

Fitting Models to Data: C-peptide Kinetics in Humans

Real World Biomedical Modeling Techniques
Through Case Studies – Module 3
IEEE/EMBS and BMES Pre-Conference Workshop
October 22, 2002 • Westin Galleria • Houston, TX

Modeling Ingredients

- Data
 - Independent and dependent variables
 - Data measurement error (uncertainty)
- Model
 - Compartmental or other
- Procedures for:
 - Matching model and data
 - Inspecting the results (diagnostics)

Modeling Results

- Model predictions
 - Either matched to the data (Fit) or not (Solve)
- Optimized parameter estimates
 - Obtained through fitting
- Uncertainty on parameter estimates
 - Usually expressed as percent Coefficient Of Variation (CV)
 - Obtained how?

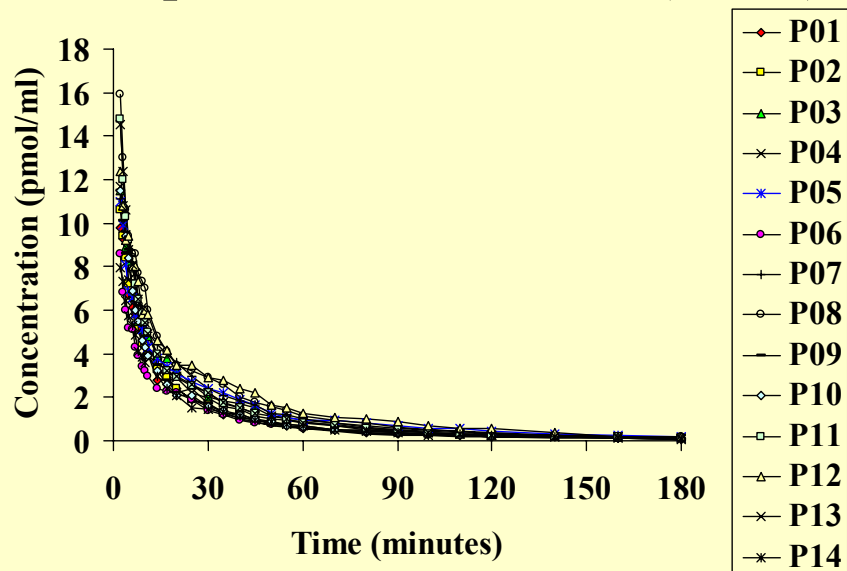
Multicompartmental System

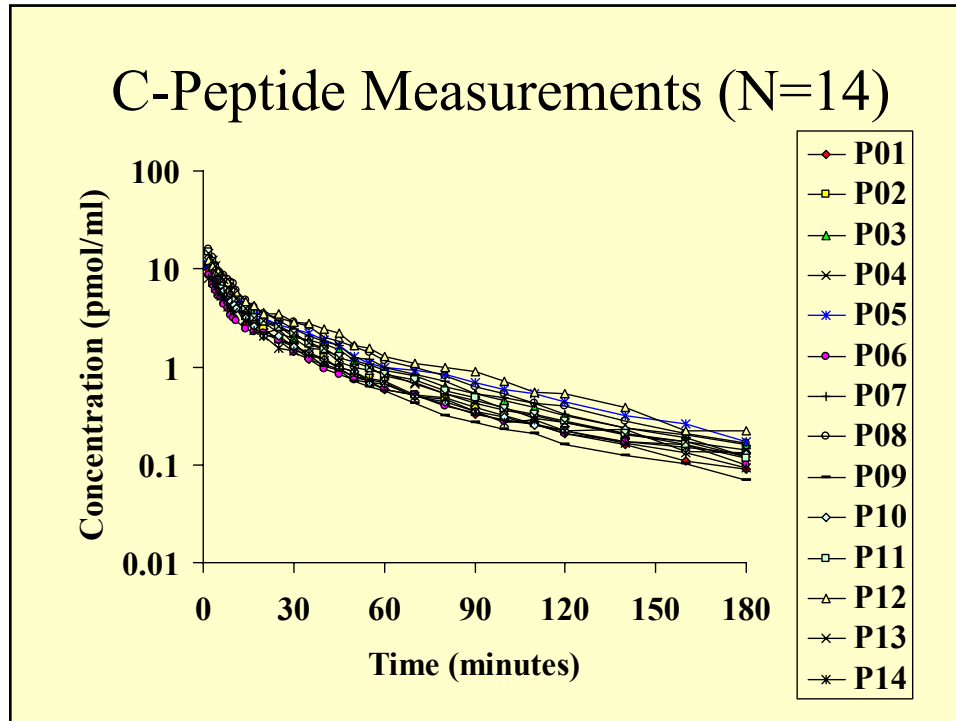
- Example: biosynthetic C-peptide kinetics
- Data are from: Sparacino G, Tombolato C, Cobelli C. Maximum-likelihood versus maximum a posteriori parameter estimation of physiological system models: the C-peptide impulse response case study. IEEE Trans Biomed Eng. 2000 Jun;47(6):801-11.

