Non-linear Regression Analysis with Fitter Software Application

Alexey Pomerantsev

Semenov Institute of Chemical Physics
Russian Chemometrics Society

Agenda

1. Introduction
2. TGA Example
3. NLR Basics
4. Multicollinearity
5. Prediction
6. Testing
7. Bayesian Estimation
8. Conclusions

1. Introduction

Linear and Non-linear Regressions

	Linear	Non Linear
Formula	$f = a_1\varphi_1(x) + \dots + a_p\varphi_p(x)$	any $f(\boldsymbol{a},\boldsymbol{x})$
Example	$f = a\exp(-20x)$	$f = \exp(-ax)$
Source	Soft	Hard
Choice	Easy ?	Difficult ?
Dimension	Large	Small
Multicollinearity	Excess of parameters	Lack of data
Interpretation	Well-known	Uncommon
Purpose	Interpolation	Extrapolation
Soft Tools	Many	Few

Close relatives?

2. Thermo Gravimetric Analysis Example

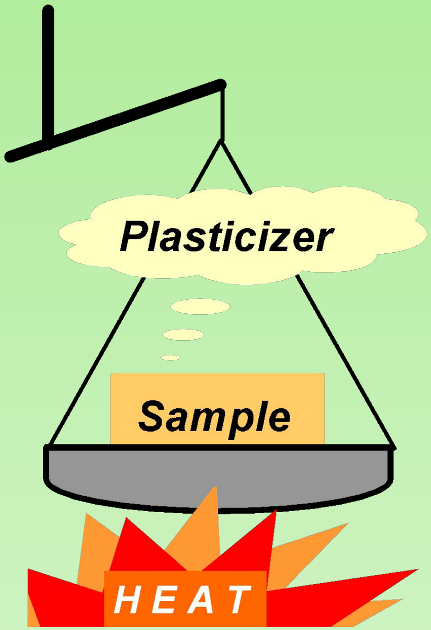
Object	**PVC Cable Isolation**
Goal	**Service-Life Prediction**
Experiment	**Thermo Gravimetric Method**
Tool	**Non-Linear Regression and Fitter**

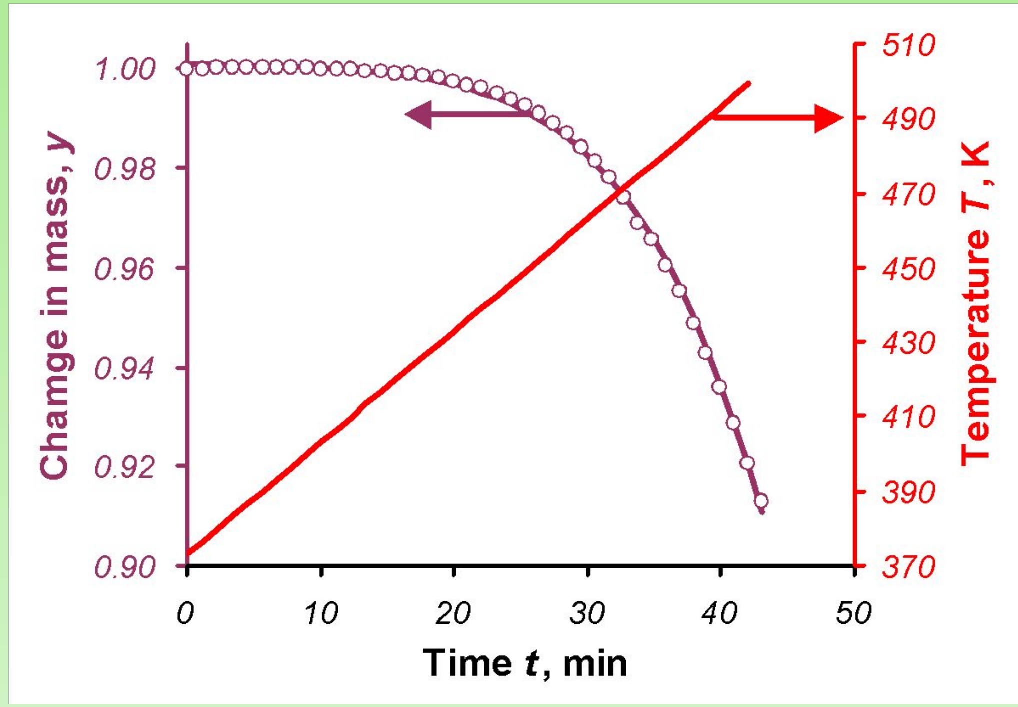
Let's see it!

TGA Experiment and Data

TGA Experiment

TGA Data





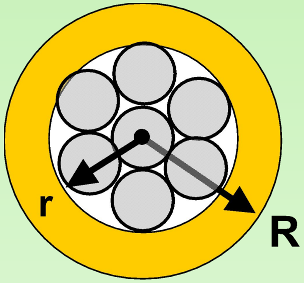
This is Experiment! Not a Hell of Flame!

TGA Example Variables

Measured		Estimated	
Response		**Intermediate**	
$y = m/m_0$	Change in mass	C	Plasticizer concentration
Predictors		**Parameters**	
t	Time	y_0	Initial value of y
C_0	Initial concentration	k	Evaporation rate constant
v	Heating rate	E	Activation energy
T_0	Initial temperature		
F	Sample specific surface		

Small size problem!

Sample specific surface $F = \frac{S}{V} = \frac{2R}{R^2 - r^2}$



Plasticizer Evaporation Model

Evaporation Law

$$\frac{dy}{dt} = -kC, \quad y(0) = y_0$$

Volume Change

$$C = 1 - \frac{1 - C_0}{y}$$

Diffusion is not relevant!

