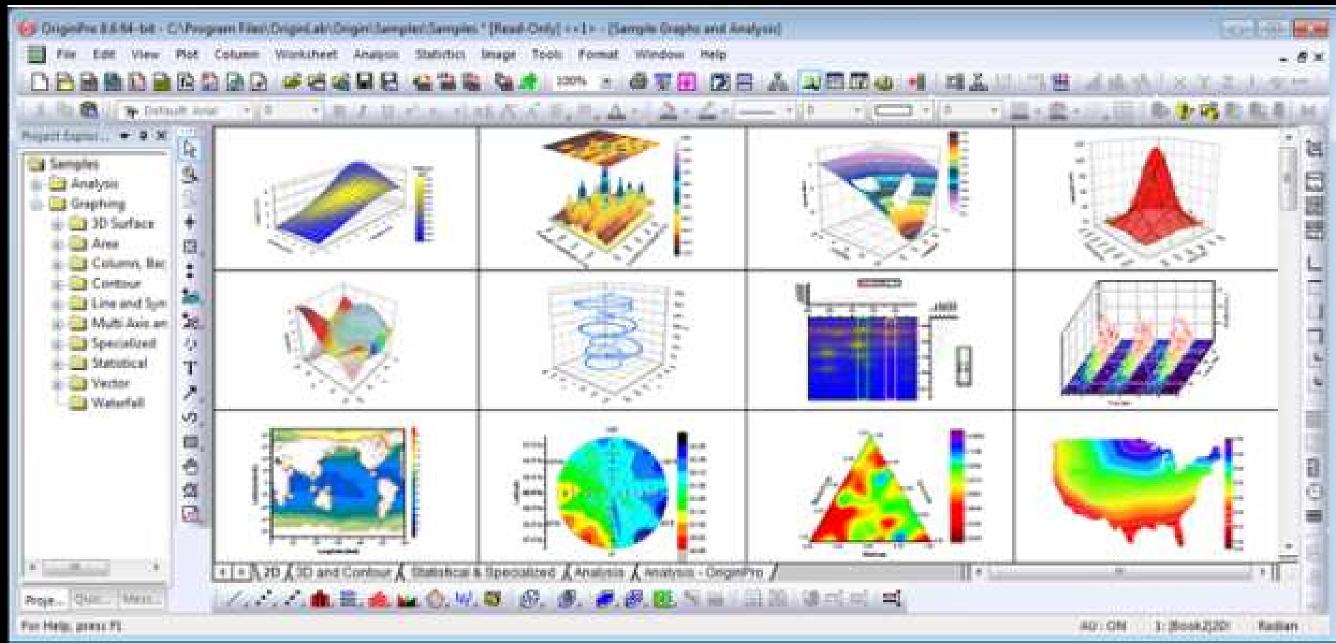
OriginLab Data Analysis and Graphing Software