Biosynthetic C-Peptide Kinetics

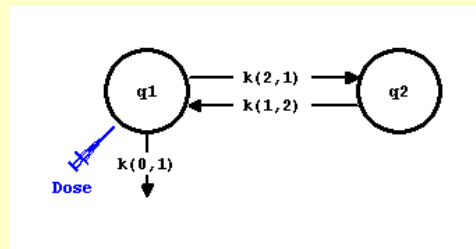
- 14 normal human subjects
- A bolus (average mass of about 50,000 pmol) of biosynthetic CP was administered intravenously
- In order to avoid the confounding effect of endogenously secreted CP, CP pancreatic secretion was suppressed through a somatostatin infusion started 2 hours before the experiment

C-Peptide Measurements (N=14)





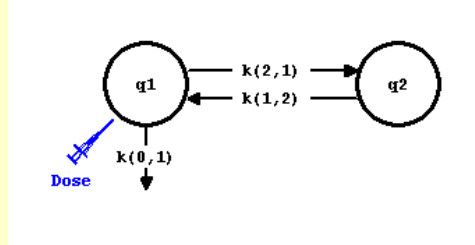
Multicompartmental Model = Balance of Mass



$$\frac{dq_1}{dt} = -[k(0,1) + k(2,1)]q_1(t) + k(1,2)q_2(t) + \text{Dose}(t)$$

$$\frac{dq_2}{dt} = k(2,1)q_1(t) - k(1,2)q_2(t)$$

Very Important: Conventions



- In SAAM II, which follows the engineering and mathematics convention, $k(i,j)$ is the transfer rate from compartment j to i
- This is opposite to the pharmacokinetics convention, where $k(i,j)$ goes from i to j
- Renaming the parameters is always an option

Compartmental Model

- The generic differential equation in a n -compartment model

$$\frac{dq_i}{dt} = -q_i(t) \sum_{\substack{j=0 \\ j \neq i}}^n k(j,i) + \sum_{\substack{j=1 \\ j \neq i}}^n k(i,j)q_j(t) + ex_i$$

Rate
of
change

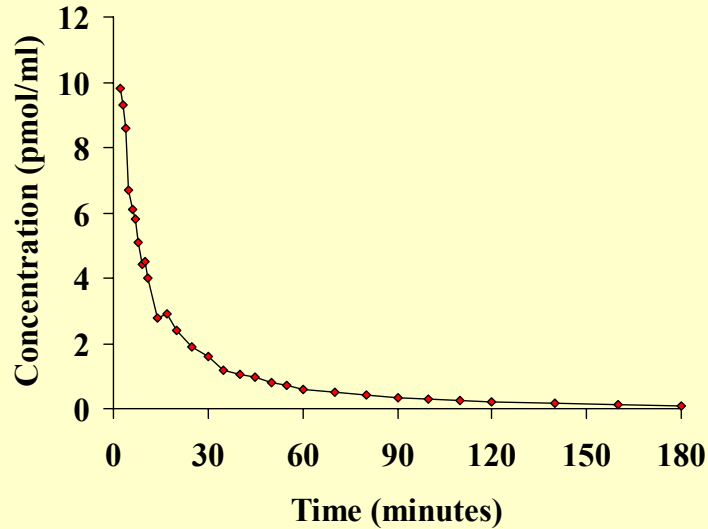
What goes out...

What comes in...

KEY: Balance of Mass

- The $k(i,j)$ are transfer rates. They can be linear or nonlinear. If they are linear, the model is a linear multicompartmental model.

Back to the Data...

