exp(-k_2 t).
\end{aligned}$$

In the special case $X_0 > 0, Y_0 = 0, Z_0 = 0$ the expression for $Y(t)$ reduces to

$$\begin{aligned}
Y(t) &= \frac{X_0 k_1}{k_2 - k_1}[\exp(-k_1 t) - \exp(k_2 t)]\\
&= X_0 kt \exp(-kt) \text{ if } k = k_1 = k_2.
\end{aligned}$$

A similar expression is often encountered in pharmacokinetics, for instance if $Y(t)$ is the concentration of a substance in the blood after it ingested at $t = 0$, then absorbed from the stomach with rate constant k_1 but eliminated from the system with rate constant k_2. However several complications of this simple scheme are often encountered.

1. There may be insufficient data to characterize the early data points.
2. There may be insufficient time to record the final data points.
3. Data may arise from a repeated experiment with insufficient time for complete washout.
4. $Y(0)$ and/or $Z(0)$ may not be zero.
5. There may be additional steps requiring additional exponential terms.
6. Instead of rising then falling as in Type 5 the observed response may be a decrease followed by an increase as with Type 6.

Program **exfit** can attempt to deal with such issues by allowing additional exponential terms to be added and by allowing a relaxation phase to follow an initial fitting phase. For instance, this simplified three parameter model $g(t)$ can be fitted first

$$g(t) = A[\exp(-k_1 t) - \exp(k_2 t)]$$

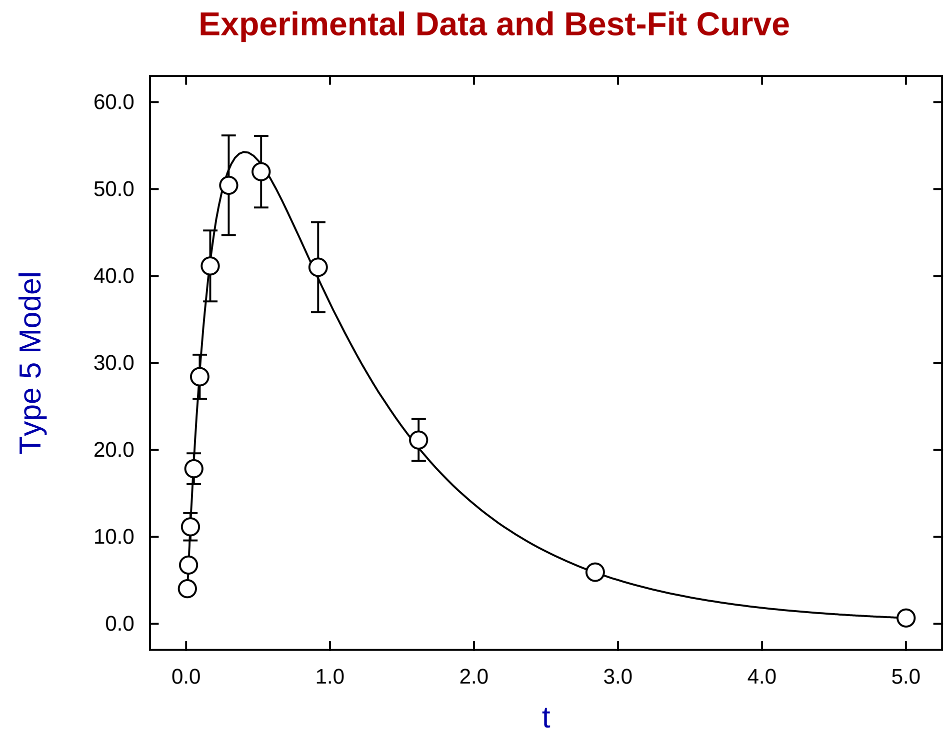
followed by using the parameter estimates from this fit as starting estimates to fit the richer four parameter model $h(t)$