Základy programovania pre fyzikov

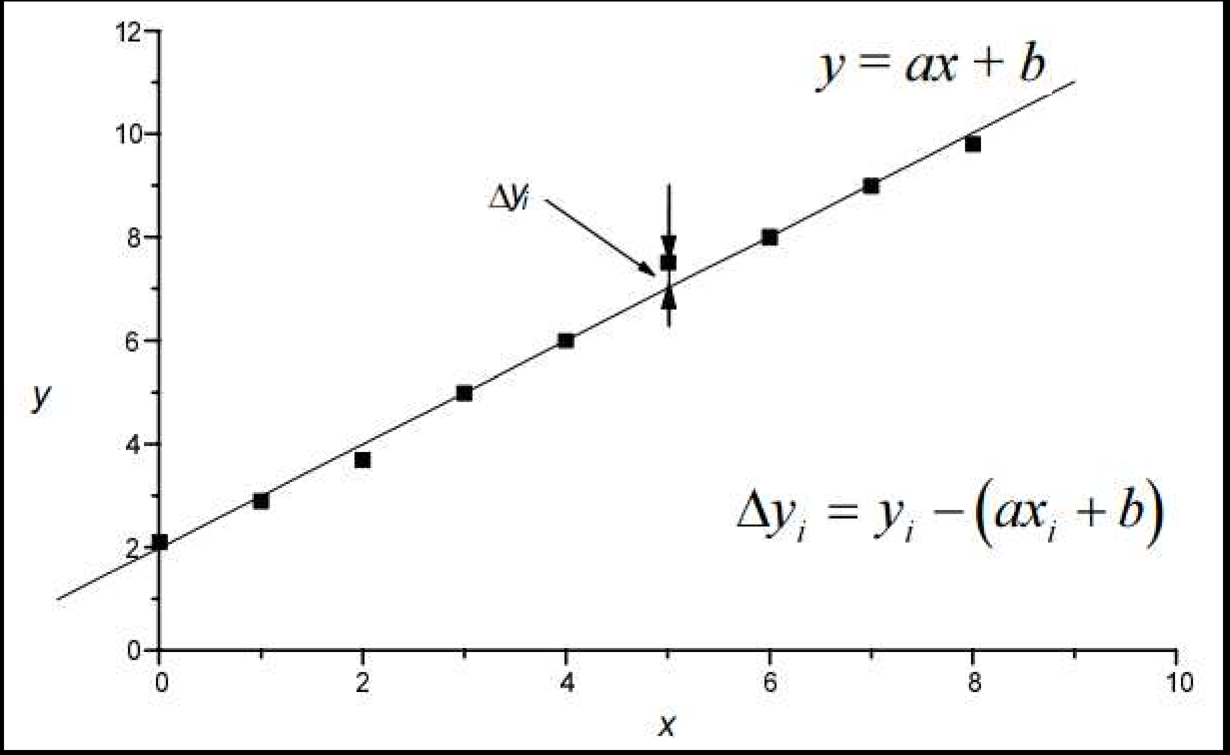
Erik Čižmár

Univerzita P.J. Šafárika

v Košiciach

Základy lineárnej regresie

- Metóda najmenších štvorcov



$$S(a,b) = \sum_{i=1}^{n}\left[y_i - (ax_i + b)\right]^2 = \min$$

Základy lineárnej regresie

- Metóda najmenších štvorcov

$$\frac{\partial S}{\partial a} = 0, \qquad \frac{\partial S}{\partial b} = 0$$

$$\sum_{i=1}^{n} \left\{ 2\left[y_i - (ax_i + b) \right] (-x_i) \right\} = 0$$

$$\sum_{i=1}^{n} \left\{ 2\left[y_i - (ax_i + b) \right] (-1) \right\} = 0$$

$$a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i$$

$$a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i$$

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2}$$

$$b = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2}$$

$$\boxed{b = \frac{\sum_{i=1}^{n} y_i}{n} - a\frac{\sum_{i=1}^{n} x_i}{n} = \bar{y} - a\bar{x}}$$

Základy lineárnej regresie

- Linearizácia funkcií ..... $y(x)$, a, b ......$Y = AX + B$
  - $y = a\, e^{bx}$

    $\ln(y) = \ln(a) + b\,x$ $\quad$ $Y = \ln(y)$, $X = x$ $\quad$ $A = \ln(a)$, $B = b$
  - $y = a\, x^{b}$

    $\ln(y) = \ln(a) + b\ln(x)$ $\quad$ $Y = \ln(y)$, $X = \ln(x)$ $\quad$ $A = \ln(a)$, $B = b$
  - $y = a\, b^{x}$, $b > 0$, $b \neq 1$

    $\ln(y) = \ln(a) + \ln(b)\,x$ $\quad$ $Y = \ln(y)$, $X = x$ $\quad$ $A = \ln(a)$, $B = \ln(b)$
  - $y = a + b/x$ $\quad$ $Y = y$, $X = 1/x$ $\quad$ $A = a$, $B = b$
  - $y = 1/(a + bx)$ $\quad$ $Y = 1/y$, $X = x$ $\quad$ $A = a$, $B = b$