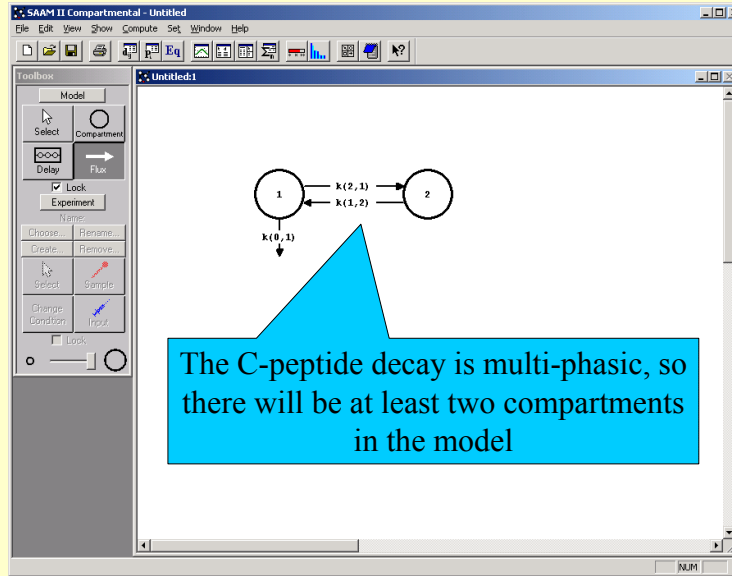


Data Format

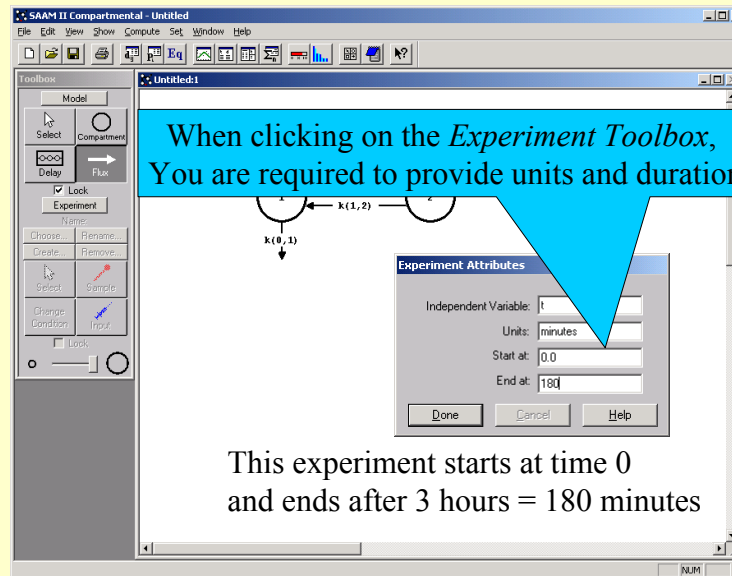
```
# Subject 01
# Dose 49650
DATA
(FSD 0.1)
t cpeptide
2 9.8
3 9.3
4 8.6
5 6.7
6 6.1
7 5.8
8 5.1
9 4.4
10 4.5
11 4
14 2.8
17 2.9
20 2.4
25 1.88
30 1.58
35 1.2
40 1.05
45 0.95
50 0.79
55 0.72
60 0.58
70 0.52
80 0.41
90 0.33
100 0.29
110 0.26
120 0.21
140 0.16
160 0.11
180 0.09
END
```

- Data are in column format, as we have seen
- Note that both the subject ID and the dose are with the data, but commented out (it pays to keep all this information together in one place)
- Units are: minutes for time, pmol/ml for C-peptide concentration
- The data measurement error is constant fractional standard deviation (FSD) equal to 10% of the measurement value
- We will see later the implications of this...

Define Your System Model



Define the Experiment Duration...



This experiment starts at time 0
and ends after 3 hours = 180 minutes

... and give it a Name!

The screenshot shows the SAAM II software interface. A compartmental model diagram is visible with two compartments, 1 and 2. Transitions are labeled with rate constants: $k(2,1)$ from compartment 2 to 1, $k(1,2)$ from compartment 1 to 2, and $k(0,1)$ from compartment 1 to the environment. A 'Create Experiment' dialog box is open, with 'New Name' set to 'Cpeplid' and 'Type' set to 'Tracer'. A blue callout box points to the dialog with the text: 'The experiment name is just one way to remember which variables you are analyzing with this file'.

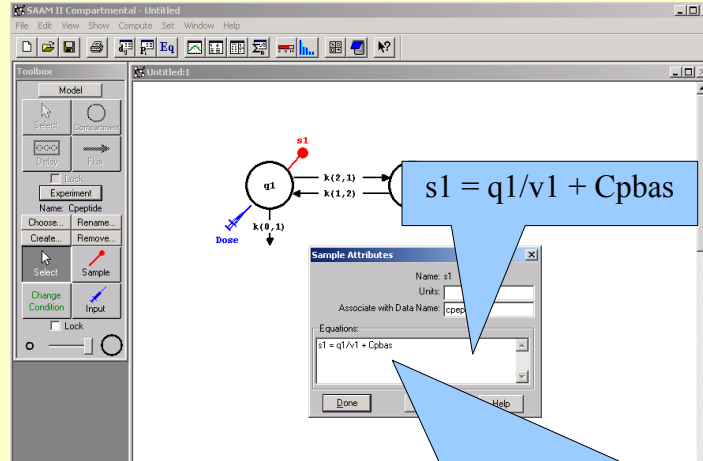
Exogenous Input (ex1)