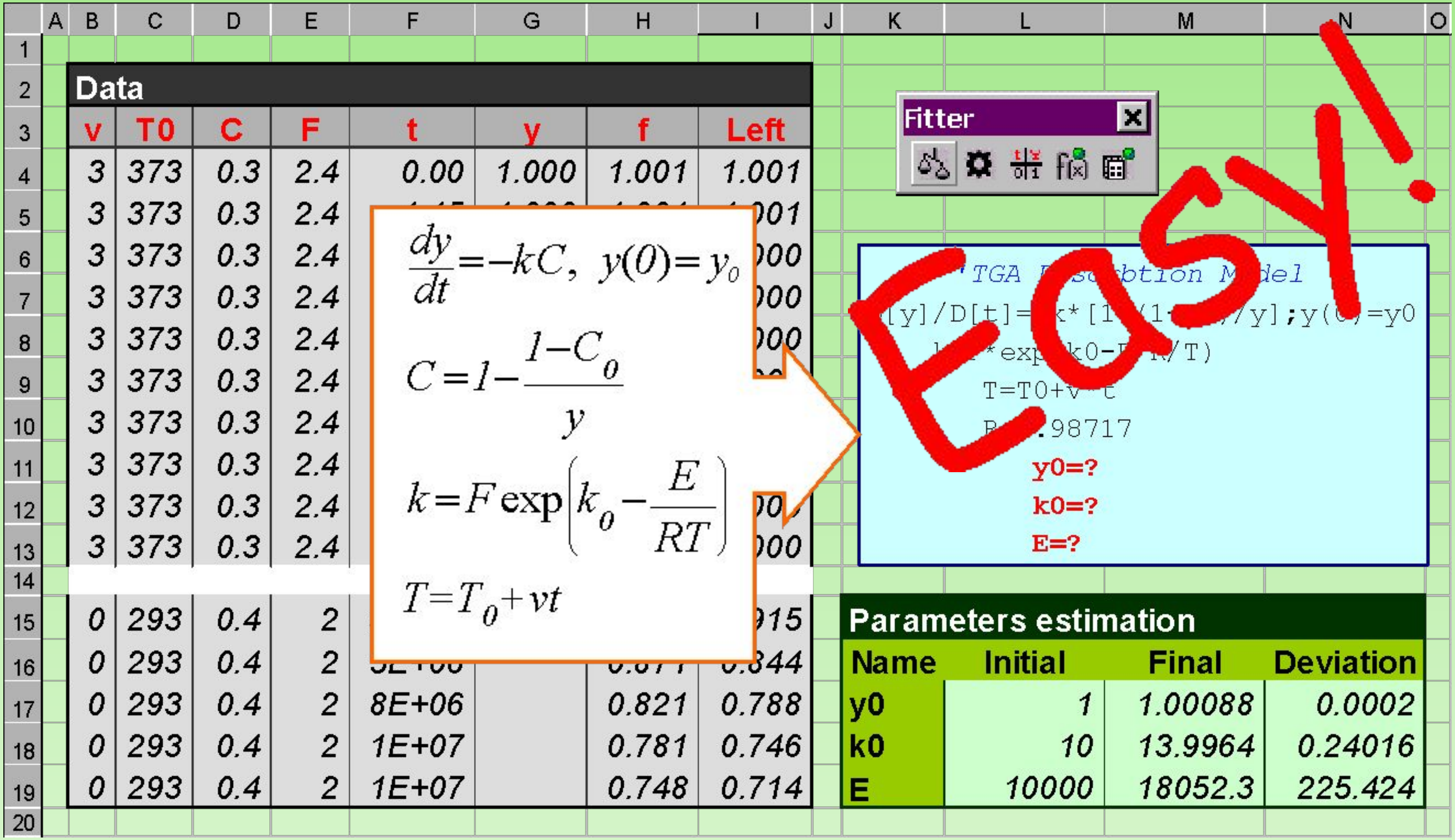
The Arrhenius law

$$k = F\exp\left(k_0 - \frac{E}{RT}\right)$$

Temperature growth

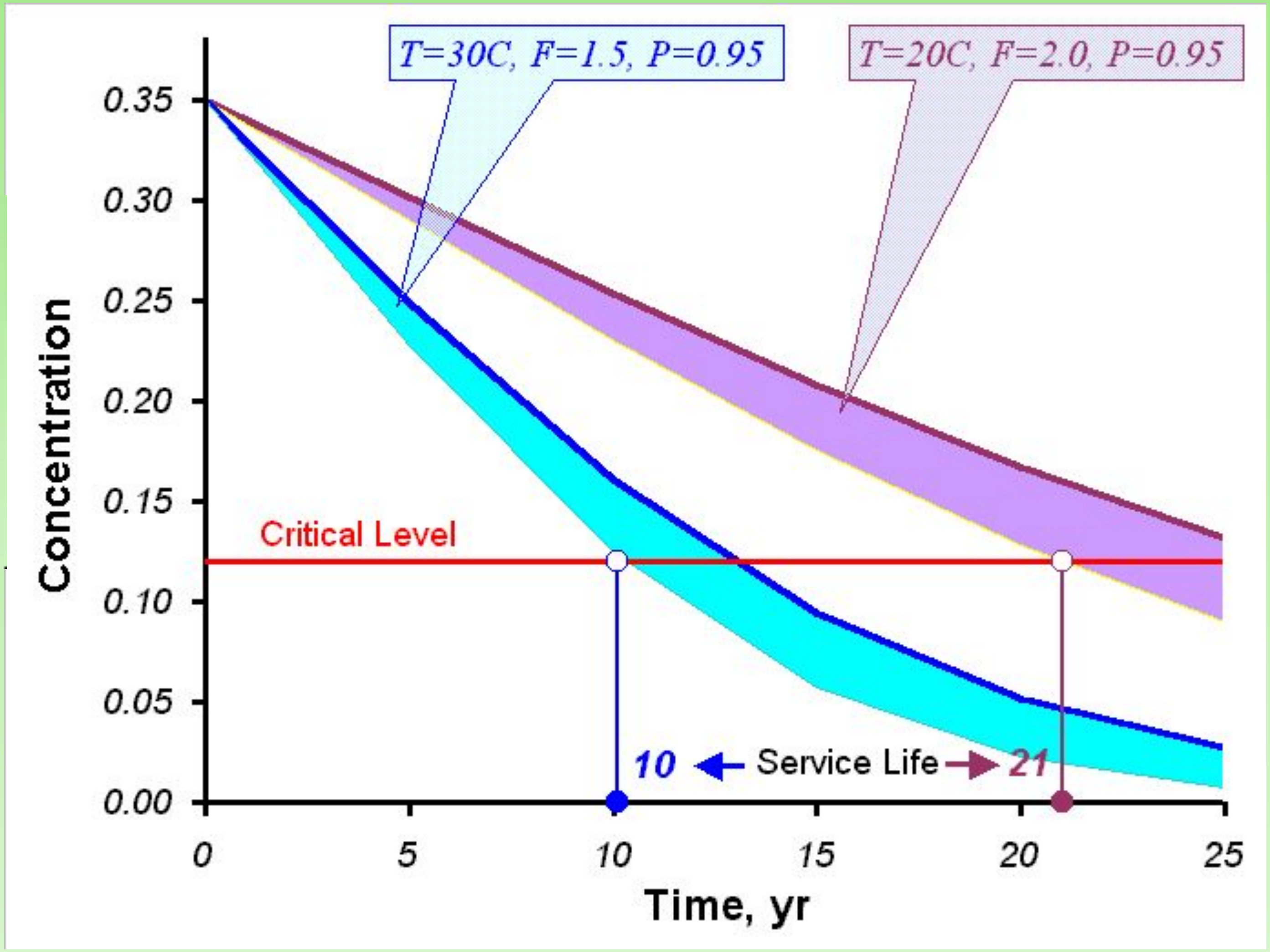
$$T = T_0 + vt$$

Fitter Worksheet for TGA Example



Service Life Prediction by TGA Data

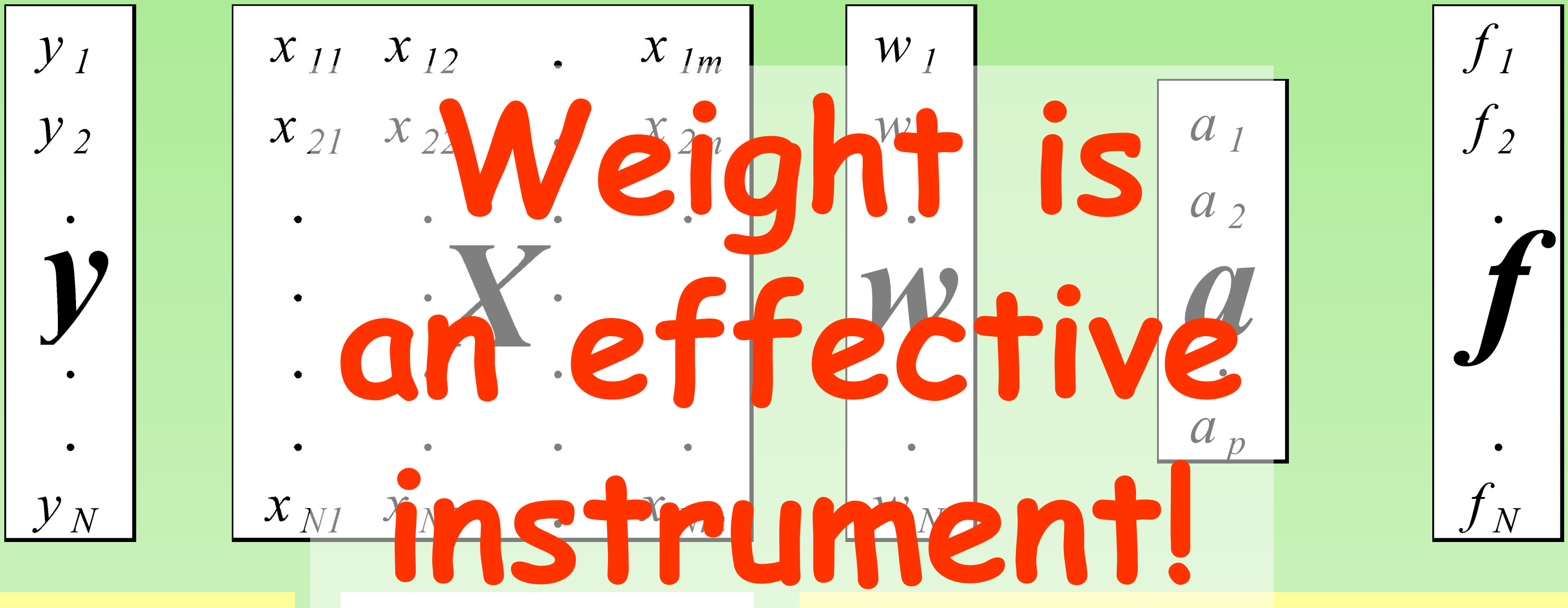
By Science
not
by guess!