Kvadratická regresia

$$y_i = a_0 + a_1 x_i + a_2 x_i^2 \qquad i = 1, \ldots, n$$

$$\frac{\partial S}{\partial a_i} = 0, \qquad i = 1, \ldots, n$$

$$\sum_{i=1}^{n} y_i = a_2 \sum_{i=1}^{n} x_i^2 + a_1 \sum_{i=1}^{n} x_i + n a_0$$

$$\sum_{i=1}^{n} x_i y_i = a_2 \sum_{i=1}^{n} x_i^3 + a_1 \sum_{i=1}^{n} x_i^2 + a_0 \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i^2 y_i = a_2 \sum_{i=1}^{n} x_i^4 + a_1 \sum_{i=1}^{n} x_i^3 + a_0 \sum_{i=1}^{n} x_i^2$$

Všeobecný tvar lineárnej regresie

$$f(x) = c_1 f_1(x) + c_2 f_2(x) + \ldots + c_k f_k(x)$$

$y = ax + b$ $\quad$ $f_1(x) = x$ a $f_2(x) = 1$

$y = a/x + bx^2$ $\quad$ $f_1(x) = 1/x$ a $f_2(x) = x^2$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix} p_1 & 0 & \ldots & 0 \\ 0 & p_2 & \ldots & 0 \\ \vdots & & & \\ 0 & 0 & \ldots & p_n \end{pmatrix}$$

$$\mathbf{X} = \begin{pmatrix} f_1(x_1) & f_2(x_1) & \ldots & f_k(x_1) \\ f_1(x_2) & f_2(x_2) & \ldots & f_k(x_2) \\ \vdots & & & \\ f_1(x_n) & f_2(x_n) & \ldots & f_k(x_n) \end{pmatrix}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} f_1(x_1) & f_2(x_1) & \ldots & f_k(x_1) \\ f_1(x_2) & f_2(x_2) & \ldots & f_k(x_2) \\ \vdots & & & \\ f_1(x_n) & f_2(x_n) & \ldots & f_k(x_n) \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$$

p^i - váha meraného bodu

$\Rightarrow$ $\boldsymbol{y = Xc}$

Všeobecný tvar lineárnej regresie

$$S(c_1, c_2, \dots, c_k) = \sum_{i=1}^{n} \left\{ p_i \left[y_i - \left(x_{i1}c_1 + x_{i2}c_2 + \dots + x_{jk}c_k \right) \right]^2 \right\} = \min$$

$$\frac{\partial S}{\partial c_1} = 2\sum_{i=1}^{n} \left\{ p_i \left[y_i - x_{i1}c_1 - x_{i2}c_2 - \dots - x_{ik}c_k \right]^2 (-x_{i1}) \right\} = 0$$

$$\frac{\partial S}{\partial c_2} = 2\sum_{i=1}^{n} \left\{ p_i \left[y_i - x_{i1}c_1 - x_{i2}c_2 - \dots - x_{ik}c_k \right]^2 (-x_{i2}) \right\} = 0$$

$$\vdots$$

$$\frac{\partial S}{\partial c_k} = 2\sum_{i=1}^{n} \left\{ p_i \left[y_i - x_{i1}c_1 - x_{i2}c_2 - \dots - x_{ik}c_k \right]^2 (-x_{ik}) \right\} = 0$$

$$c_1 \sum_{i=1}^{n} x_{i1}p_i x_{i1} + c_2 \sum_{i=1}^{n} x_{i1}p_i x_{i2} + \dots + c_k \sum_{i=1}^{n} x_{i1}p_i x_{ik} = \sum_{i=1}^{n} x_{i1}p_i y_i$$

$$c_1 \sum_{i=1}^{n} x_{i2}p_i x_{i1} + c_2 \sum_{i=1}^{n} x_{i2}p_i x_{i2} + \dots + c_k \sum_{i=1}^{n} x_{i2}p_i x_{ik} = \sum_{i=1}^{n} x_{i2}p_i y_i$$

$$\vdots$$

$$c_1 \sum_{i=1}^{n} x_{ik}p_i x_{i1} + c_2 \sum_{i=1}^{n} x_{ik}p_i x_{i2} + \dots + c_k \sum_{i=1}^{n} x_{ik}p_i x_{ik} = \sum_{i=1}^{n} x_{ik}p_i y_i$$

$$\left(\boldsymbol{X}^T \boldsymbol{P} \boldsymbol{X} \right) \boldsymbol{c} = \boldsymbol{X}^T \boldsymbol{P} \boldsymbol{y}$$

$$\Downarrow$$

$$\boxed{\boldsymbol{c} = \left(\boldsymbol{X}^T \boldsymbol{P} \boldsymbol{X} \right)^{-1} \boldsymbol{X}^T \boldsymbol{P} \boldsymbol{y}}$$