The screenshot shows the 'Exogenous Input' dialog box. The 'Name' is 'Dose' and the 'Reference Name' is empty. The 'Units' field is also empty. A table shows the input parameters:

Type	Initial	Constant	Start	Stop	Repeat	Every	Nr. Repeats
Bolus	4.97e+4	-	0.000	-	-	-	-

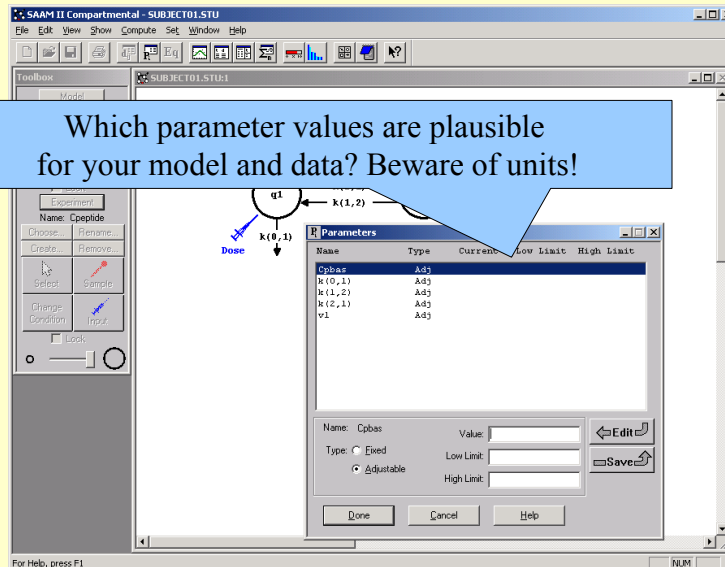
Below the table, the 'Input Type' is set to 'Bolus'. The 'Initial Amount' is 49650. The 'Constant Rate', 'Event Start', and 'Event Stop' fields are empty. A blue callout box points to the 'Initial Amount' field with the text: 'The dose is a pulse dose (bolus) of 49650 pmol administered at time 0 – it can be defined directly'. Buttons for 'Save', 'Edit', and 'Add' are visible on the right side of the dialog.

Sample (s1)



For this data set, we need to account for a basal level of C-peptide (which is an endogenous substrate)

Provide Parameter Values



Provide Parameter Values

The screenshot shows the SAAM II software interface. The main window displays a compartmental model with two compartments, q1 and q2. A red arrow labeled 's1' points into compartment q1. A blue arrow labeled 'Dose' also points into q1. There are two bidirectional arrows between q1 and q2, labeled with rate constants k(2,1) and k(1,2). A parameter k(0,1) is shown as an arrow pointing out of q1. A 'Parameters' dialog box is open, showing a table of parameters:

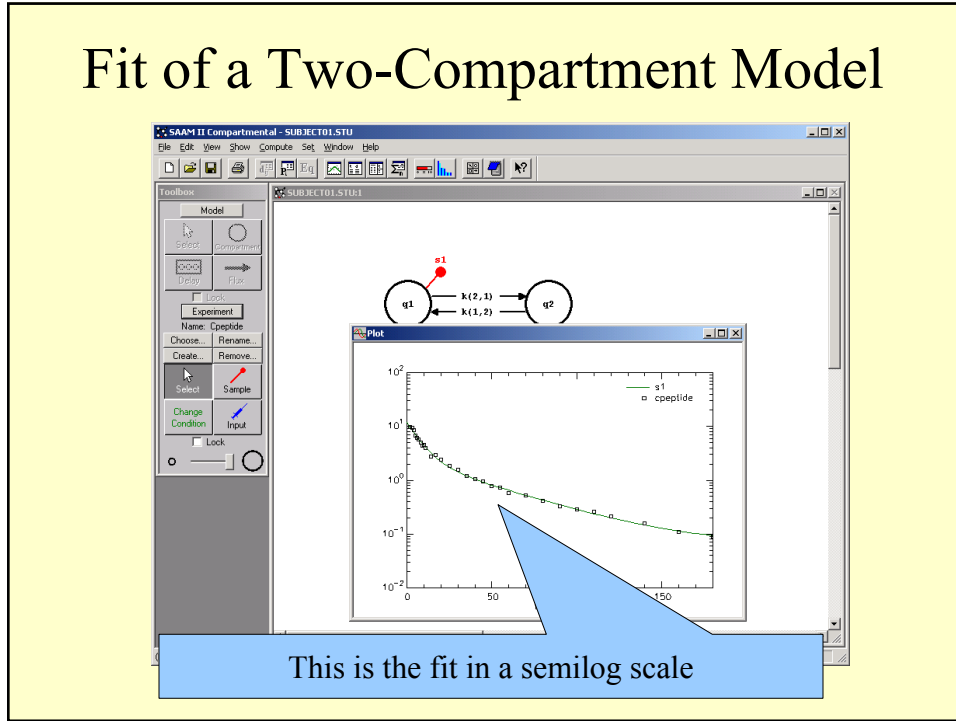
Name	Type	Current	Low Limit	High Limit
Cpbas	Adj	0.1000	0.0100	1.0000
k(0,1)	Adj	0.1000	0.0100	1.0000
k(1,2)	Adj	0.1000	0.0100	1.0000
k(2,1)	Adj	0.1000	0.0100	1.0000
v1	Adj	5000.0000	500.0000	50000.0000

