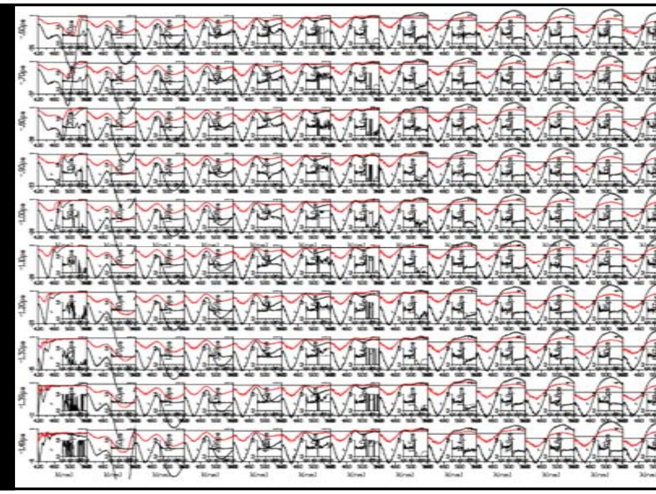
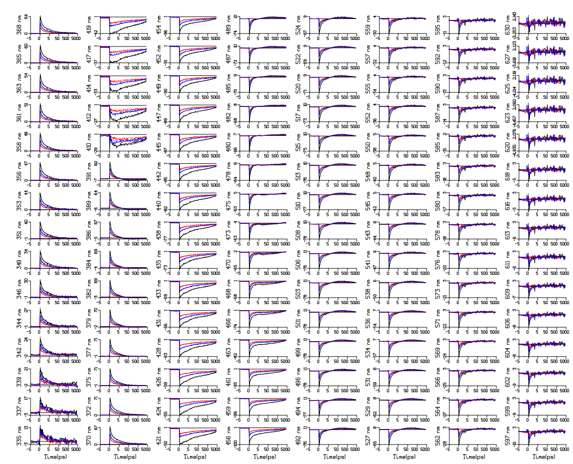
A word on data analysis

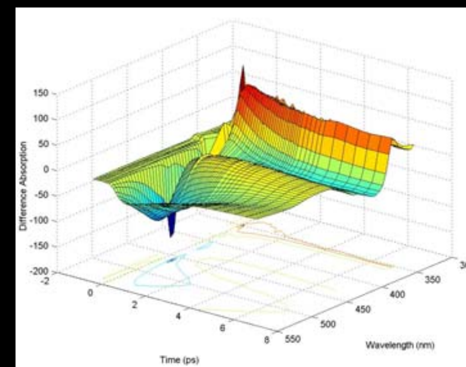


Or…

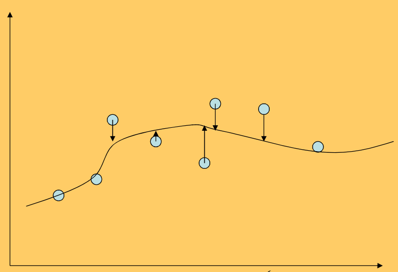


Or

Pump-probe dataset



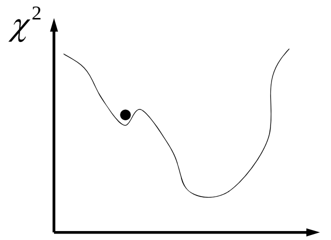
Merit function



$$\chi^2 = \sum_{j,k} \frac{\left(D_{jk} - F(t_j, \lambda_k)\right)^2}{\sigma_{jk}^2} = \sum_{j,k} \frac{\left(D_{jk} - \sum_i n_i(t) A_i(\lambda)\right)^2}{\sigma_{jk}^2}$$

Linear least squares

Global and local minima



Instrument response function

$$D(t) = \int_{-\infty}^{\infty} S(\tau) IRF(t-\tau)\, d\tau$$

Typical for time-resolved spectroscopy

$$I(t) = \frac{1}{\sqrt{2\pi}z} e^{-\frac{t^2}{2z^2}}$$

Instrument response function

$$\frac{dn}{dt} = \frac{A}{\sqrt{2\pi}z} e^{-\frac{t^2}{2z^2}} - \frac{1}{\tau_{fl}} n$$