3. NLR Basics

Data and Errors

Response	Predictors	Weights	Parameters	Fit



Absolute error	$y_i = f_i + \varepsilon_i$	Weight and variance	
Relative error	$y_i = f_i(1 + \varepsilon_i)$	$w_i^2 = \mathrm{cov}(\varepsilon_i, \varepsilon_i) = \sigma_\varepsilon^2 = \mathrm{Const}$	

Model $f(x,$

Explicit

Implicit

Diff. eq

$= b$

```
'  Explic
y=a+(b-a)
y(0)=b
```

```
'Цикл "увлажнение-сушка"
M=Sor*hev(t1-t)+Des*[hev(t-t1)+imp(t-t1)]
'Кинетика "увлажнения"
    Sor=Sor1*hev(USESor1)+Sor2*[hev(-USESor1)+imp(-USESor1)]
'Кинетика "сушки"
    Des=Des1*hev(USEDes1)+Des2*[hev(-USEDes1)+imp(-USEDes1)]
'Условие применимости асимптотик
    USESor1=Sor2-Sor1
    USEDes1=Des1-Des2
'константы и промежуточные величины
    t3=(t-t1)*hev(t-t1)
    t4=t*hev(t1-t)+t1*[hev(t-t1)+imp(t-t1)]
    P2=PI*PI
    P12=(PI)^(-0.5)
    R=r*(M1-M0)*exp(-r*t4)
    K=M1+(M0-M1)*exp(-r*t4)
    V0=M0-C0
    V1=M1-C0
'асимпто[illegible] сорбции при 0<[illegible]<tau
    Sor[illegible]C0+[illegible]*P12*(d*t[illegible]0.5[illegible][M0-C0+(M1-M0)*beta]
        b[illegible]ta=[illegible]e[illegible]([illegible])
        x[illegible]
        [illegible](a[illegible]x+[illegible]2*[illegible]x+a[illegible]*x*[illegible]*x)[illegible]([illegible]+b1*x[illegible]2*x*x+b3*x*x*x)
            a1=0.[illegible]
            a2=0.012105101749
            a3=0.009922532428
            b1=0.08480062319
            b2=0.024663459123
            b3=0.00[illegible]75499[illegible]47958
'кине[illegible]
   So[illegible]2=K-[illegible]S1
        S[illegible]
            n0=P2
            U01=[(V0*n0*d-V1[illegible]r)*exp(-n0*d*t4)+R]/(n0*d-r)
            n1=P2*9
            U11=[(V0*n1*d-V1*r)[illegible]exp(-n1*[illegible])+R]/(n1*d-r)
            n2=P2*25
            U21=[(V0*n2*d-V1*[illegible]exp(-n2*[illegible])+R]/(n2*d-r)
            [illegible]9
            [illegible]3[illegible][[illegible]n3*[illegible]1*r[illegible]n3*[illegible])+R]/(n3*d-r)
            [illegible]4=[illegible]2*[illegible]1
            U41[illegible][([illegible]*n4*d-V1*r)*exp(-n4*d*t4)+R]/(n4*d-r)
'асимптотика десорбции при t1<t<t1+tau
    Des1=K*[1-4*P12*(d*t3)^0.5]-8*S1
'кинетика десорбции при t1+tau<t
    Des2=8*S2
        S2=U02/n0+U12/n1+U22/n2+U32/n3+U42/n4
            U02=(K-U01)*exp(-n0*d*t3)
            U12=(K-U11)*exp(-n1*d*t3)
            U22=(K-U21)*exp(-n2*d*t3)
            U32=(K-U31)*exp(-n3*d*t3)
            U42=(K-U41)*exp(-n4*d*t3)
'неизвестные параметры
            d=?
            M0=?
            M1=?
            C0=?
            r=?
            t1=?
```

Rather complex model!

Data & Model Prepared for Fitter

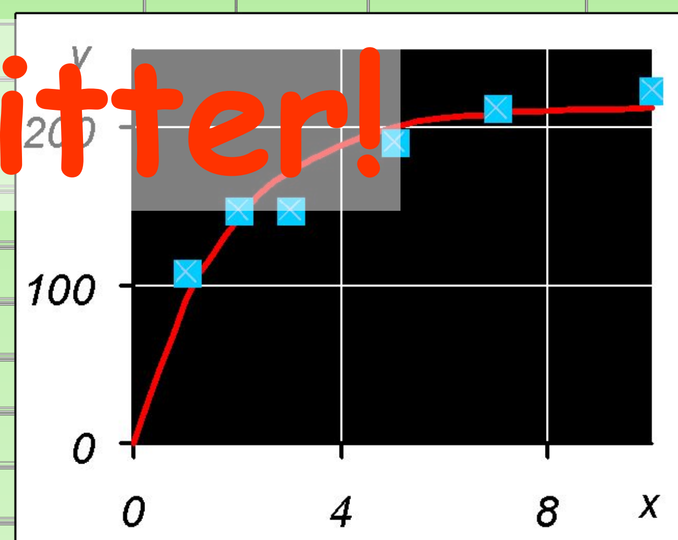
Predictor | Response | Weight | Values | Comment

BoxBod Data			
x	y	w	f
0		0	0.00
1	109	1	90.11
2	149	1	142.24
3	149	1	172.41
5	191	1	199.95
7	213	1	209.17
10	224	1	212.91

Parameters	
a	213.80941
b	0.5472375

Apply Fitter!

```
'BoxBOD model
y=a*[1-exp(-b*x)]
a=?
b=?
```



Fitting | Parameters | Equation

Objective Function $Q(\boldsymbol{a})$

Sum of squares	$S(a) = \sum_{i=1}^{N} w_i^2 (y_i - f_i)^2 g_i^2$
Bayesian term	$B(\boldsymbol{a}) = s_0^2 \left[N_0 + (\boldsymbol{a}-\boldsymbol{b})^t H (\boldsymbol{a}-\boldsymbol{b}) \right]$
Objective function	$Q(\boldsymbol{a}) = S(\boldsymbol{a}) + B(\boldsymbol{a})$

Objective function Q is a sum of squares and may be more…

Parameter estimates

$$\hat{\boldsymbol{a}} = \arg\min \; Q(\boldsymbol{a})$$

Weighted variance estimate

$$s^2 = \frac{S(\hat{\boldsymbol{a}})}{N_f} \qquad N_f = N - p$$

Very Important Matrix A

Hesse's matrix	$A_{\alpha\beta} = \frac{1}{2}\frac{\partial^2 Q(\boldsymbol{a})}{\partial a_\alpha \partial a_\beta}, \quad \alpha, \beta = 1, \ldots, p$
Gauss' approximation	$\boldsymbol{A} \approx \boldsymbol{V}^t \boldsymbol{V}$ ($\boldsymbol{X}^t \boldsymbol{X}$ in linear regression)
	$V_{\alpha i} = w_i \frac{\partial f(\boldsymbol{x}_i, \boldsymbol{a})}{\partial a_\alpha} \quad \alpha = 1, \ldots, p; \quad i = 1, \ldots, N$
Model derivatives	
	$\boldsymbol{C} = s^2 \boldsymbol{A}^{-1} = \boldsymbol{F}^{-1}$
Covariance matrix	$\boldsymbol{F} = s^{-2} \boldsymbol{A} = \boldsymbol{C}^{-1}$
F matrix	

Matrix A is the cause of troubles..

Quality of Estimation

Covariances Matrix	$\operatorname{cov}(\hat{\boldsymbol{a}}, \hat{\boldsymbol{a}}) = \boldsymbol{C} = s^2 \boldsymbol{A}^{-1}$
Deviations Vector	$\operatorname{dev}(\hat{a}_\alpha) = \sqrt{C_{\alpha\alpha}}$
Correlations Matrix	$\operatorname{cor}(\hat{a}_\alpha, \hat{a}_\beta) = \dfrac{C_{\alpha\beta}}{\sqrt{C_{\alpha\alpha} C_{\beta\beta}}}$
Final Objective Value	$Q(\hat{\boldsymbol{a}}) = \min Q(\boldsymbol{a})$
Error Variance and Number Degrees of Freedom	$s^2 = \dfrac{S(\hat{\boldsymbol{a}})}{N_f}$ $\quad N_f = N - p$

Matrix A is the measure of quality!

