The SAEM algorithm and its implementation in MONOLIX 2.2

Marc Lavielle[1,2,3]
Hector Mesa[1,6]
France Mentré[4,5]

[1]University Paris 11 [2]University Paris 5 [3]INRIA Futurs
[4]INSERM U738 [5]University Paris 7 [6]University of La Habana

The SAEM algorithm

A stochastic version of the EM algorithm for **Maximum Likelihood Estimation** in the general framework of incomplete data models.

- Delyon B., Lavielle M. and Moulines E., *Convergence of a stochastic approximation version of the EM algorithm*, The Annals of Statistics, vol. 27, no. 1, pp 94–128, 1999.

The model

$$y_{ij} = f(x_{ij}, \phi_i) + \varepsilon_{ij} \quad , \; 1 \leq i \leq N \quad , \quad 1 \leq j \leq n_i$$

$y_{ij} \in \mathbb{R}$ is the jth observation of subject i,
N is the number of subjects
n_i is the number of observations of subject i.

The regression variables, or design variables, (x_{ij}) are **known**,
$x = (Dose \; , \; Time)$ for PK and PKPD models,
$x = Concentration$ for PD models, ...

The individual PKPD parameters (ϕ_i) are **unknown**,

The basic model

$$y_{ij} = f(x_{ij}, \phi_i) + \varepsilon_{ij} \quad , \; 1 \leq i \leq N \quad , \quad 1 \leq j \leq n_i$$

The vector ϕ_i of individual parameters is assumed to be Gaussian:

$$\phi_i = \mu + \eta_i \quad \text{with} \quad \eta_i \sim_{i.i.d.} \mathcal{N}(0, \Gamma)$$

μ: **unknown vector of population parameters** (*the fixed effects*),
(η_i): **unknown random vectors** (*the random effects*).

The sequence (ε_{ij}) is assumed to be Gaussian:

$$\varepsilon_{ij} \sim_{i.i.d.} \mathcal{N}(0, \sigma^2)$$

The incomplete data model

$$y_{ij} = f(x_{ij}, \phi_i) + \varepsilon_{ij} \quad , \; 1 \leq i \leq N \quad , \quad 1 \leq j \leq n_i$$

We are in a classical framework of *"incomplete data"*:

- the mesurement $y = (y_{ij}, 1 \leq i \leq N \, , 1 \leq j \leq n_i)$ are the *"observed data"*
- the individual random parameters $\phi = (\phi_i, \, 1 \leq i \leq N)$, are the *"non observed data"*,
- the *"complete data"* of the model is (y, ϕ).

Our purpose is to compute the maximum likelihood estimator of the unknown set of parameters $\theta = (\mu, \Gamma, \sigma^2)$, by maximizing the likelihood of the observations $\ell(y, \theta)$, without any approximation on the model.

The incomplete data model

$$y_{ij} = f(x_{ij}, \phi_i) + \varepsilon_{ij} \quad , \; 1 \leq i \leq N \quad , \quad 1 \leq j \leq n_i$$

We are in a classical framework of *"incomplete data"*:

- the mesurement $y = (y_{ij}, 1 \leq i \leq N \, , 1 \leq j \leq n_i)$ are the *"observed data"*
- the individual random parameters $\phi = (\phi_i, \, 1 \leq i \leq N)$, are the *"non observed data"*,
- the *"complete data"* of the model is (y, ϕ).

Our purpose is to compute the maximum likelihood estimator of the unknown set of parameters $\theta = (\mu, \Gamma, \sigma^2)$, by maximizing the likelihood of the observations $\ell(y, \theta)$, without any approximation on the model.

Some existing algorithms

- **FO, FOCE:** based on the linearization of the model,
  - theoretical drawbacks: no well-known statistical properties of the algorithm,
  - practical drawbacks: very sensitive to the initial guess, does not always converge, poor estimation of some parameters,...

- **Gaussian quadrature:** based on a numerical approximation of the likelihood
  - nice theoretical properties: maximum likelihood estimation is performed,
  - practical drawbacks : limited to few random effects.