$$n(t) = G(t) \cdot e^{-\frac{t}{\tau_{fl}}}$$

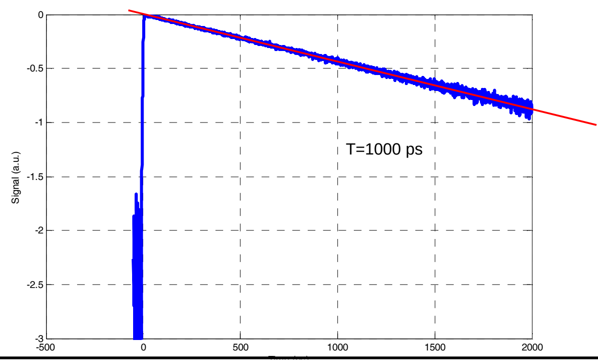
$$G(t) \sim 1 + erf\left(\frac{t}{z}\right)$$

$$erf(t) \equiv \frac{2}{\sqrt{\pi}} \int_0^t e^{-t^2} dt$$

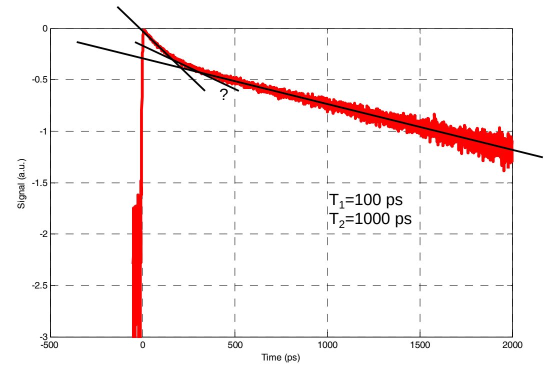
Gaussian instrument response

$$F(t,\lambda) = \frac{1}{2}\left(1 + erf\left(\frac{t}{z}\right)\right) \sum_{n=1}^{N} A_n(\lambda) e^{-\frac{t}{\tau_n}}$$

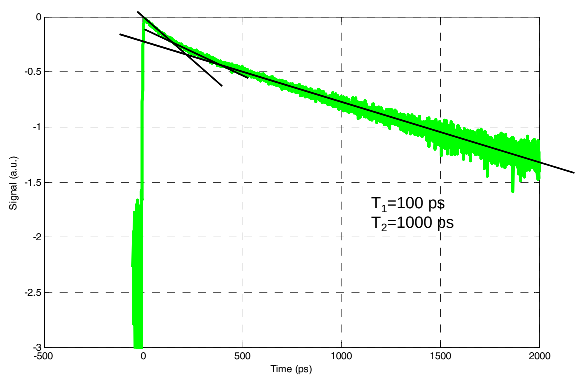
One, two three exponents



One, two three exponents



One, two three exponents



Singular value decomposition

$$\begin{pmatrix} \\ \mathrm{A} \\ \\ \end{pmatrix} = \begin{pmatrix} \\ \mathrm{U} \\ \\ \end{pmatrix} \cdot \begin{pmatrix} w_0 & & & \\ & w_1 & & \\ & & \cdots & \\ & & \cdots & \\ & & & w_{N-1} \end{pmatrix} \cdot \begin{pmatrix} \mathrm{V}^T \end{pmatrix}$$

$$\begin{pmatrix} \mathrm{A} \end{pmatrix} = \begin{pmatrix} \mathrm{U} \end{pmatrix} \cdot \begin{pmatrix} w_0 & & & \\ & w_1 & & \\ & & \cdots & \\ & & \cdots & \\ & & & w_{N-1} \end{pmatrix} \cdot \begin{pmatrix} \mathrm{V}^T \end{pmatrix}$$

$$\mathbf{V} \cdot \mathbf{V}^T = 1 \qquad \mathbf{V}^T \cdot \mathbf{V} = 1 \qquad \mathbf{U}^T \cdot \mathbf{U} = 1$$

$$\sum_{j=0}^{N-1} V_{jk} V_{jn} = \delta_{kn} \qquad \begin{array}{l} 0 \le k \le N-1 \\ 0 \le n \le N-1 \end{array} \qquad \sum_{i=0}^{M-1} U_{ik} U_{in} = \delta_{kn} \qquad \begin{array}{l} 0 \le k \le N-1 \\ 0 \le n \le N-1 \end{array}$$

Singular value decomposition

• Divides a matrix into 'most important columns', 'most important rows' and 'importance coefficients' (singular values);

• Great for solving linear systems of equations

$$\mathbf{Ax} = \mathbf{b}$$

in least-squares sense, i.e. finds **x**, such that

$$\left|\mathbf{Ax} - \mathbf{b}\right|$$

is minimized.