Search by Gradient Method

$$Q(\boldsymbol{a}) \approx Q(\boldsymbol{a}_n) + \boldsymbol{b}_n^t(\boldsymbol{a} - \boldsymbol{a}_n) + \frac{1}{2}(\boldsymbol{a} - \boldsymbol{a}_n)^t A(\boldsymbol{a} - \boldsymbol{a}_n)$$

$$\boldsymbol{a}_{n+1} = \boldsymbol{a}_n + A^+\boldsymbol{b}_n$$

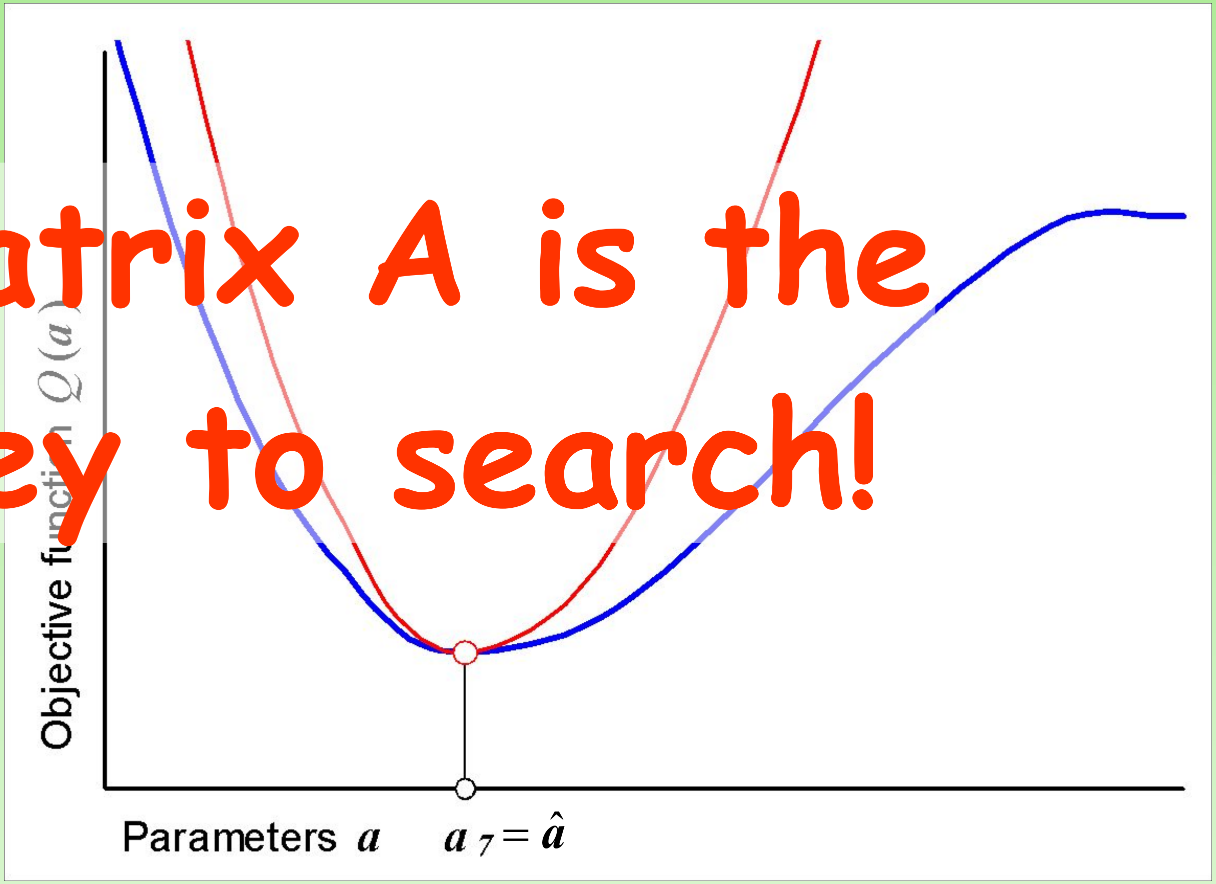
$$A^+A = I$$

Search problems

Initial point $\boldsymbol{a}_0$

$$\det(A) \approx 0$$

Local minima



Matrix A is the key to search!

4. Multicollinearity

Multicollinearity: View

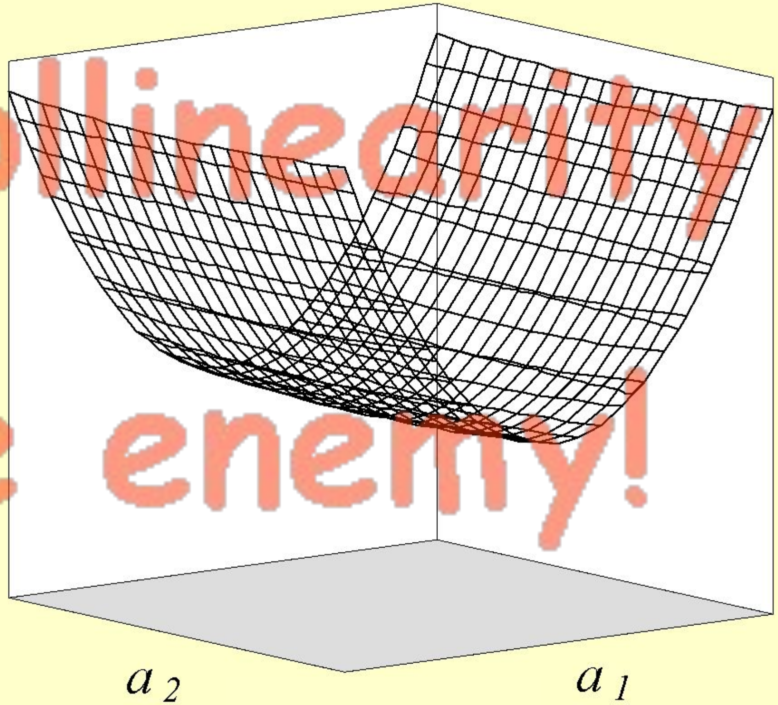
Multicollinearity is degradation of matrix $\boldsymbol{A}$

Spread of eigenvalues

$$\mathrm{N}(A) = \log_{10}\left(\frac{\lambda_{\max}}{\lambda_{\min}}\right)$$

is a measure of degradation

Objective function $Q(\boldsymbol{a})$

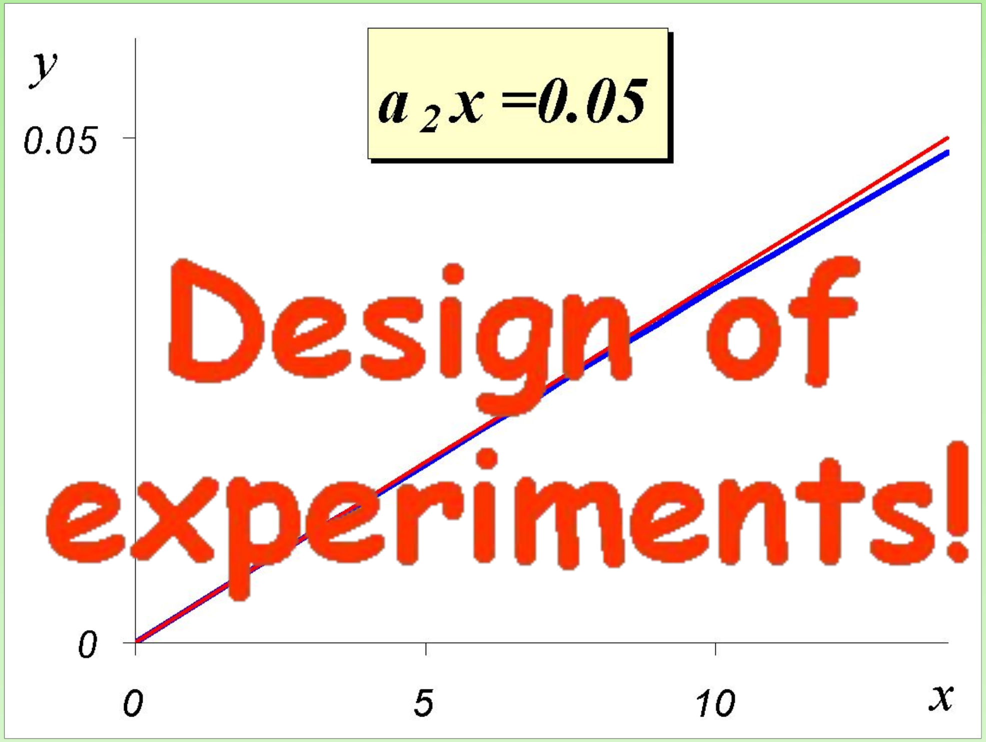


$\mathbf{N}(A) = 7$

Multicollinearity is the enemy!

Multicollinearity: Source

“Hard” multicollinearity	“Soft” multicollinearity
$y = a_1 a_2 x$	$y = a_1\left(1 - e^{-a_2 x}\right) \approx a_1 a_2 x$ at $a_2 x << 1$



Data & Model Preprocessing

$$((a+b)+c)+d \neq a+(b+(c+d)) \quad \text{as} \quad 1+10^{-20}=1$$

Representation of a number in computer with 64 bits

This is a hard job!

1	0	0	1	…	0	1	1

<u>Target</u> is to make Hesse matrix $\boldsymbol{A}$ regular and decrease $\mathrm{N}(\boldsymbol{A})$

Release the Computer!