Some existing algorithms

- **FO, FOCE:** based on the linearization of the model,
  - theoretical drawbacks: no well-known statistical properties of the algorithm,
  - practical drawbacks: very sensitive to the initial guess, does not always converge, poor estimation of some parameters,...
- **Gaussian quadrature:** based on a numerical approximation of the likelihood
  - nice theoretical properties: maximum likelihood estimation is performed,
  - practical drawbacks : limited to few random effects.

EM-type algorithms

MCPEM, MCEM, PEM, SAEM, ...

- use simulated sequences of random effects (no limitation),
- very good results in the practice,
- few theoretical results.

The SAEM algorithm

- **Convergence of SAEM theoretically demonstrated:**

  Delyon B., Lavielle M. and Moulines E., *Convergence of a stochastic approximation version of the EM algorithm*, The Annals of Statistics, vol. 27, no. 1, pp 94–128, 1999.

  Kuhn E., Lavielle M. *Coupling a stochastic approximation version of EM with a MCMC procedure* ESAIM P&S, vol.8, pp 115-131, 2004.

- **Intensive use of powerful and well-known algorithms:**
  MCMC, Simulated annealing, Importance Sampling, . . .
- **Very good practical properties:**
  - SAEM converges succesfully with complex models and provides accurate estimations of the parameters of the model, the standard errors, the likelihood function,
  - SAEM is few sensitive to the initialization,
  - SAEM is fast.

The SAEM algorithm

- **Convergence of SAEM theoretically demonstrated:**

  Delyon B., Lavielle M. and Moulines E., *Convergence of a stochastic approximation version of the EM algorithm*, The Annals of Statistics, vol. 27, no. 1, pp 94–128, 1999.

  Kuhn E., Lavielle M. *Coupling a stochastic approximation version of EM with a MCMC procedure* ESAIM P&S, vol.8, pp 115-131, 2004.

- **Intensive use of powerful and well-known algorithms:**
  MCMC, Simulated annealing, Importance Sampling, . . .

- **Very good practical properties:**
  - SAEM converges succesfully with complex models and provides accurate estimations of the parameters of the model, the standard errors, the likelihood function,
  - SAEM is few sensitive to the initialization,
  - SAEM is fast.

The SAEM algorithm

- **Convergence of SAEM theoretically demonstrated:**

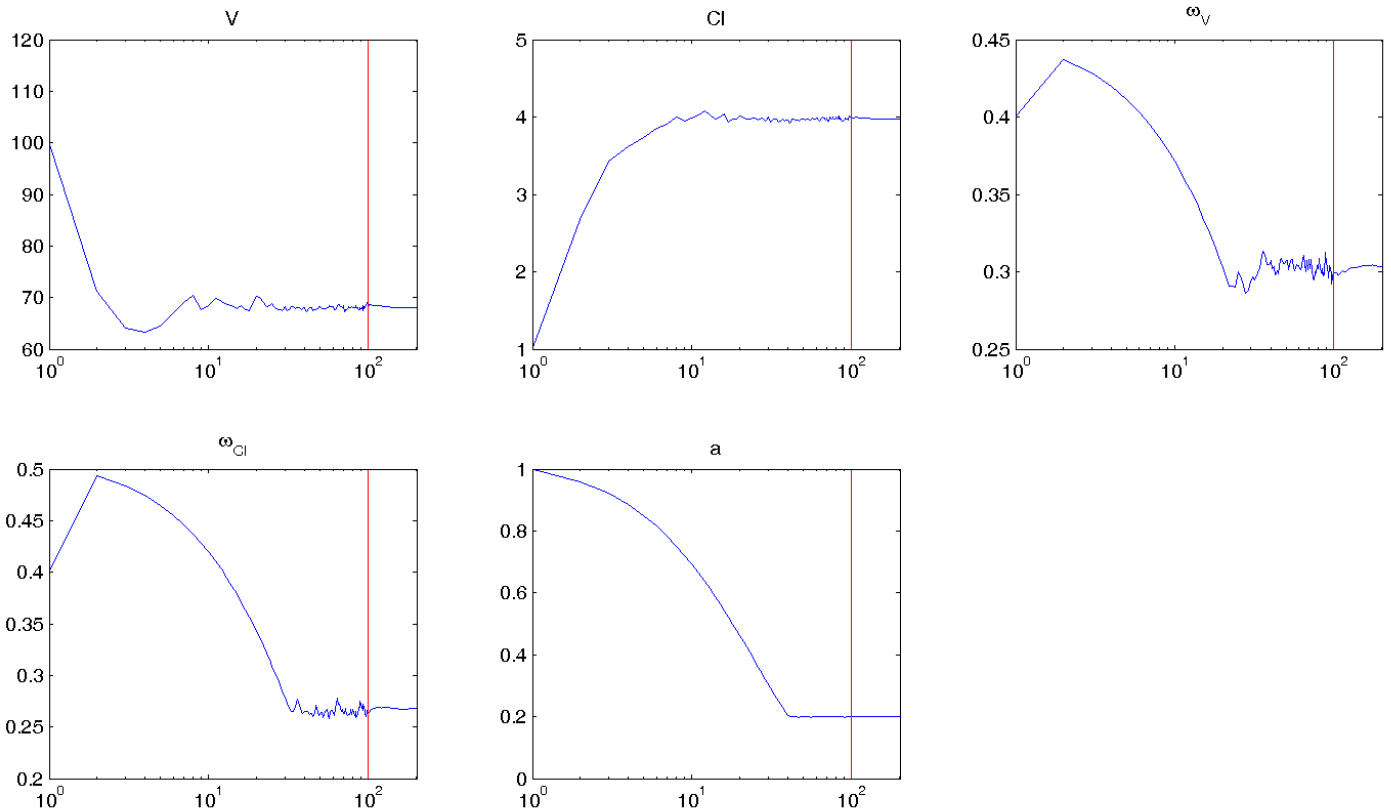
  Delyon B., Lavielle M. and Moulines E., *Convergence of a stochastic approximation version of the EM algorithm*, The Annals of Statistics, vol. 27, no. 1, pp 94–128, 1999.