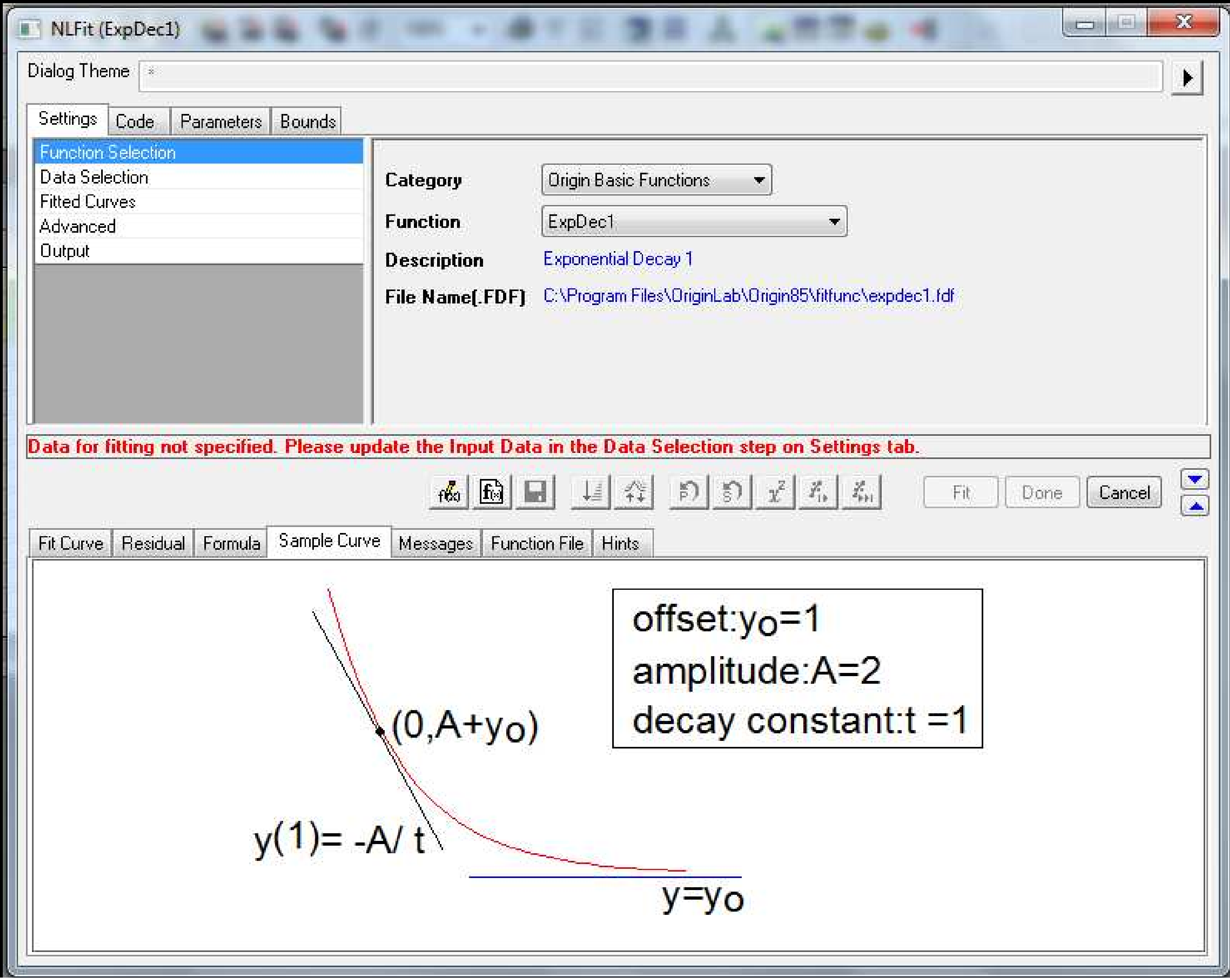
Nelineárna regresia

- Ak neviem funkciu, ktorá popisuje moje dáta popísať ako lineárnu kombináciu bázových funkcií (ktoré nezávisia od hľadaných parametrov), neviem priamo spočítať hľadané parametre ako pri lineárnej regresii.
- Pre nájdenie minima sumy štvorcov následne používam iteratívne metódy – postupne sa mení sada hľadaných parametrov a na základe zmeny sumy štvorcov v ďalšom kroku zmením parametre
  - suma štvorcov je plocha v *n*-rozmernom priestore parametrov (ak mám *n* hľadaných parametrov) - podľa zmeny sklonu tejto plochy iterujem k parametrom zodpovedajúcim minimu
  - Levenberg-Marquardt
  - Simplex

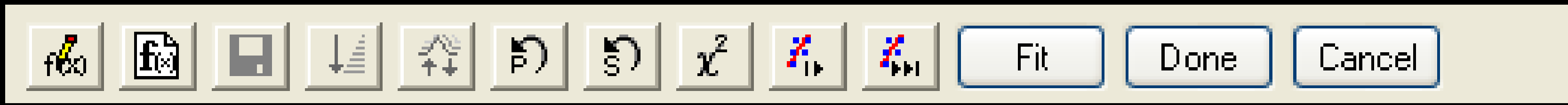
Fitovanie v Origine

- **Linear Fit**
- **Polynomial Fit**
- **Multiple Linear Regression – sada dát, ktorá je popísaná jedným modelom so spoločným parametrom – napr. poľná závislosť magnetizácie magnetickej látky zmeraná pri rôznych teplotách, kde spoločným parametrom je výmenná interakcia medzi magnetickými momentmi**
- **Nonlinear Curve Fit – nelineárna regresia funkciou jednej alebo viacerých premenných, fitovanie fukcií komplexných premenných**