Means	Data Scaling	$\boldsymbol{X} \rightarrow m\boldsymbol{X}$
	Data Centering	$\boldsymbol{X} \rightarrow \boldsymbol{X}-\boldsymbol{X}_0$
	Model Adjusting	$\boldsymbol{a} \rightarrow \phi(\boldsymbol{a}) \quad \boldsymbol{x} \rightarrow \psi(\boldsymbol{x}) \quad y \rightarrow \chi(\boldsymbol{y})$

Example: The Arrhenius Law

Standard form	Adjusted form
$k \exp\left(-\frac{E}{RT}\right)$	$\exp(a_1 - a_2 X)$
Scaling & centering	$X = \frac{1000}{T} - X_0$ $X_0 = \frac{1}{n}\sum_{i=1}^{n}\frac{1000}{T_i}$
Adjusting	$a_1 = \ln(k) - a_2 X_0$ $a_2 = \frac{E}{1000R}$
$k \approx 10^{+11}$ $E \approx 10^{+4}$	$a_1 \approx 1$ $a_2 \approx 10$
$\mathrm{N}(\boldsymbol{A}) = 20$	$\mathrm{N}(\boldsymbol{A}) = 2$

Simple transformation with great effect!

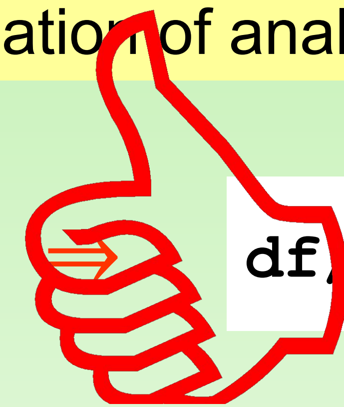
Derivative Calculation and Precision

N(A)	A^{-1}	$y=f(a,x)$	$0=f(y,a,x)$	$dy/dx=f(y,a,x)$
10	10+2=12	12+0=12	12+2=14	14+2=16
6	6+2=8	8+0=8	8+2=10	10+2=12

1) Numerical calculation of difference derivatives

$$\frac{\partial f}{\partial a} \approx \frac{f(a+h)-f(a)}{h}$$

2) Auto calculation of analytical derivatives

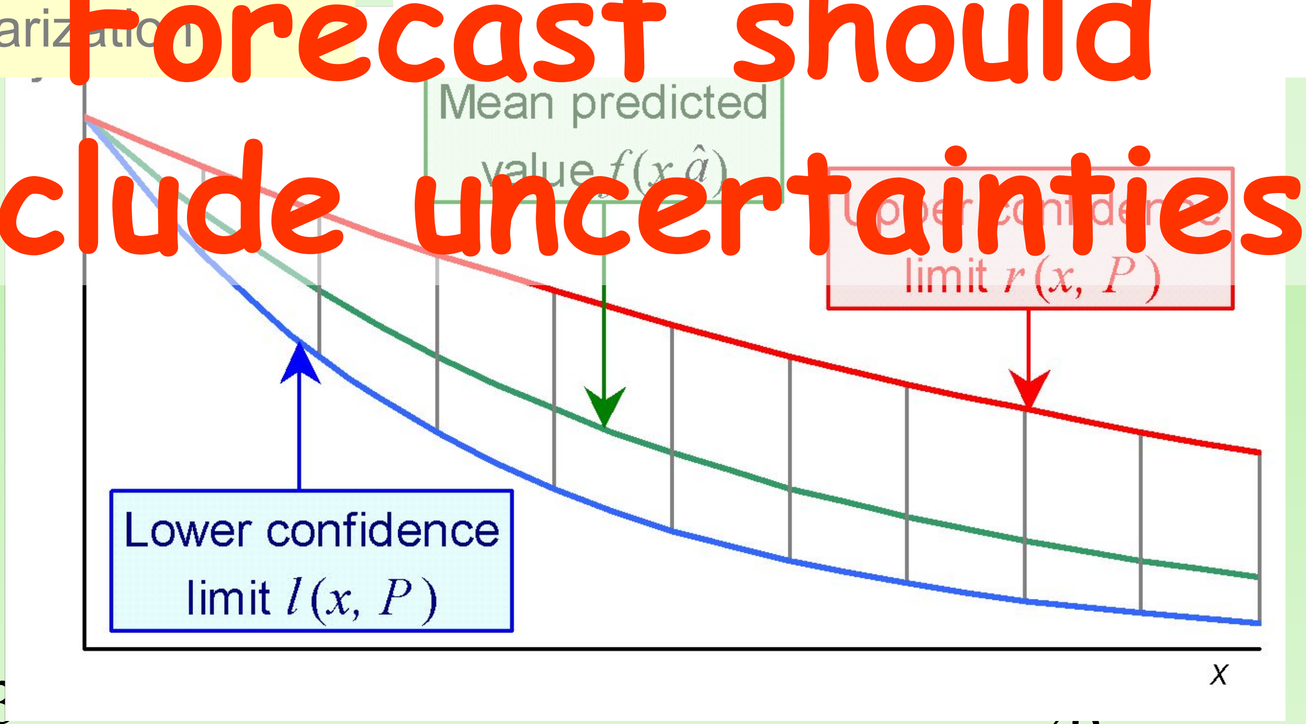


`f=exp(-a*t)` ⇒ `df/da=-t*exp(-a*t)`

5. Prediction

Reliable Prediction

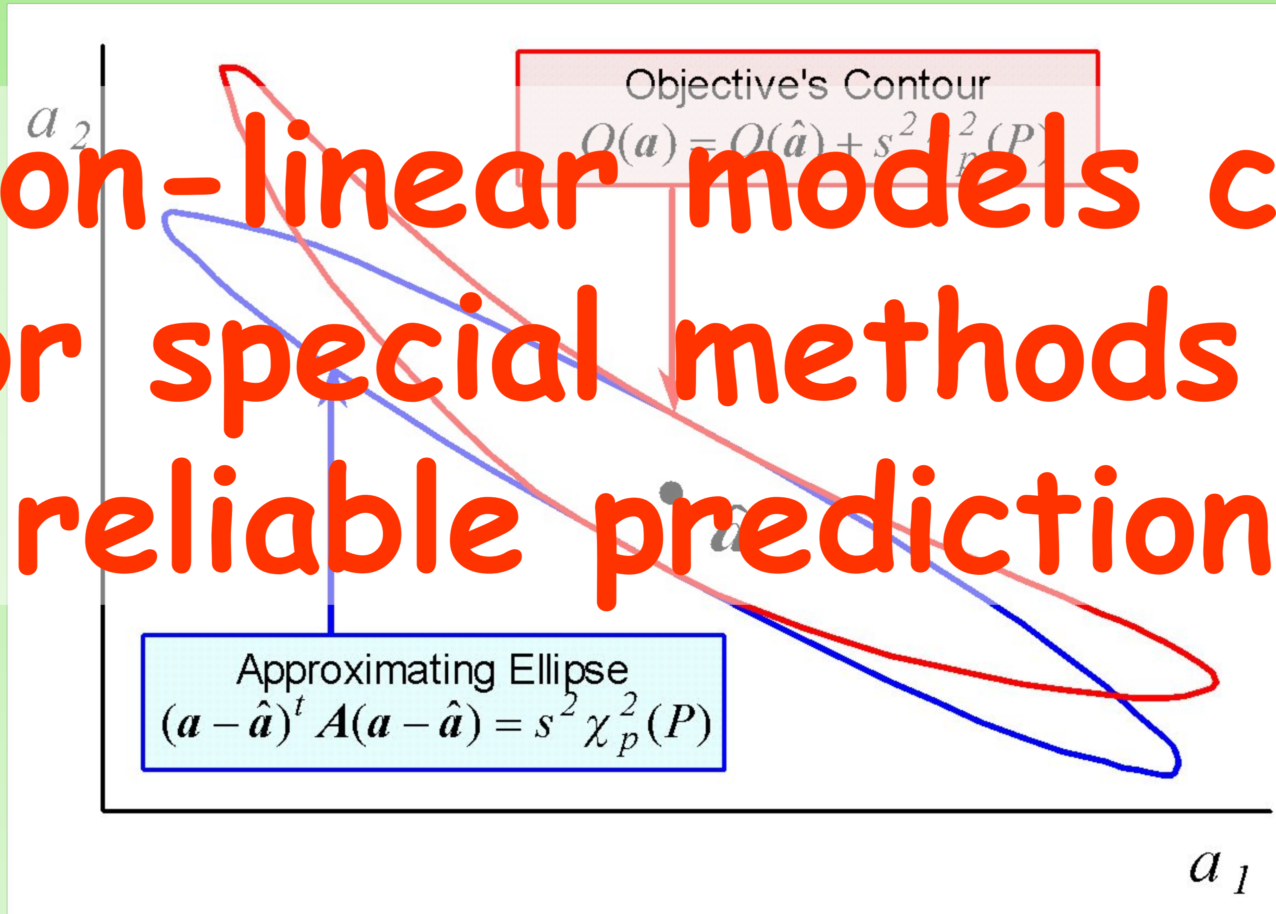
Estimate of response	$\hat{y} = f(\boldsymbol{x}, \hat{\boldsymbol{a}})$
Confidence limits	$\text{Prob}\{l(\boldsymbol{x}, P) < f(\boldsymbol{x}, \boldsymbol{a}) < r(\boldsymbol{x}, P)\} \geq P$
Linearization	$r(\boldsymbol{x}, P) = f(\boldsymbol{x}, \hat{\boldsymbol{a}}) + g(P)\sqrt{\boldsymbol{v}^t \boldsymbol{C} \boldsymbol{v}}$



Forecast should include uncertainties!