Below the table, a blue callout box contains the text: "Here are some reasonable starting values: Cpbas=0.1 k01=0.1 k12=0.1 k21=0.1 v1=5000".

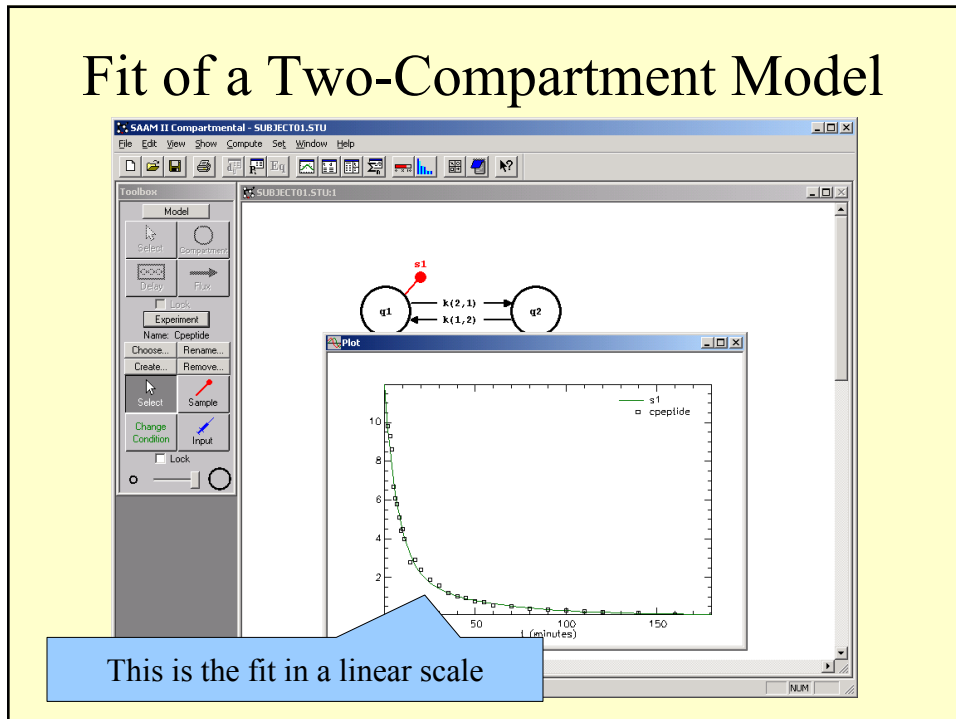
Solve – How Does It Look?

The screenshot shows the same SAAM II software interface, but now with a 'Plot' window open. The plot shows the concentration of 'cpeptide' (represented by open squares) over time. The y-axis is on a logarithmic scale from 10⁰ to 10². The x-axis represents time. The data points show a characteristic exponential decay curve. A blue callout box points to the plot with the text: "This is the solution in a semilog scale".