Solving linear system

$$\mathbf{A} = \mathbf{U} \cdot diag\left[\boldsymbol{\omega} \cdot\right] \mathbf{V}^T$$

$$\mathbf{V} \cdot diag\left[\frac{1}{\boldsymbol{\omega}}\right] \cdot \mathbf{U}^T \cdot \mathbf{U} \cdot \boldsymbol{\omega} \cdot \mathbf{V}^T = 1$$

$$\mathbf{A}^{-1} = \mathbf{V} \cdot diag\left[\frac{1}{\boldsymbol{\omega}}\right] \cdot \mathbf{U}^T$$

U and V columns/rows are orthogonal

$$\sum_{i=0}^{M-1} U_{ik} U_{in} = \delta_{kn} \qquad \begin{array}{l} 0 \le k \le N-1 \\ 0 \le n \le N-1 \end{array}$$

$$\sum_{j=0}^{N-1} V_{jk} V_{jn} = \delta_{kn} \qquad \begin{array}{l} 0 \le k \le N-1 \\ 0 \le n \le N-1 \end{array}$$

General linear least squares

- Linear combination of model functions:

$$y(x) = \sum_{k=0}^{M-1} a_k X_k(x)$$

- Chi square is the merit function, as before:

$$\chi^2 = \sum_{i=0}^{N-1} \left[\frac{y_i - \sum_{k=0}^{M-1} a_k X_k(x_i)}{\sigma_i} \right]^2$$

General linear least squares

- Normal equations are obtained by taking a derivative of chi square and zeroing it:

$$0 = \sum_{i=0}^{N-1} \frac{1}{\sigma_i^2} \left[y_i - \sum_{j=0}^{M-1} a_j X_j(x_i) \right] X_k(x_i)$$

$$\sum_{j=0}^{M-1} \alpha_{kj} a_j = \beta_k$$

$$\alpha_{kj} = \sum_{i=0}^{N-1} \frac{X_j(x_i) X_k(x_i)}{\sigma_i^2}$$

$$\beta_k = \sum_{i=0}^{N-1} \frac{y_i X_k(x_i)}{\sigma_i^2}$$

General linear least squares

- The system of equations constructed in such a way is called *normal equations.* Its solution is equivalent to solving the fitting problem.

$$\sum_{j=0}^{M-1} \alpha_{kj} a_j = \beta_k$$

Nonlinear least squares a.k.a Levenberg-Marquardt

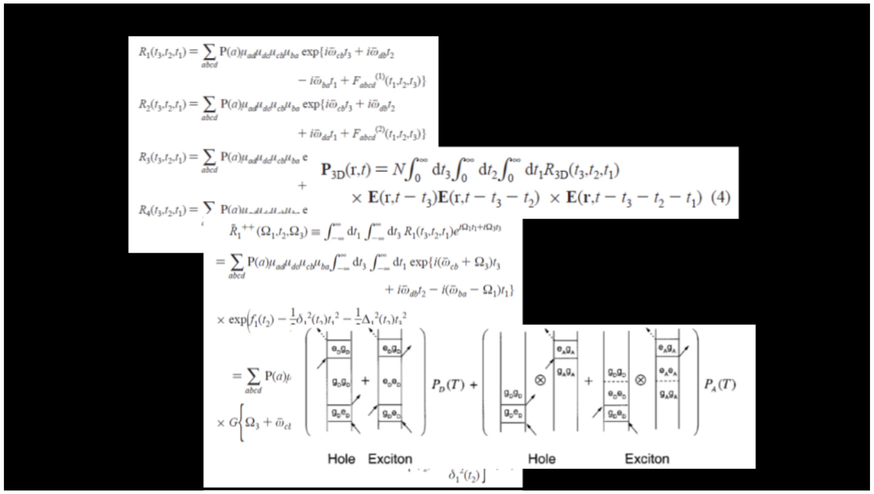
- Guess initial values and look for the closest minimum…