Nonlinearity and Simulation

$$\boldsymbol{a}^* \sim N(\hat{\boldsymbol{a}}, \boldsymbol{C}) \rightarrow \boldsymbol{a}_1^*, \ldots, \boldsymbol{a}_M^* \rightarrow f_1^*, \ldots, f_M^* \rightarrow r(P, \boldsymbol{x})$$



Non-linear models call for special methods of reliable prediction!

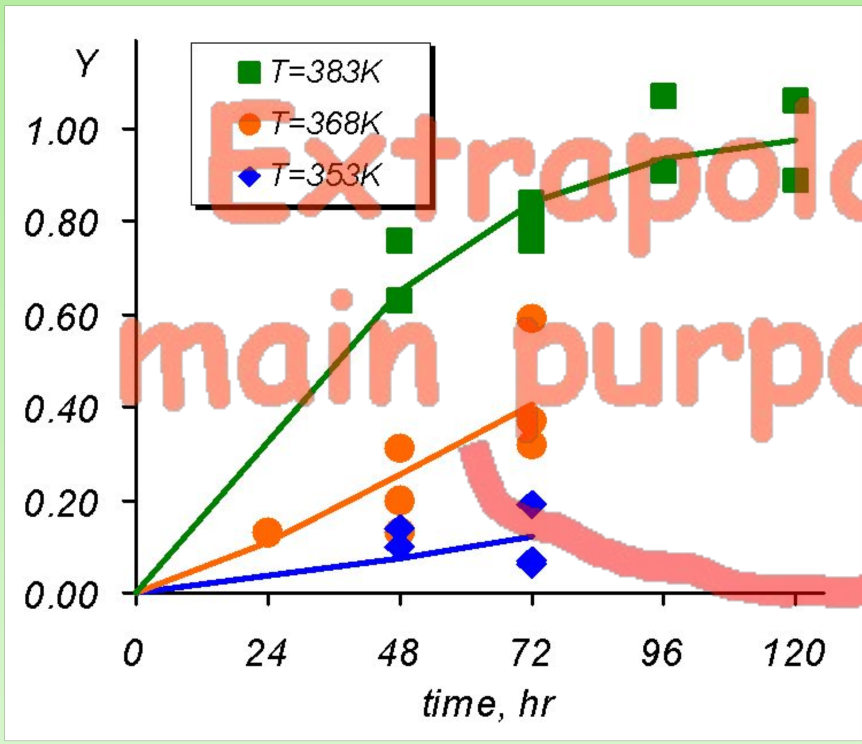
Prediction: Example

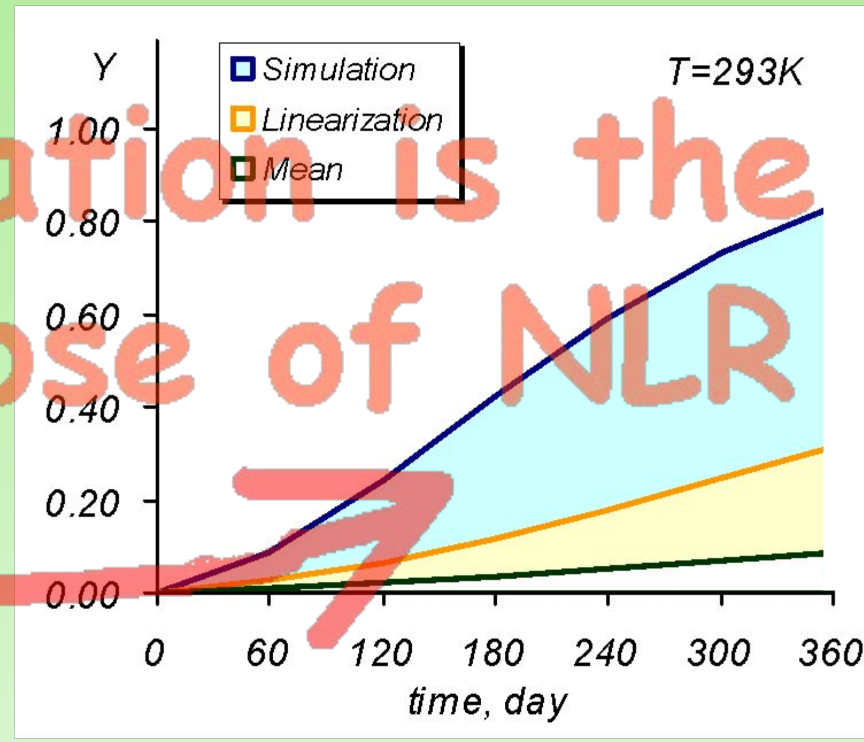
Model of rubber aging

$$Y = 1 - e^{-(kt)^a}, \qquad k = e^{k_0 - \frac{E}{RT}}$$

Accelerated aging tests

Upper confidence limits





Extrapolation is the main purpose of NLR !