  Kuhn E., Lavielle M. *Coupling a stochastic approximation version of EM with a MCMC procedure* ESAIM P&S, vol.8, pp 115-131, 2004.

- **Intensive use of powerful and well-known algorithms:**
  MCMC, Simulated annealing, Importance Sampling, . . .
- **Very good practical properties:**
  - SAEM converges succesfully with complex models and provides accurate estimations of the parameters of the model, the standard errors, the likelihood function,
  - SAEM is few sensitive to the initialization,
  - SAEM is fast.

Convergence of SAEM: an example

Fitting an IV Bolus model with SAEM
Estimation of $(V, Cl, \omega_V, \omega_{Cl}, \sigma)$



Some extensions

- **Models with data below LOQ (left-censored data),**

  Samson A., Lavielle M., Mentré F. *Extension of the SAEM algorithm to left-censored data in non-linear mixed-effects model: application to HIV dynamics models*, Computational Statistics and Data Analysis, vol. 51, pp. 1562–1574, 2006.

- **Models defined by Ordinary Differential Equations,**

  Donnet S., Samson A. *Estimation of parameters in incomplete data models defined by dynamical systems*, Jour. of Stat. Planning and Inference (to appear), 2007.

- **Models defined by Stochastic Differential Equations,**

  Donnet S., Samson A. *Parametric inference for mixed models defined by stochastic differential equations*, ESAIM P&S (to appear), 2007.