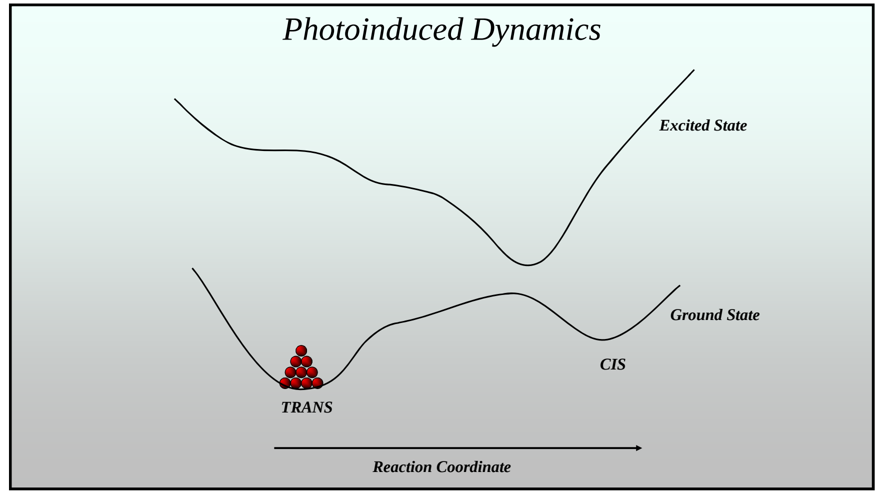
Nonlinear least squares a.k.a Levenberg-Marquardt

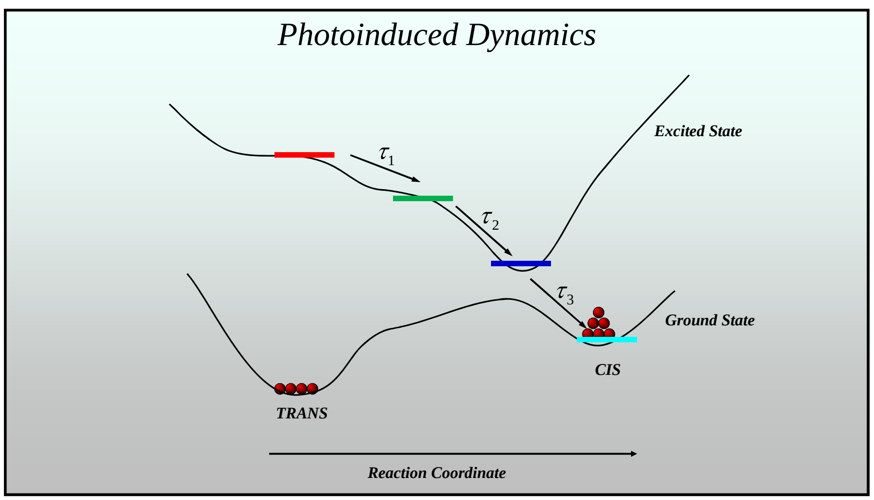
- If the parameter guess is close to the minimum, use Taylor expansion to the quadratic order.
- If guess is bad – go in the direction of steepest descent.
- L-M method is a continuous variation between these two approaches.

Models are reflections of reality in our minds

Phenomenological	Complicated
Intuitive	First principles based
Simplistic	Meaningful
Good description of data	Unintuitive
	Far away from data





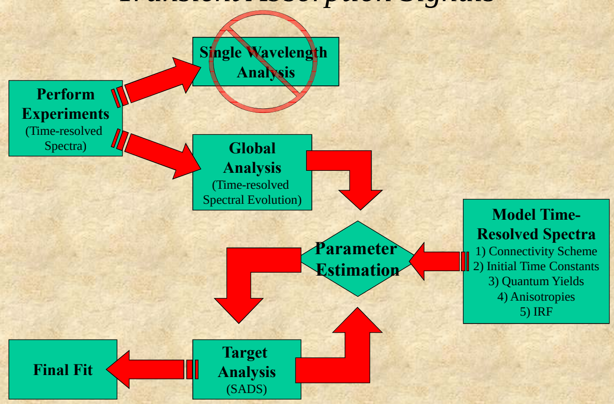


Global analysis is a 'pinball machine' approximation of ultrafast data

Use it when:

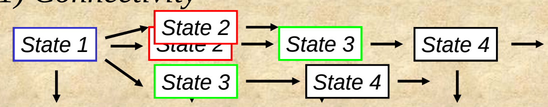
- You do not know any better.
- You need to parametrize large datasets concisely.
- You need to present and interpret the data to people without hardcore physics background.