Fit of a Two-Compartment Model



Fit of a Two-Compartment Model



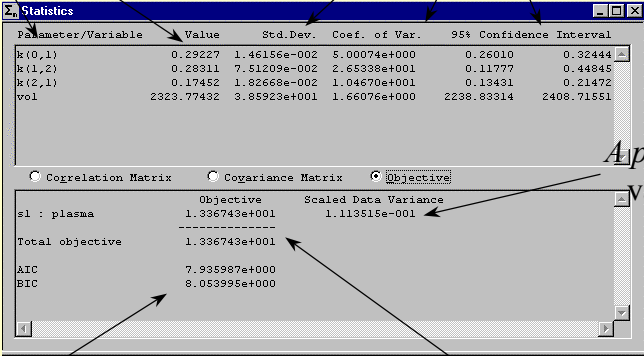
Output Following a Fit

- Parameter estimate and the statistics
- Correlation Matrix
- Objective function value
- Values for AIC and BIC: test for model parsimony (Module 6)
- Residual information
- Error messages if the fit is not successful

Statistics Window Output

Parameter names
and optimal values

Standard Deviations, Coefficients
of Variation and 95% Confidence Intervals



Parameter/Variable	Value	Std.Dev.	Coef. of Var.	95% Confidence Interval	
k(0,1)	0.29227	1.46166e-002	5.00074e+000	0.26010	0.32444
k(1,2)	0.28311	7.51209e-002	2.65338e+001	0.11777	0.44845
k(2,1)	0.17452	1.82668e-002	1.04670e+001	0.13431	0.21472
vol	2323.77432	3.85923e+001	1.66076e+000	2238.83314	2408.71551

	Objective	Scaled Data Variance
s1 : plasma	1.336743e+001	1.113515e-001
Total objective	1.336743e+001	
AIC	7.935987e+000	
BIC	8.053995e+000	

A posteriori
variance

Parsimony criteria:
Akaike (AIC) and Schwarz-Bayesian (BIC)

Objective function
minimal value

Parameter Precisions

$$CV[\theta] = \frac{SD[\theta]}{\theta}$$

- Usually they are expressed as %CV (coefficient of variation or fractional standard deviation times 100)
- Bad precisions ($CV > 100\%$) are often a symptom of a model which is too complex to be supported by the data
- They are reported in the Statistics Window
 - “Coef. Of Var.” is %CV
 - “Std.Dev.” is the SD

Residual and Weighted Residual

- Residuals and weighted residuals are components of the Objective Function. After a successful fit to a set of data, they can be used to test for goodness-of-fit.

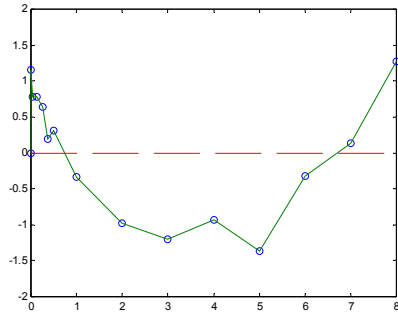
The residual (simple difference between data and model)

$$\text{res}(t_i) = y(t_i) - s(t_i, \theta)$$

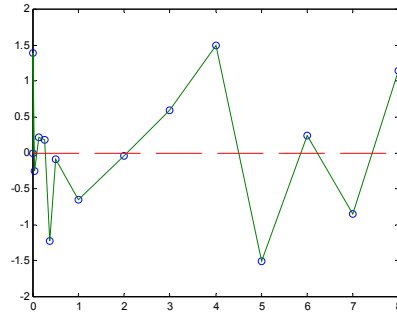
The weighted residual (weighted difference)

$$\text{wres}(t_i) = \frac{y(t_i) - s(t_i, \theta)}{\sigma(t_i, \theta)}$$

Weighted Residual Trends



- This set of weighted residuals supports model rejection

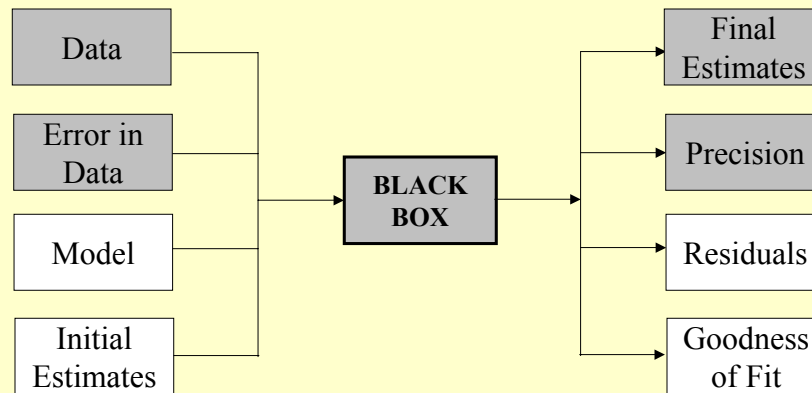


- This set of weighted residuals supports model acceptance

Hands-On: C-Peptide Data

Subject#	k(0,1) (CV)	k(1,2) (CV)	k(2,1) (CV)	v1 (CV)	Objective
01					
02					
03					
04					
05					
06					
07					
08					
09					
10					
11					
12					
13					
14					