Evaluation of the PK library of MONOLIX 2.2

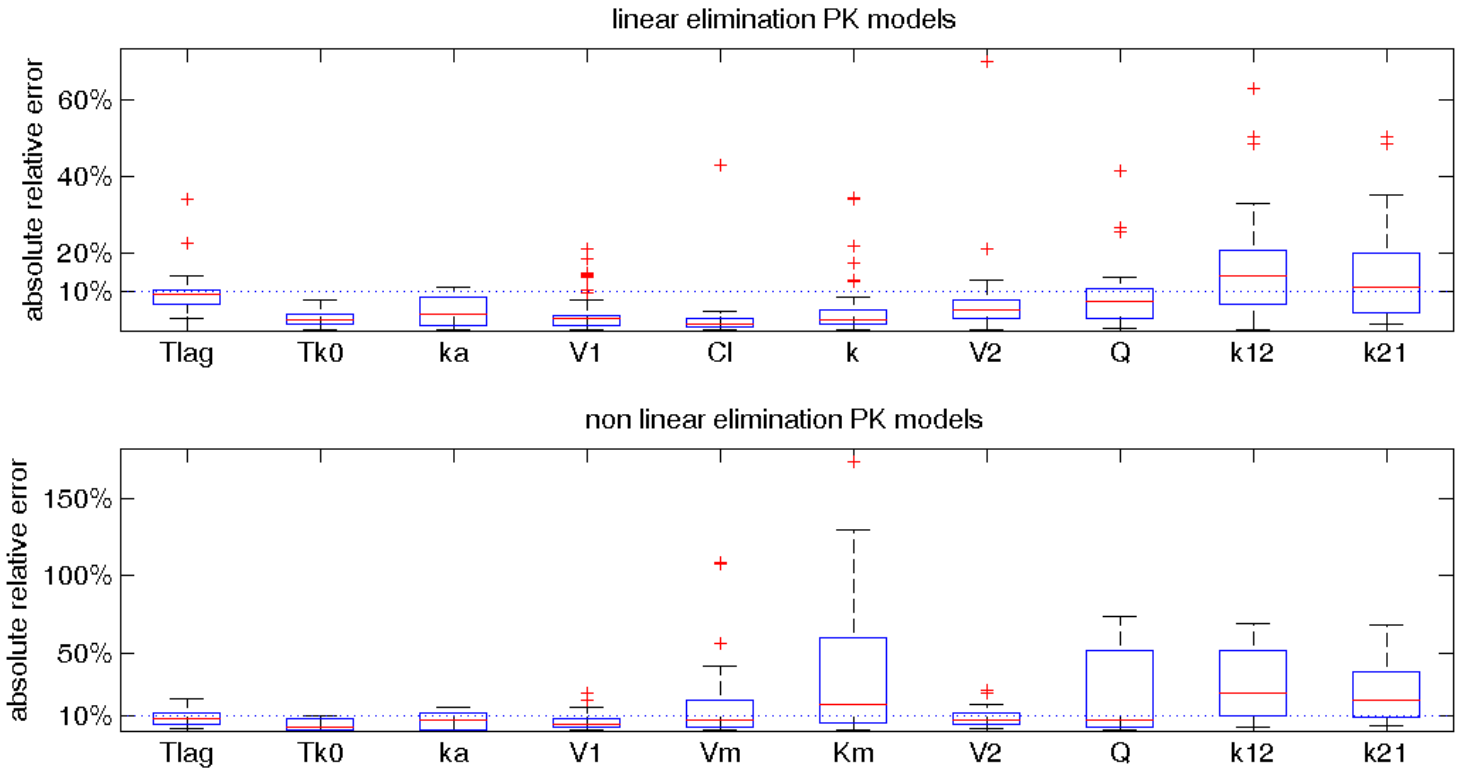
This work was done in collaboration with *Exprimo NV* to test exhaustively all the available models in the PK library of MONOLIX 2.2.

- **42 models tested** (1 and 2 compartment models, several parametrizations, linear and nonlinear eliminations),
- **84 datasets simulated** (single dose, multiple doses, steady-state),
- **150 runs**,
- The NONMEM software version V level 1 and VI version 1 were evaluated (Exprimo),
- The MONOLIX software version 2.2 was evaluated (M. Lavielle).