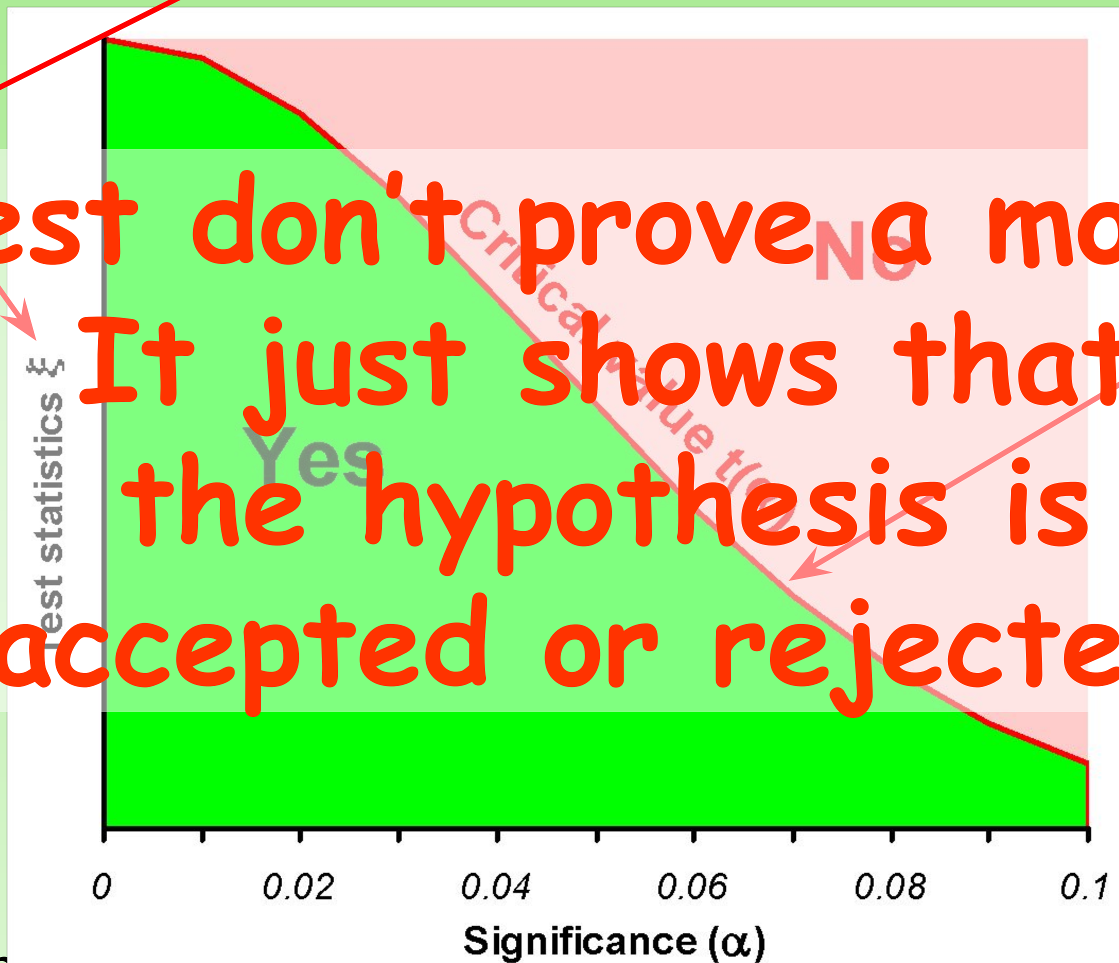
6. Testing

Hypotheses Testing

Test statistics ξ is compared with critical value $t(\alpha)$

From experiment

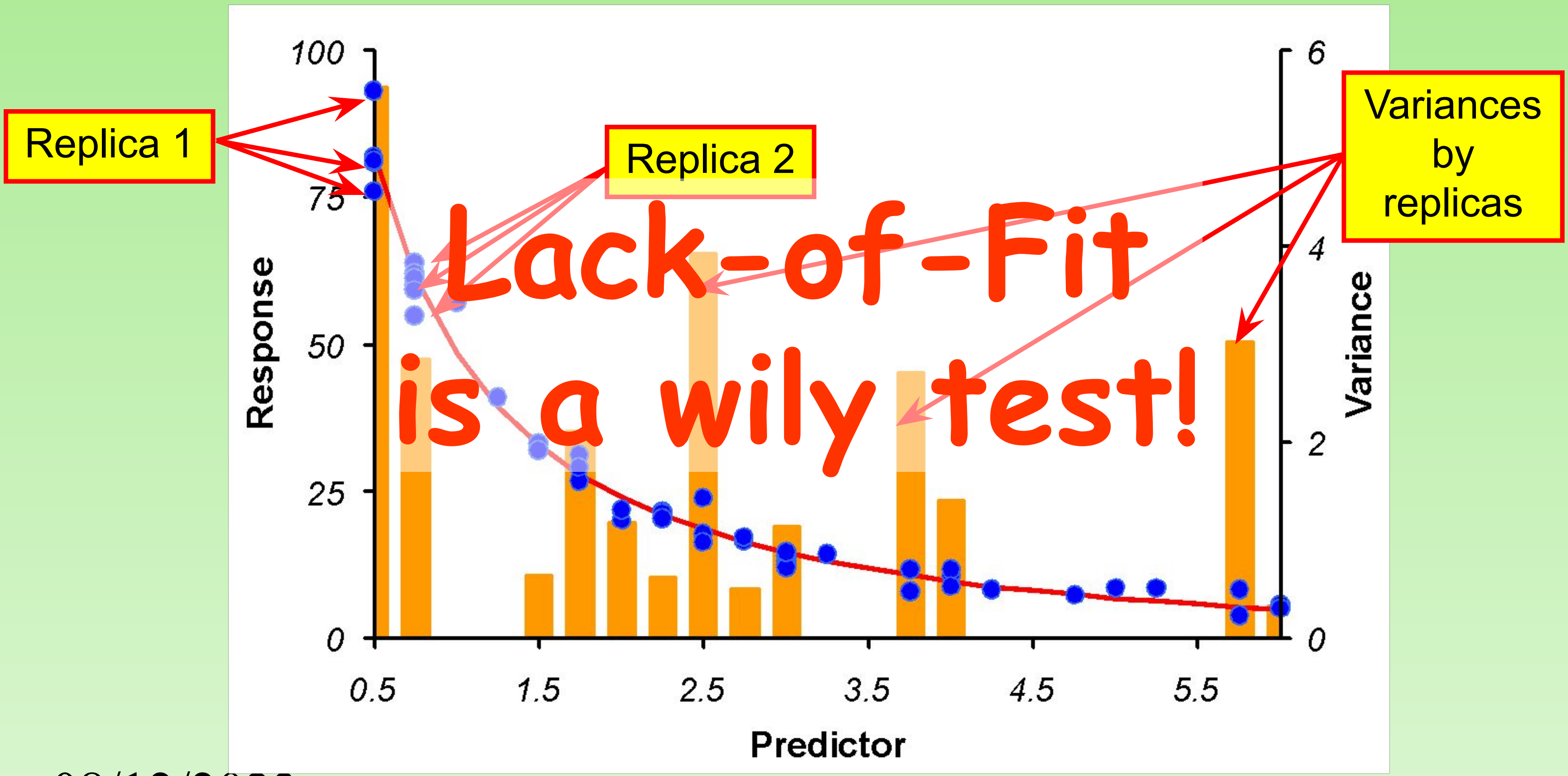
From theory



Test don't prove a model!
It just shows that the hypothesis is accepted or rejected!

Lack-of-Fit and Variances Tests

These hypotheses are based on variances and they can't be tested without **replicas**!