$$h(t) = A_1 \exp(-k_1 t) + A_2 \exp(-k_2 t)$$

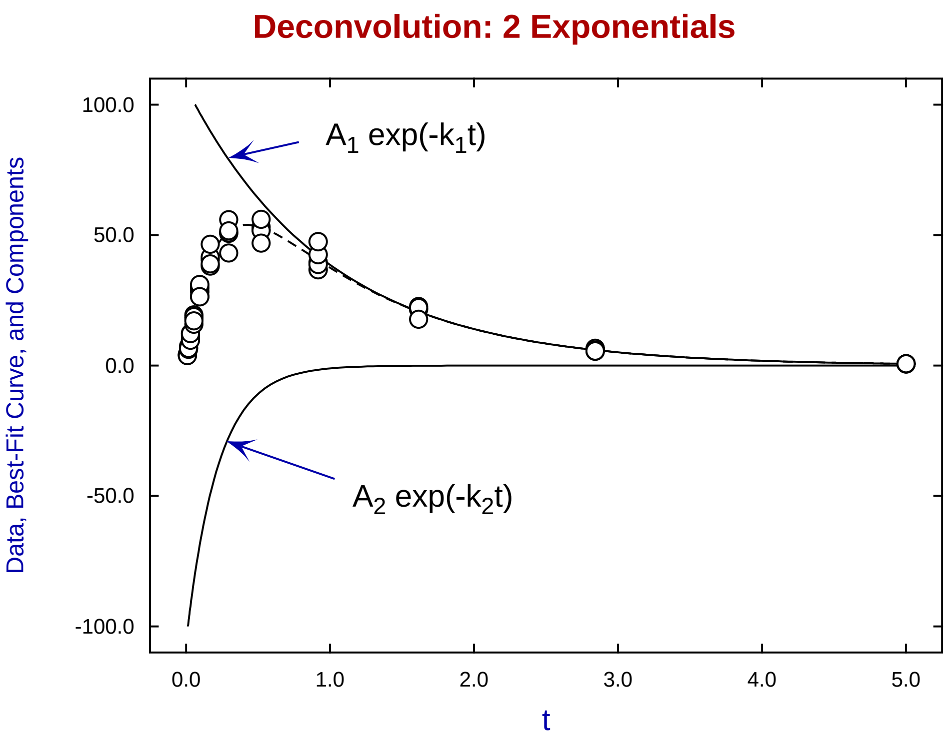
which will attempt to retain the up-down character supposed in the data. This relaxation method should be used when fitting Type 5 and Type 6 exponential models. However, in order to deal with models such as these it is best to use the greater versatility of program **qnfit**.

To illustrate fitting such a model the exact data in test file `exfit.tf5` was input into SIMFIT program **adderr** and five replicates per point were generated with 7.5% relative error with weights calculated from the sets of five replicates at each time point.

The simulated data and best-fit Type 5 curve from program **exfit** was as follows.



The graphical deconvolution of the best fit Type 5 model is shown next.



Theory

SimFIT program **exfit** will attempt to fit the multi-exponential model

$$f(t) = \sum_{i=1}^{n} A_i \exp(-k_i t) + C$$

by scaling the data, calculating starting estimates, then performing unconstrained nonlinear regression. It will only succeed if the data are extensive and accurate over a wide time range, the absolute values of the amplitudes A_i are similar, the rate constants k_i are sufficiently distinct, and the value of n modest, say $n \le 3$. For more extreme conditions it may be necessary to input starting estimates interactively, or preferably use the advanced programs **qnfit** or **deqsol**.

The reason why users have to choose which of the six exponential types to fit is that the scaling, calculation of starting estimates, and parameterization of the model has considerable influence on the success of optimization and model discrimination. To appreciate this, consider a system where several irreversible first order processes are taking place as in this scheme.

$$\begin{aligned} &X_1 \stackrel{k_1}{\rightarrow} Y_1 \\ &X_2 \stackrel{k_2}{\rightarrow} Y_2 \\ &\dots \\ &X_n \stackrel{k_n}{\rightarrow} Y_n. \end{aligned}$$

For each independent component we would have the solutions

$$\begin{aligned} X_i &= X_{i0} \exp(-k_i t) \\ Y_i &= Y_{i0} + X_{i0}(1 - exp(-kt)) \end{aligned}$$

so that fitting a model to $\sum_{i=1}^{n} X_i(t)$ would require a Type 1 model, while fitting $\sum_{i=1}^{n} Y_i(t)$ would require a Type 4 model or a Type 3 model if $Y_{i0} = 0$.

As explained in Example 2, a consecutive scheme like

$$X_1 \stackrel{k_1}{\rightarrow} X_2 \stackrel{k_2}{\rightarrow} X_3 \dots \stackrel{k_n}{\rightarrow} X_n$$

would also lead to a different type of solution with n exponentials of Type 5 or Type 6.

The actual experimental situation could be further complicated for these reasons.

- Most processes fitted by exponential models are not irreversible but also involve backward flux.
- Reversible consecutive processes require $n \ge 2$ but here the exponential terms are not simple but involve calculating the eigenvalues for the system at each iteration.
- Cyclical consecutive processes can also lead to complex eigenvalues and oscillating solutions.

In such situations it is far better to model the situation as a set of differential equations and simulate then fit these using SimFIT program **deqsol**.