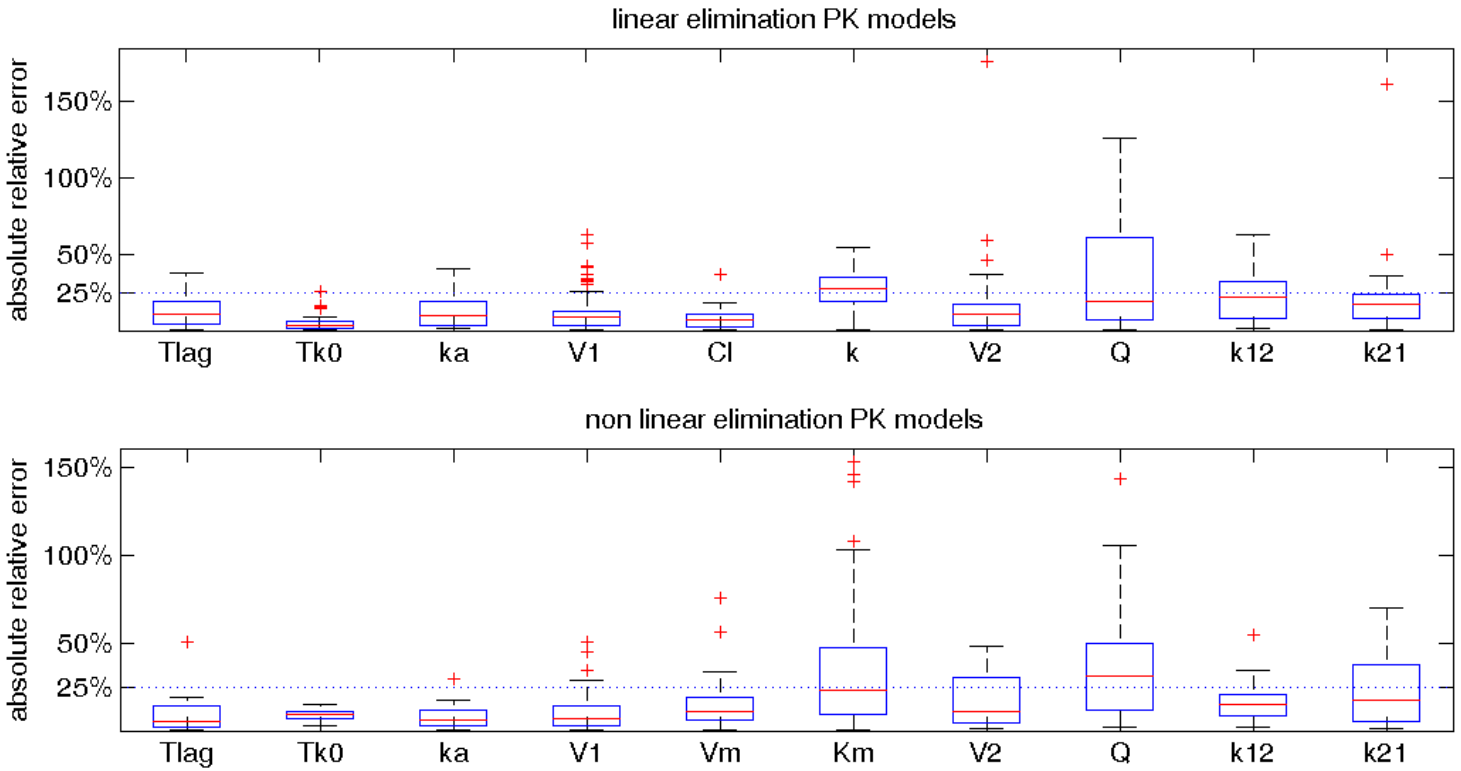
% of successful runs with MONOLIX (150 runs)

Estimation of the population parameters	100%
Estimation of the variability	100%
Estimation of the standard errors	100%
Estimation of the objective function	100%