Parameter Estimation: Fitting Models to Data



Model Assumptions

- We assume that the measurement value at the i -th time point is given in vector form by:

$$y(t_i) = s(t_i, \theta) + e_i, \quad i=1, \dots, N$$

where:

- y is the random vector of measurements
- s is a known function describing the kinetics (the sample)
- θ is the unknown vector of kinetic model parameters
- e_i is the measurement error random vector

Black Box in SAAM II: The Objective Function

- The objective function is a function which provides a measure of the difference between the model prediction and the data.
- Given an objective function, optimization is the mathematics that is used to minimize this function, i.e. figure out the parameter values that minimize the difference between the model predictions and the data.

The Objective Function in SAAM II

$$\text{OBJ} = \sum_{i=1}^n \left[\frac{y(t_i) - s(t_i, \theta)}{\sigma(t_i, \theta)} \right]^2 + \ln[\sigma^2(t_i, \theta)]$$

- The objective function penalizes the distance between the model prediction and the data
- The optimal weighting scheme is given by the measurement error variance: this way, the parameter confidence limits will be a function of data reliability

Some Definitions

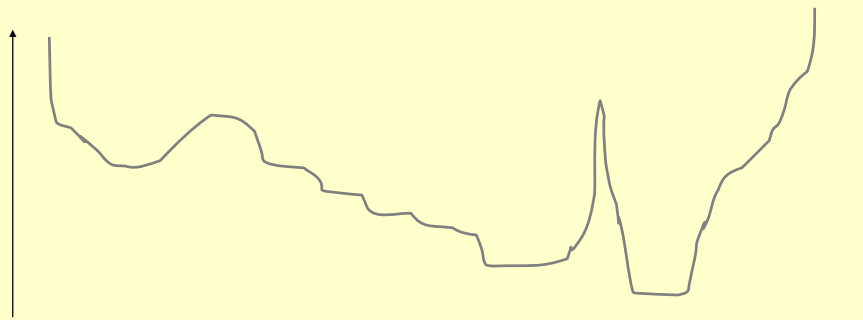
$$\text{OBJ} = \sum_{i=1}^n \left[\frac{y(t_i) - s(t_i, \theta)}{\sigma(t_i, \theta)} \right]^2 + \ln[\sigma^2(t_i, \theta)]$$

- OBJ is called the **extended least squares** objective function
- σ = standard deviation (the “error bar”)
- σ^2 = variance (the standard deviation squared)
- You can see that the inverse of the variance appears at the denominator of the objective
- The inverse of the variance is called the **weight**:
 $w_i = 1/\sigma^2$

Optimization

$$\hat{\theta} = \arg \min_{\theta} \text{OBJ}$$

OBJ



The minimum is found through iterations! Beware of local minima!

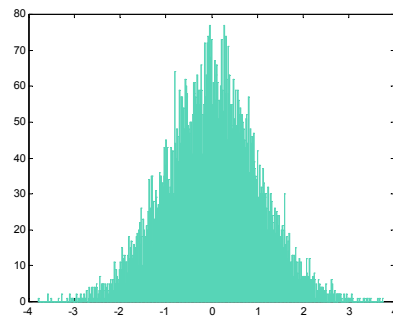
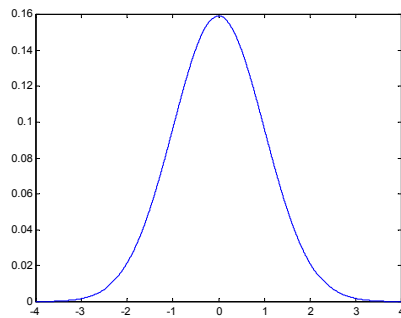
Basic Assumptions on the Measurement Error

The measurement error is assumed to be:

- Zero mean
(there are no offsets in the data!)
- Gaussian
- Independent from one sample to the other
- With given (known) variance structure

$$e \sim N(0, \sigma^2)$$

A Gaussian Distribution



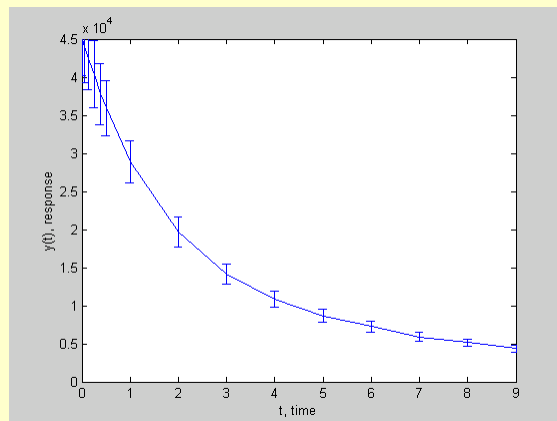
- Continuous Gaussian (normal) distribution with mean zero and variance 1
- Histogram of 1,000 discrete samples from the distribution on the left