Hierarchical Modeling of Dispersed Transient Absorption Signals



Three Principle Objectives of Global Analysis

1) Connectivity



2) Timescales

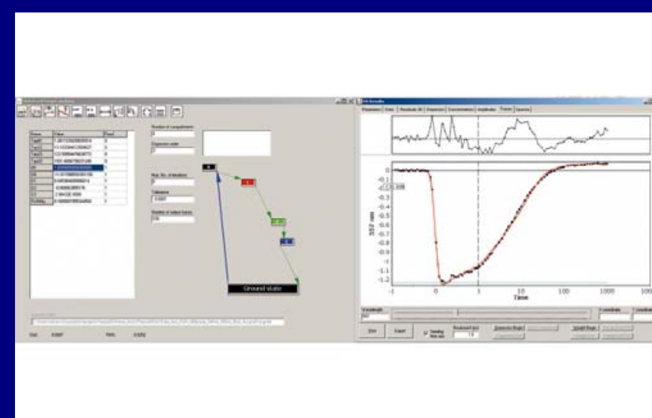
$$\dot{n}_i(t) = \sum_j k_{ij} n_j(t)$$

3) Spectra

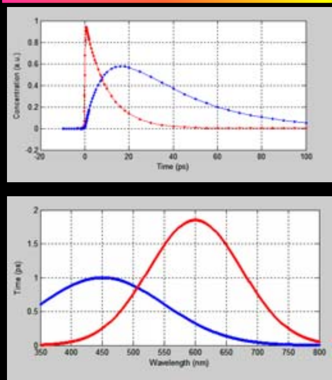
$$D(\lambda, t; k_{ij}) = \sum_i A_i(\lambda) n_i(t; k_{ij})$$

Available options:

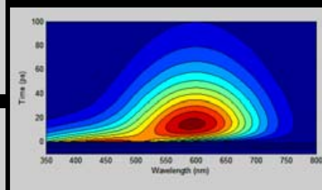
- Glotaran – VU Amsterdam
- CarpetView – LIGHT CONVERSION
- Jasper van Thor's Matlab® based package
- A number of groups have developed their own software



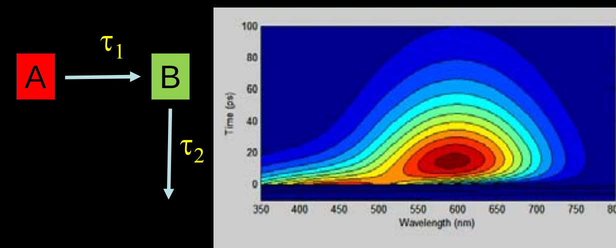
Time-resolved fluorescence dataset (fake data)



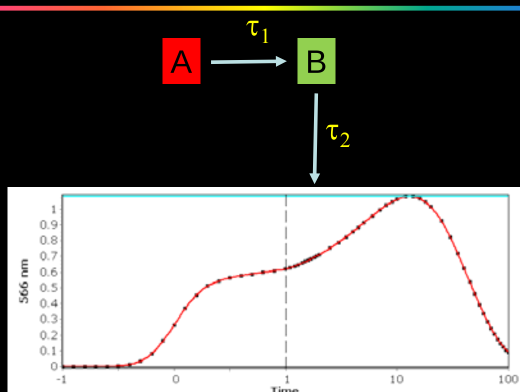
Time constants: 10 and 30 ps, IRF width: 0.2 ps



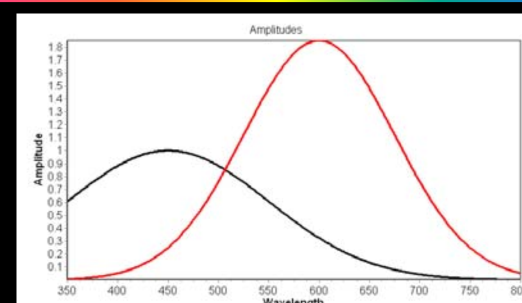
Let's fit it using sequential model…



Let's fit it using sequential model…



Let's fit it using sequential model…

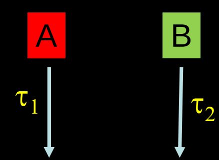


Time constants: 10 and 30 ps, IRF width: 0.2 ps, just what we put in.

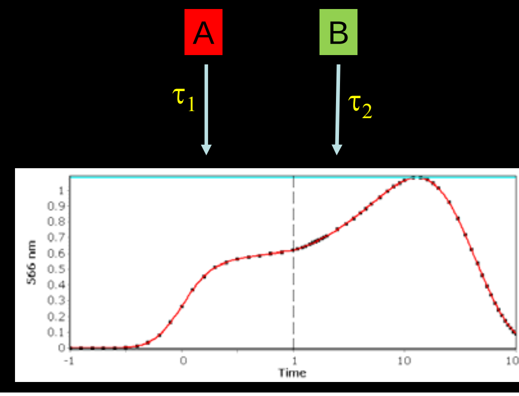
So, we nailed it, right?...