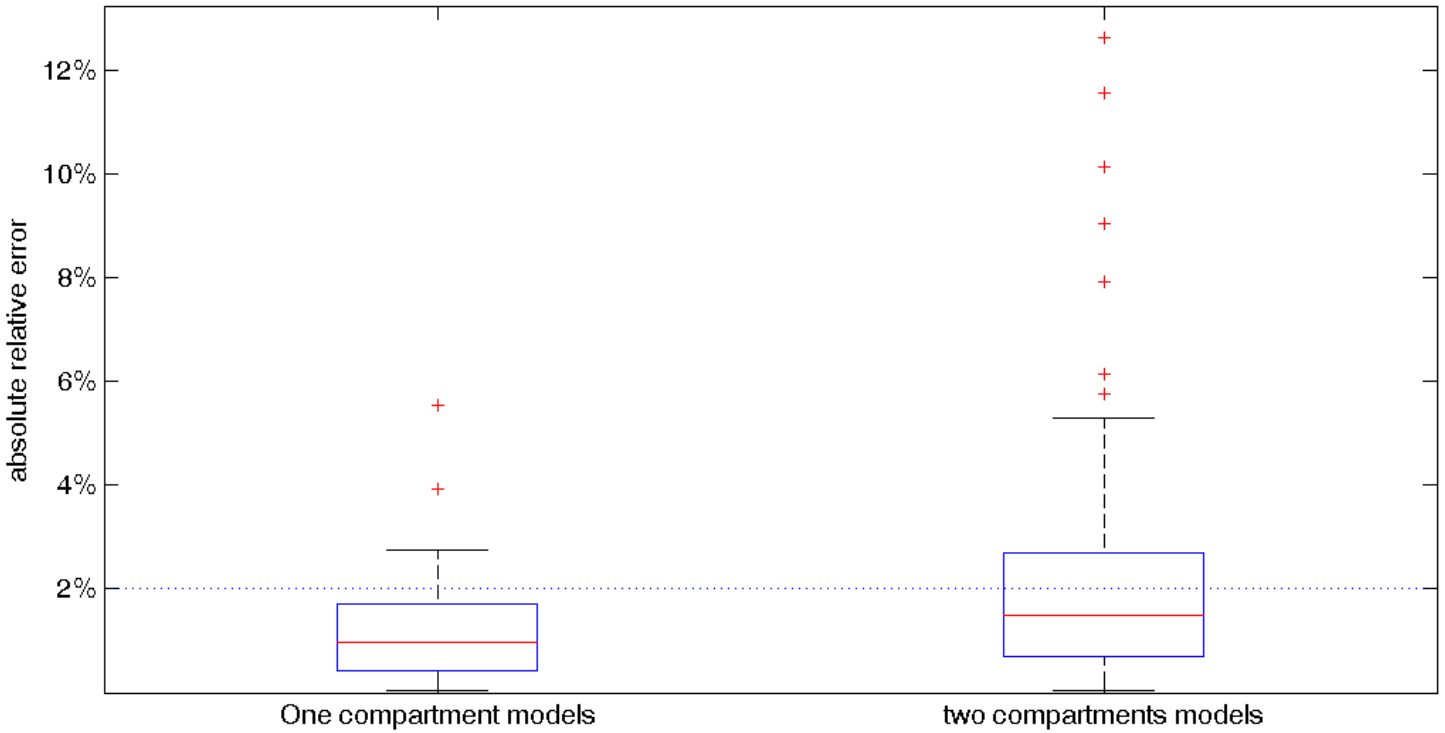
Estimation of the fixed effects (absolute relative error)



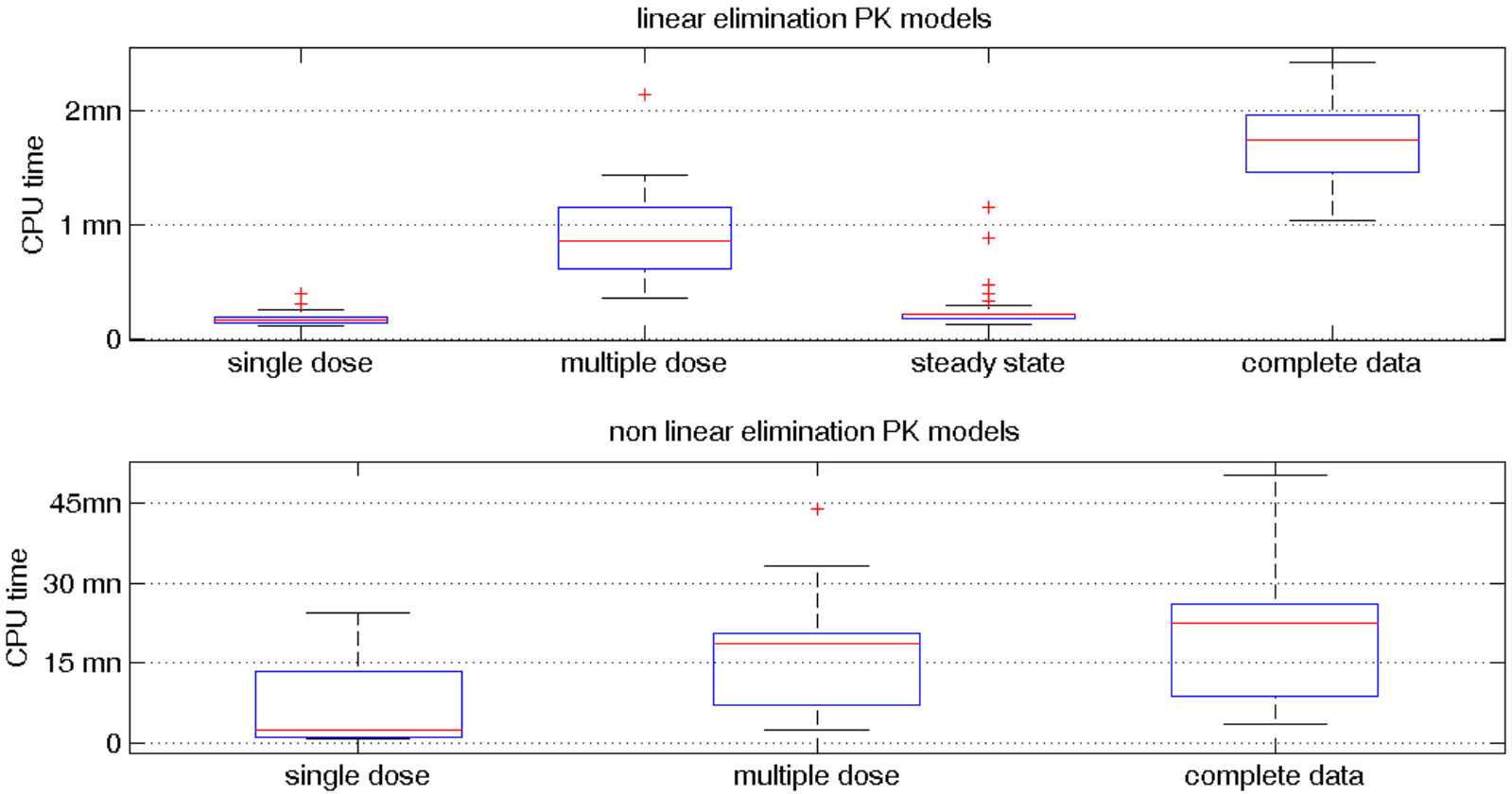
Estimation of the standard deviation of the random effects (absolute relative error)