Fitovanie v Origine



Fitovanie v Origine



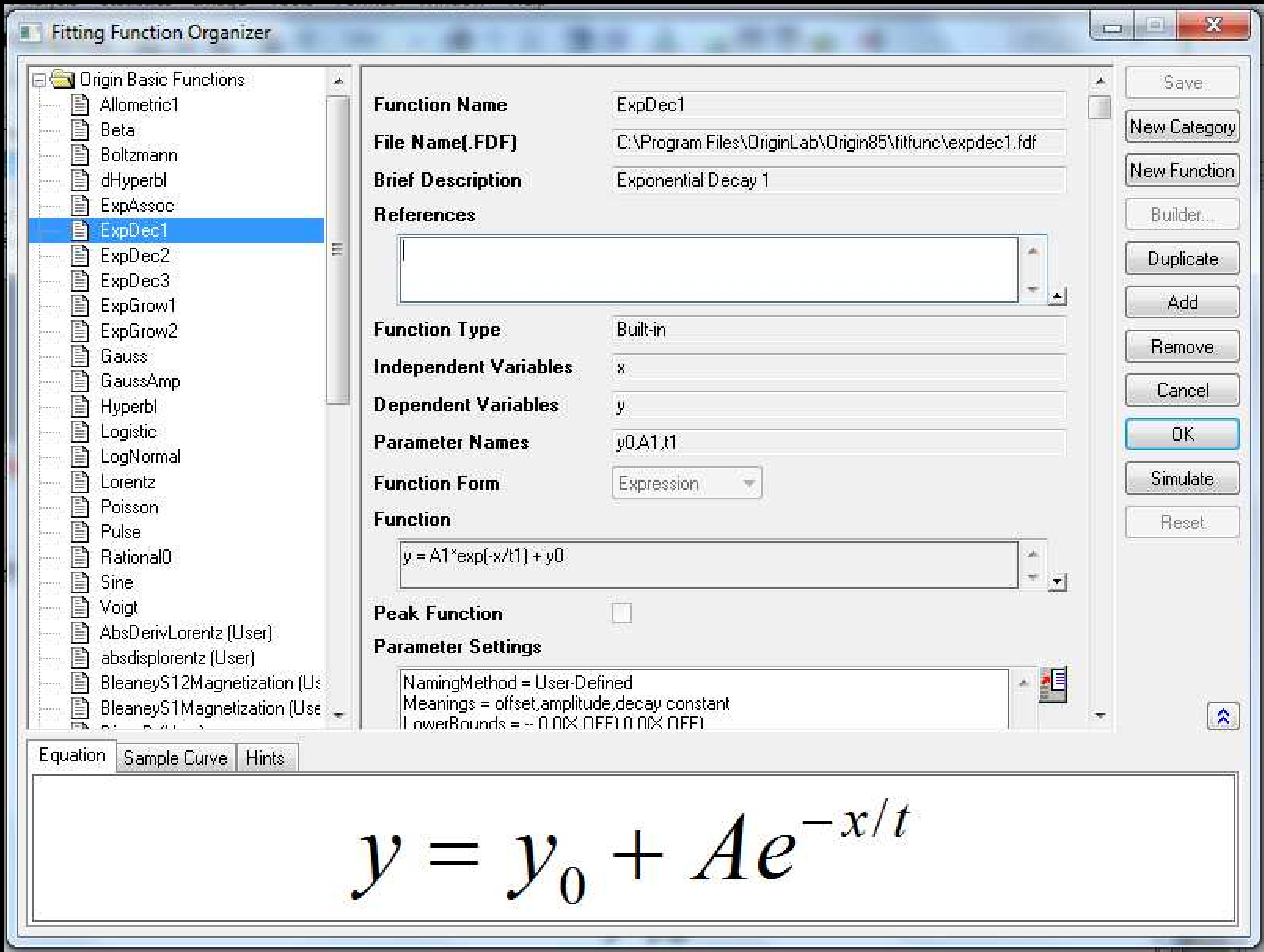
	Edit Fitting Functions
	Create New Fitting Functions
	Save FDF File

χ^2	Calculate Chi-Square
	1 Iteration
	Fit until converged

	Initialize Parameters

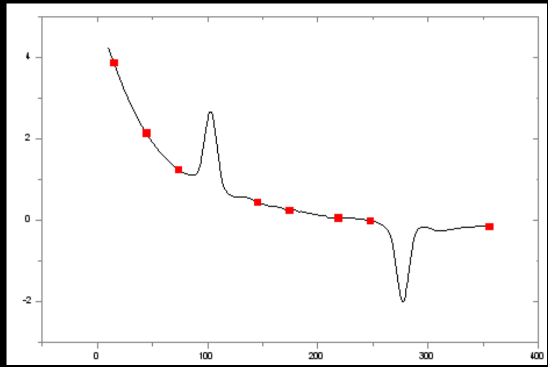
	Simplex

Editovanie funkcií
- Fitting Function Organizer



Fitovanie píkových závislostí

- Spektroskopické merania jeden alebo viac maxím - píkov
- Funkcie typu Gaussián, Lorentzián, Voigth, Laplace ....
- Baseline – signál pozadia



- Určenie polohy píkov
- Plocha pod píkom – intenzita signálu – aké „množstvo“ daného fyzikálneho javu sme pozorovali – relatívne, absolútne
- Polšírka píku – ako veľmi je pozorovaný fyzikálny proces „rozmazaný“ - napr. pri magnetickej rezoznancii nie všetky častice rezonujú pri rovnako veľkom magnetickom poli
- Single Peak Fit alebo Multiple Peak Fit

Fitovanie píkových závislostí

- Peak Analyzer
  - Baseline – Constant, User Defined , Dataset
  - Create Baseline – interpolácia
  - Baseline Treatment – odpočítanie
  - Find Peaks – maximum, 1. derivácia, 2.derivácia
  - Intergrate peaks – intenzita, polšírka