SAAM II Weighting Schemes

- SD: Constant Standard Deviation
 - error in the data = constant = SD
- FSD: Constant Fractional Standard Deviation
 - error in the data = constant percent = CV% of the data/model value (CV and FSD are synonyms)
- POIS: Poisson Error
 - error in the data = $[A * y^{1/2}]$
- GEN: General Variance Model
 - error in the data = $[A + B * y^C]$
 - A, B and C are constants
 - y is either the data or the model-predicted value

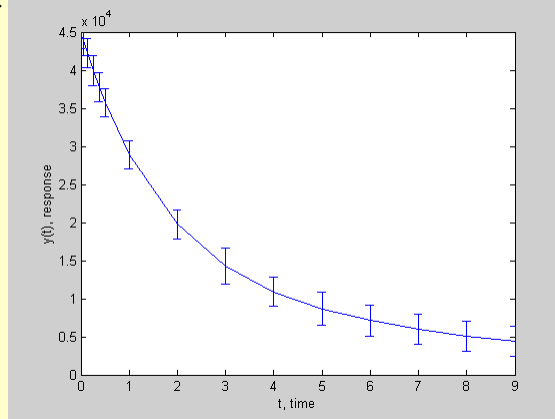
Error models: constant FSD

- The standard deviation of the error is a constant percent of the measurement
- *Small* data contribute *more*
- e.g. for $FSD = 0.05$
- 100 ± 5 has weight $w = 0.04$
- 1 ± 0.05 has weight $w = 400$



Error models: constant SD

- The standard deviation of the error is constant over all measurements
- *Large* data contribute *more*
- e.g. for $SD = 0.5$
- 100 ± 0.5 has weight $w = 4$
- 1 ± 0.5 has weight $w = 4$



Effect of Changing Weights

- We will now repeat the identification of the two-compartment model of C-peptide kinetics, but we will assume constant standard deviation weights
- In the Data file, you should substitute, for
(FSD 0.1)
a new variance model
(SD 1)

Hands-On: C-Peptide Data

Subject#	k(0,1) (CV)	k(1,2) (CV)	k(2,1) (CV)	v1 (CV)	Objective
01					
02					
03					
04					
05					
06					
07					
08					
09					
10					
11					
12					
13					
14					

Relative Weights

- When relative weighting is chosen, a proportionality constant is estimated for each set, and the error that was assumed in the Data Window gets multiplied by it
- This proportionality constant is reported in the Statistics Window as the “Scaled Data Variance”
- If you do not know your error structure exactly, use relative weighting

The Variance Model: Options in SAAM II

	Model	Data
Absolute	Precise knowledge of measurement error, and high noise level in data (not robust)	Precise knowledge of measurement error, and good noise level in data (very common)
Relative	Uncertain knowledge of measurement error, and high noise level in data (most robust)	Uncertain knowledge of measurement error, and good noise level in data (SAAM II default)

Computational Settings

The screenshot shows the 'Computational Settings' dialog box with the following callouts:

- Differential Equation Integrator Choice:** Points to the Integrator dropdown menu showing 'Rosenbrock', 'Pade', and 'Runge-Kutta'.
- Number of Calculation Points:** Points to the 'Min. Nr. of Calculations Intervals' dropdown set to '20'.
- Tolerance for Differential Equation Integration:** Points to the 'Use Relative Error' checkbox and its associated value field (0.00100000).
- Number of Optimizer Iterations:** Points to the 'Max. Nr. of Fit Iterations' dropdown set to '20'.
- Model for The Data Error:** Points to the 'Variance Model' section with radio buttons for 'Data' and 'Model'.
- Tolerance for Optimization:** Points to the 'Convergence Criterion' value field (0.00010000).
- Differencing For Gradient Optimization:** Points to the 'Derivative' section with radio buttons for 'Forward' and 'Central'.
- Bayesian Estimation:** Points to the 'Include Bayesian Term' checkbox and its 'Lambda' value field (10.00000000).

What Have We Learned?

- Introduction to model fitting
- To define a multicompartment model
- The role of data measurement error and weights in parameter estimation
- Choice of variance based on knowledge of data and experiment
- Definition of parameter statistics
- Assessment of residuals