Estimation of the standard deviation of the residual error (absolute relative error)

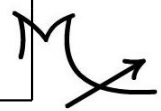


CPU time

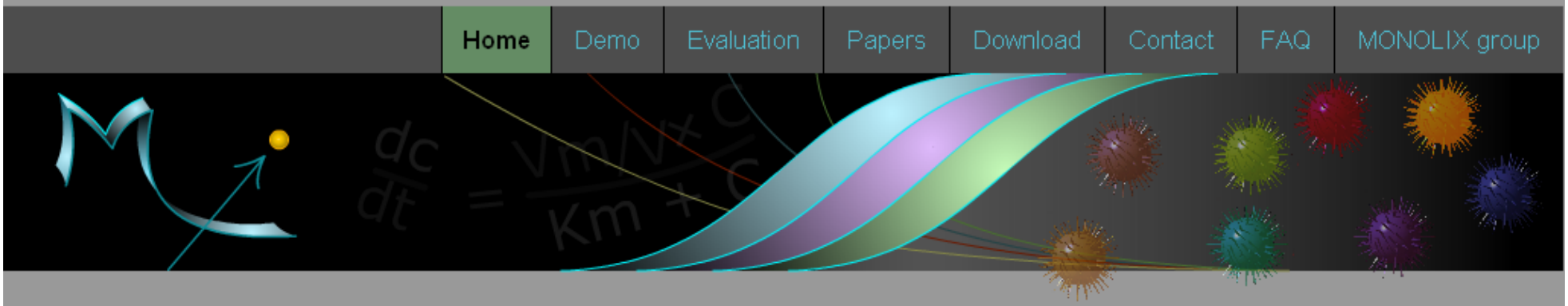

linear elimination PK models
CPU time
2mn
1 mn
0
single dose
multiple dose
steady state
complete data
non linear elimination PK models
CPU time
45mn
30 mn
15 mn
0
single dose
multiple dose
complete data

Evaluation of the PK library of MONOLIX 2.2

	Single Dose 120 subjects 2280 observations	Multiple Doses (7 doses) 120 subjects 5520 observations
IV bolus 1 compartment model **linear elimination**	6"	17"
IV bolus 1 compartment model **non linear elimination**	29"	2' 7"
1st order oral absorption 1 compartment model linear elimination **lag time**	11"	1'12"



The MONOLIX software



http://www.monolix.org