WRONG.

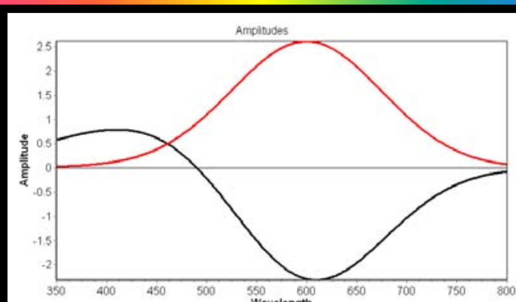
To see why, let's fit it using another model:



The fit is just as good

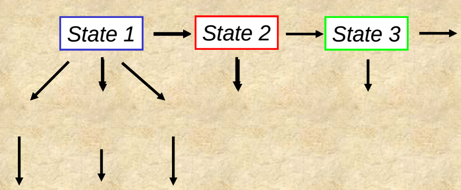


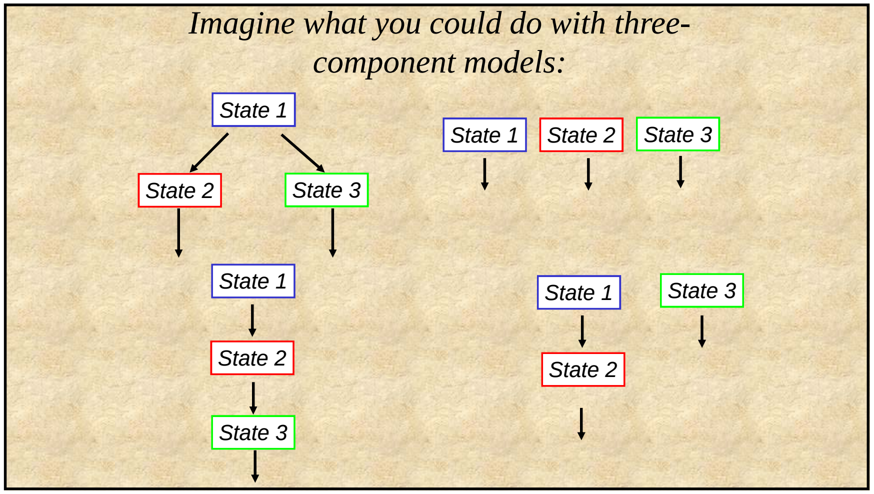
But the component spectra look different:

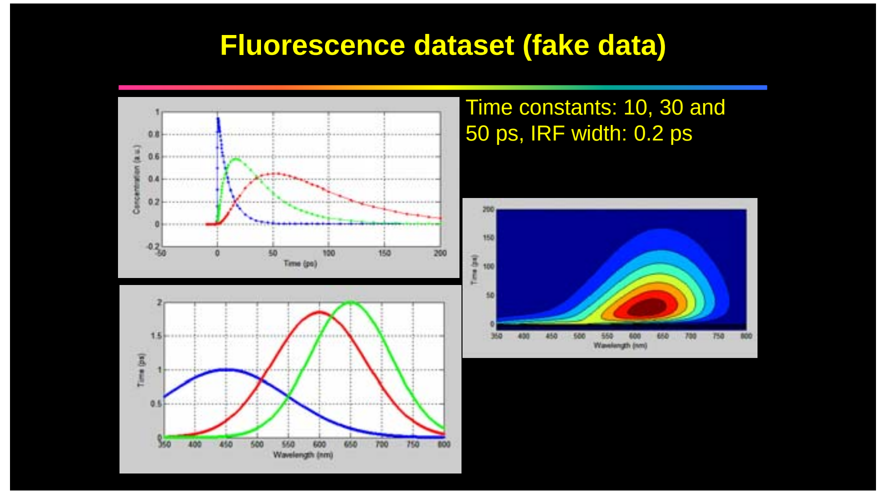


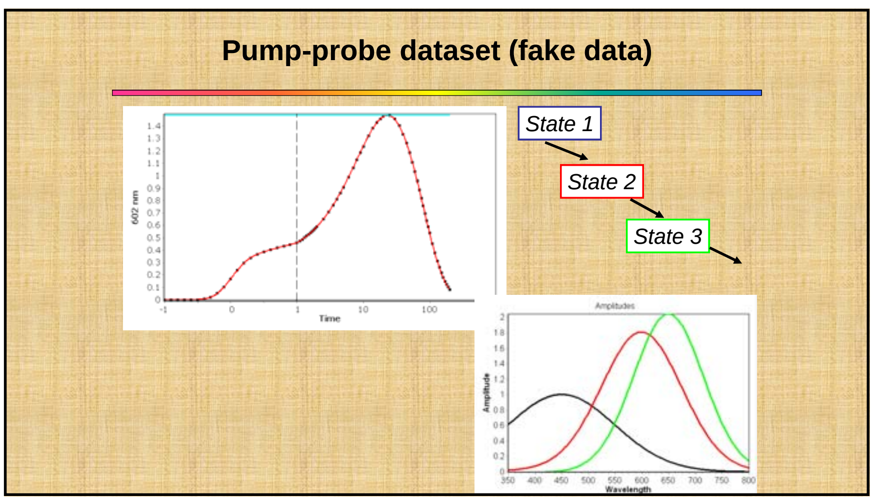
The models represent two different realities, they can't both be correct.

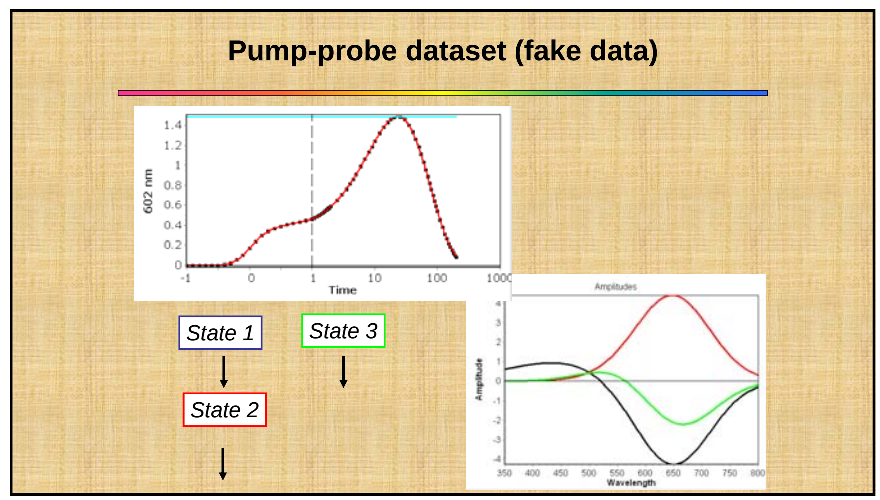
Imagine what you could do with three-component models:




Imagine what you could do with three-component models:
State 1
State 2
State 3
State 1
State 2
State 3
State 1
State 2
State 3
State 1
State 2
State 3


Fluorescence dataset (fake data)
Time constants: 10, 30 and 50 ps, IRF width: 0.2 ps

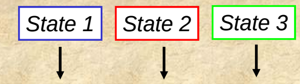

Pump-probe dataset (fake data)
State 1
State 2
State 3
602 nm
Time
Amplitudes
Amplitude
Wavelength


Pump-probe dataset (fake data)
602 nm
Time
State 1
State 3
State 2
Amplitudes
Amplitude
Wavelength

Model degeneracy

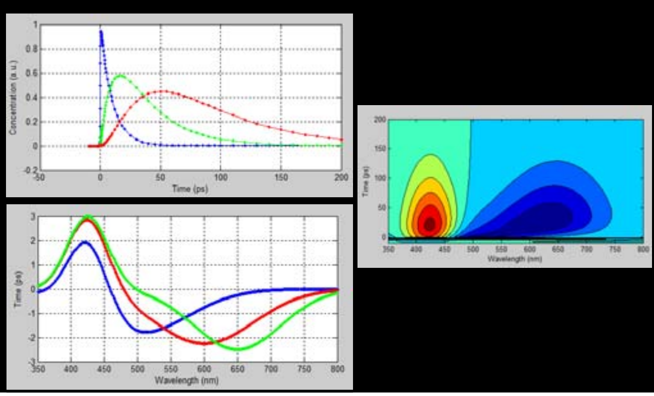
- Any model using connectivity scheme with the same rank (number of different lifetimes observed) will fit the data equally well.
- Besides the quality of the fit, the models have to be judged by the plausibility of component spectra they produce!