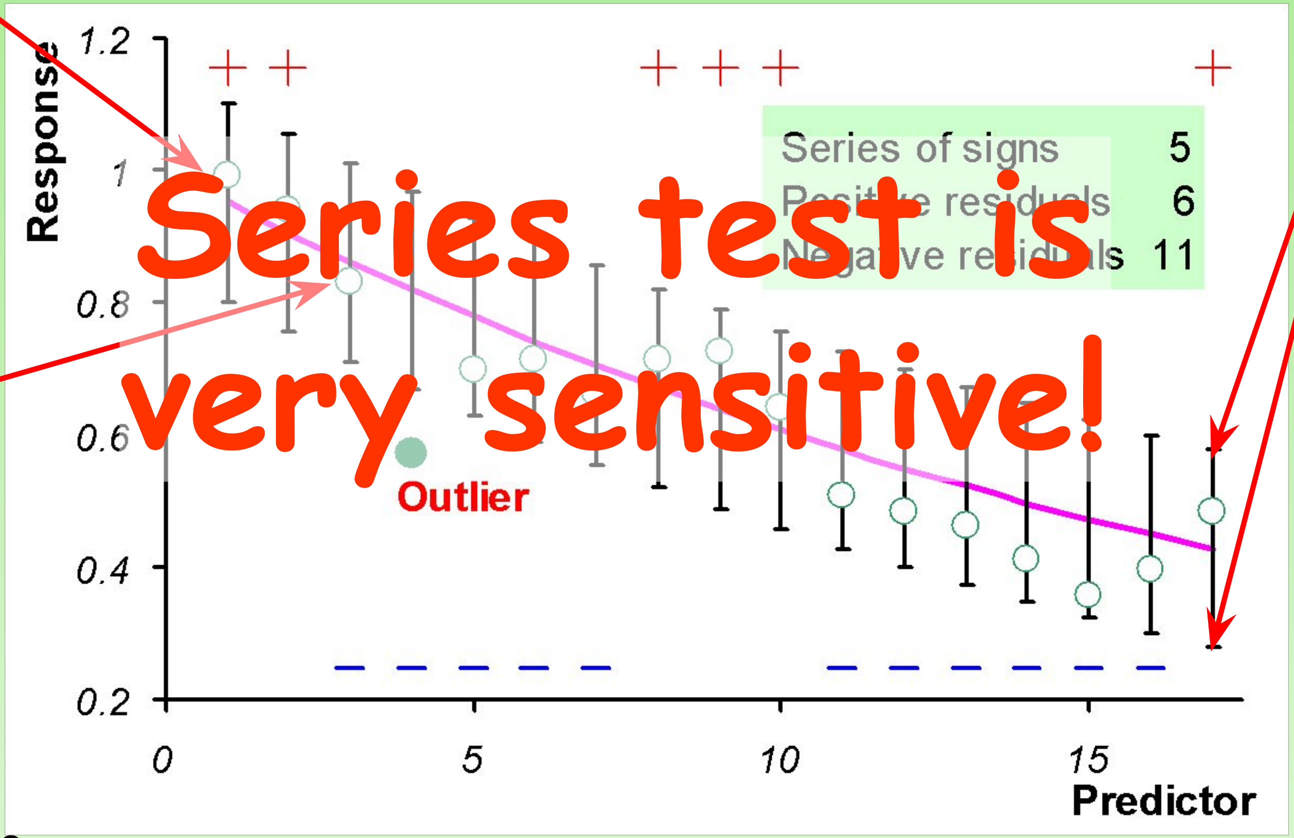
Outlier and Series Tests

These hypotheses are based on residuals and they can be tested without replicas

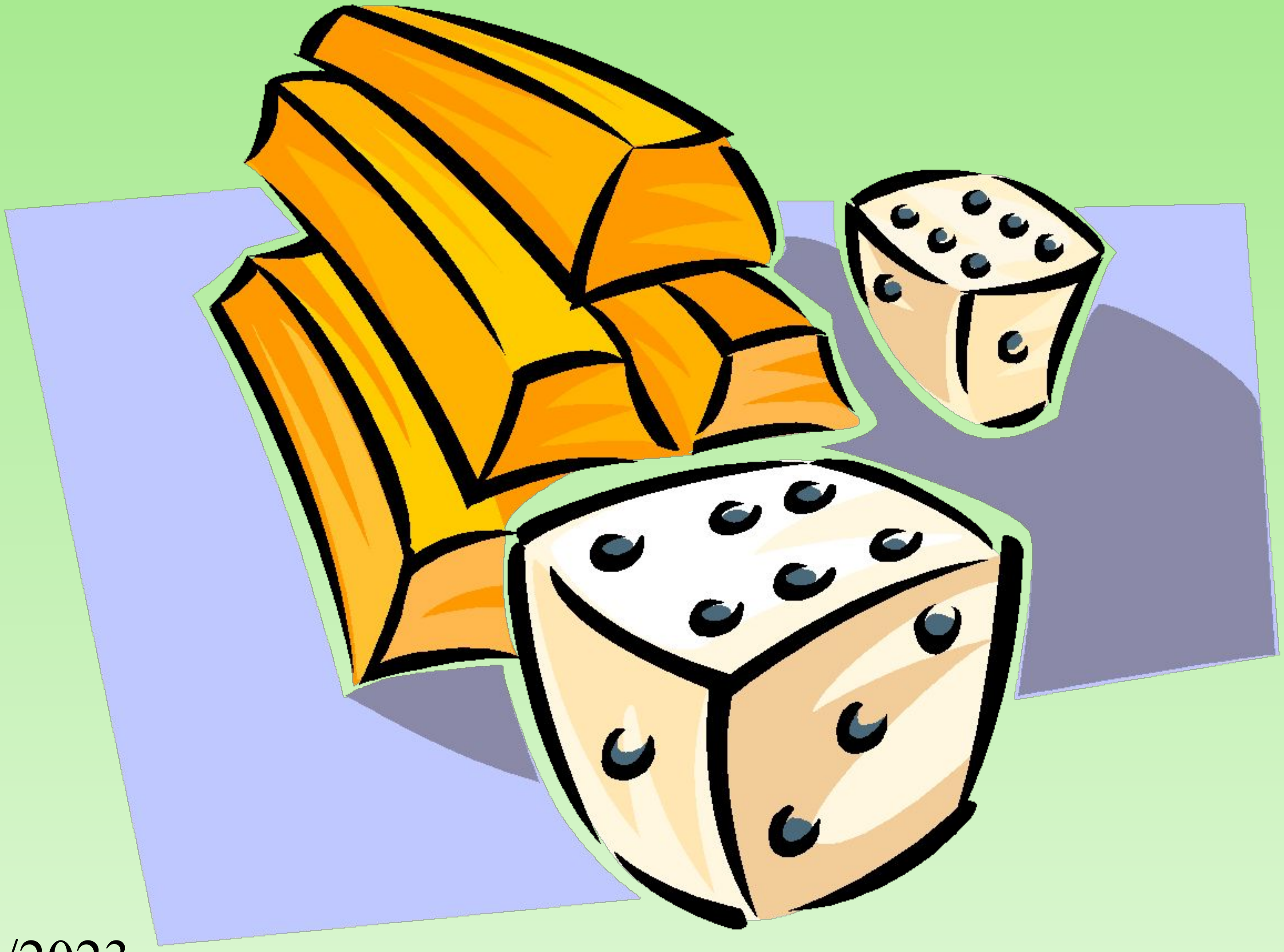
Positive residual

Negative residual

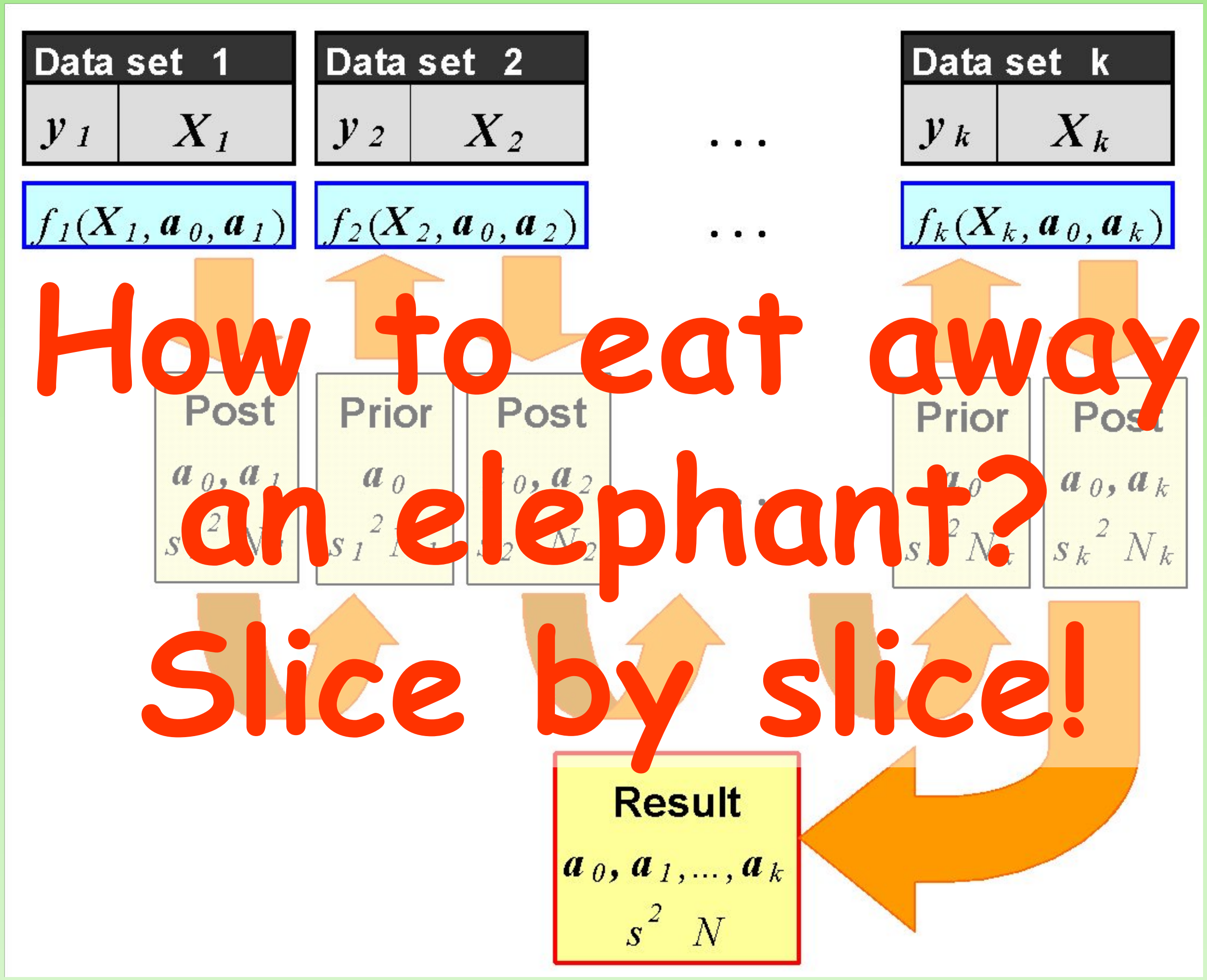
Acceptable deviation



7. Bayesian Estimation



Bayesian Estimation



Posterior and Prior Information. Type I

Posterior Information	Prior Information
Parameter estimates $\hat{\boldsymbol{a}}$	Prior parameter values $\boldsymbol{b}$
Matrix $\boldsymbol{F}$	Recalculated matrix $\boldsymbol{H}$
Error variance s^2	Prior variance value s_0^2
NDF N_f	Prior NDF N_0

The same error in each portion of data!

Objective Function

$$Q(\boldsymbol{a}) = S(\boldsymbol{a}) + B(\boldsymbol{a})$$

$$B(\boldsymbol{a}) = s_0^2 \left(N_0 + R(\boldsymbol{a})\right) \quad R(\boldsymbol{a}) = (\boldsymbol{a} - \boldsymbol{b})^t \boldsymbol{H} (\boldsymbol{a} - \boldsymbol{b})$$

Posterior and Prior Information. Type II

Posterior Information	Prior Information
Parameter estimates $\hat{\boldsymbol{a}}$	Prior parameter values $\boldsymbol{b}$
Matrix F	Recalculated matrix $\boldsymbol{H}$

$$Q(\boldsymbol{a}) = S(\boldsymbol{a})B(\boldsymbol{a})$$

Objective Function

$$B(\boldsymbol{a}) = \exp\left(\frac{R(\boldsymbol{a})}{N}\right) \qquad R(\boldsymbol{a}) = (\boldsymbol{a}-\boldsymbol{b})^t \boldsymbol{H}(\boldsymbol{a}-\boldsymbol{b})$$

Different errors in each portion of data!

8. Conclusions

Mysterious Nature

LR Model

NLR Model