Models describing data for parametrization purposes (global analysis): parallel

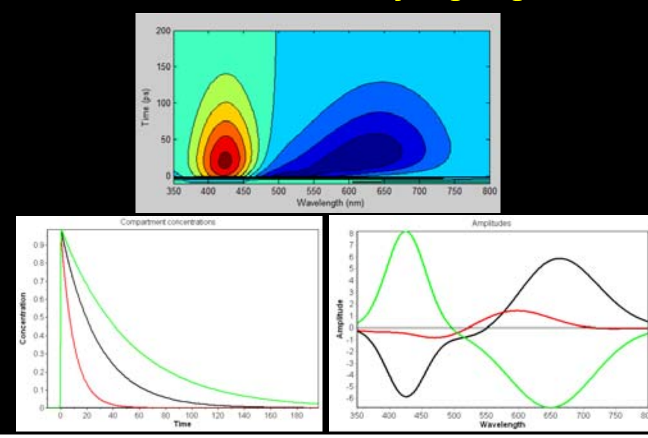


- Independent (parallel) decay model;
- Assumes independent lifetimes for different components;
- Produces *Decay-Associated Difference Spectra*, DADS (in TA) or *Decay-Associated Spectra*, DAS (in fluorescence).
- Negative amplitude means loss of (positive) signal, positive amplitude means gain (growth) of (positive signal.
- What about the signals with varying signs?

Dataset with varying signs (pump-probe)

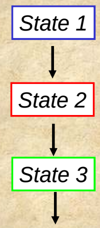


Dataset with varying signs

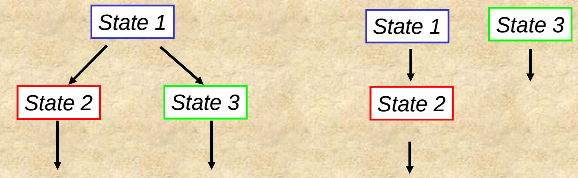


Models describing data for parametrization purposes (global analysis): sequential

- Assumes initial population put in compartment 1, and spectra evolving one into the next.
- Produces *Evolution-Associated Difference Spectra* (EADS).
- Different EADS resemble spectra observed at different times.
- Should be the first model of choice when doing preliminary analysis of TA (and probably fluorescence).



When you start to wonder…



When the different compartments are ascribed physical meanings and connectivity scheme is established using physical assumptions, you are entering the realm of *Target Analysis*.

The resulting spectra with physical meaning are called *Species-Associated Difference Spectra* (SADS)

Build your intuition about SADS:

- Fluorescence SADS should be positive.
- Upon solvation, stimulated emission shifts to the red.
- Ground state SADS are negative only in the GSB region.
- Spectral changes ascribed to different physical processes match your intuition.

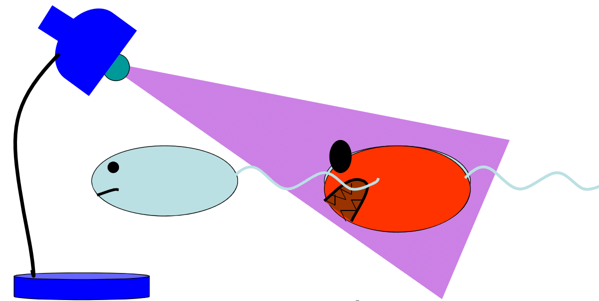
Important to remember:

- Not all kinetics are exponential, but most of what we measure can be depicted as such.
- Worse fit and reasonable spectra is better than good fit with ridiculous spectra

The Photoactive Yellow Protein Structure

- Function: phototaxis photoreceptor in *Halorhodospira halophila*
- Water-soluble protein, suitable for genetic and chemical engineering
- High-resolution structures available (~0.85Å)

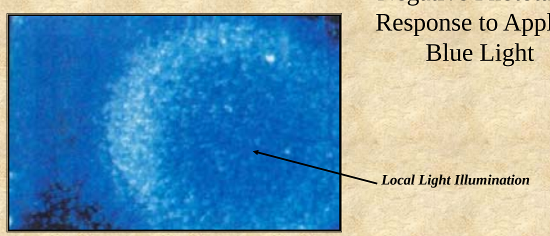
Negative phototactic response to blue light



UnHappy bacterium

Negative Phototactic Response in Halorhodospira halophila :

PYP as the signal Transducer



Negative Phototaxic Response to Applied Blue Light

Hellingwerf and co-workers: J Bacteriol, 1993. 175(10): p. 3096-104.

Case study: solvation+isomerization (live demo)