Estimation of nonlinear mixed effects model in pharmacokinetics with the SAEM algorithm implemented in MONOLIX

Pr France Mentré, INSERM U738, University Paris Diderot
Pr Marc Lavielle, INRIA, Universities Paris 5 & 11
Paris, France

Outline

1. Introduction
2. Brief history of estimation methods in NLMEM
3. Stochastic EM algorithms
4. MONOLIX software
5. Comparison of Stochastic EM algorithms to NONMEM
6. PKPD example with MONOLIX
7. Conclusion

1. Introduction

- Nonlinear mixed effects models (NLMEM) allow "population" PKPD analyses
  - Global analysis of data in all individuals
  - Rich or sparse design
- Increasingly used in clinical and non clinical drug development
  - Parameter estimation
  - Model selection
  - Covariate testing
  - Predictions & Simulations

  ➢ Good estimation methods needed
- Focus here on Maximum Likelihood Estimation (MLE) parametric methods

2. Brief history of estimation methods

NON linear Mixed Effects Model

L Sheiner & S Beal, UCSF

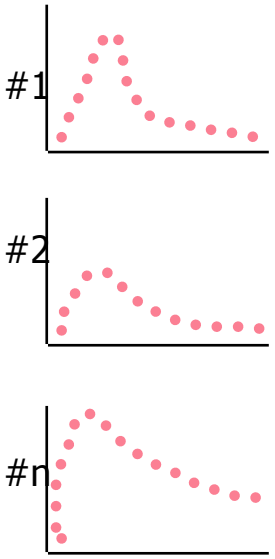
- **1972**: The concept and the **FO method**

  Sheiner, Rosenberg & Melmon (1972). Modelling of individual pharmacokinetics for computer aided drug dosage. *Comput Biomed Res*, 5:441-59.
- **1977**: The first case study

  Sheiner, Rosenberg & Marathe (1977). Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *J Pharmacokin Biopharm*, 5: 445-479.
- **1980**: NONMEM - An IBM-specific software

  Beal & Sheiner (1980). The NONMEM system. *American Statistician, 34:118-19*.
  Beal & Sheiner (1982). Estimating population kinetics. *Crit Rev Biomed Eng*, 8:195-222.

Standard Two-Stage approach (STS)

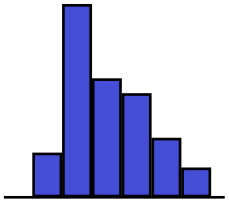


stage 1

Individual fitting Non Linear regression

Subject	**Parameters estimate**
#1	12.3
#2	21.9
#n	16.1

stage 2

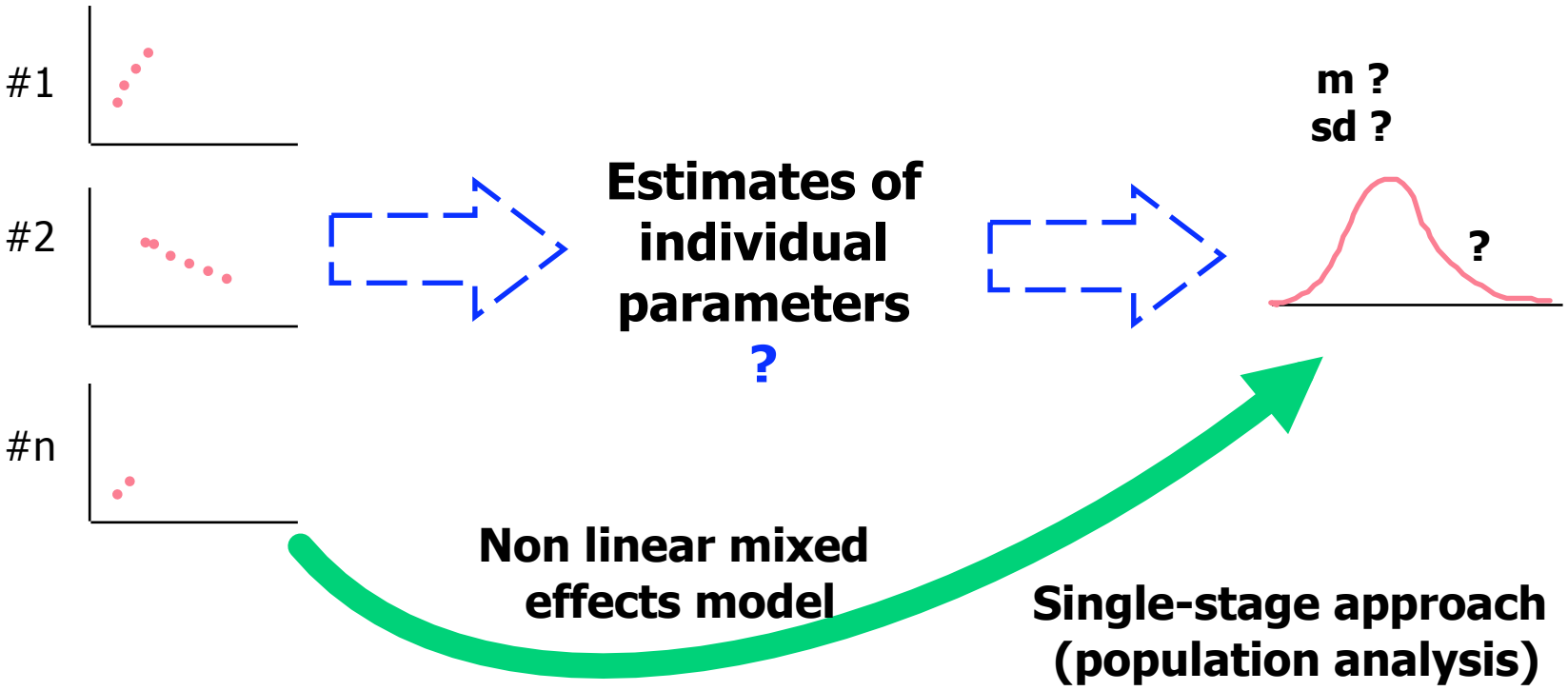


m , sd

Descriptive statistics, linear stepwise regression for covariate effect

From Steimer (1992): « Population models and methods, with emphasis on pharmacokinetics », in M. Rowland and L. Aarons (eds), *New strategies in drug development and clinical evaluation, the population approach*

Population approach



From Steimer (1992) : « Population models and methods, with emphasis on pharmacokinetics », in M. Rowland and L. Aarons (eds), *New strategies in drug development and clinical evaluation, the population approach*

The population approach

- N individuals (i = 1, ..., N)
- Structural model f: same shape in all individuals

$$y_{ij} = f(\theta_i, t_{ij}) + g(\theta_i, t_{ij})\, \varepsilon_{ij} \quad (j = 1, \ldots, n_i)$$

- Assumption on the individual parameters

$\theta_i = \mu + \eta_i$ or $\theta_i = \mu \exp(\eta_i)$

μ = "mean" parameters (fixed effects)

η_i = individual random effects

$\eta_i \sim$ Normal distribution with mean 0 and variance Ω

Ω: inter-individual variability

The FO method (1)

- Estimation of population parameters by maximum likelihood
  - find parameters that maximise the probability density function of the observations given the model
  - good statistical properties of ML estimator

- Problem: No closed form of the likelihood

$$y_{ij} = f(\mu + \eta_i , t_{ij}) + g(\mu + \eta_i , t_{ij})\ \varepsilon_{ij}$$

- First order linearisation of the model around $\eta = 0$

$$y_{ij} \approx f(\mu, t_{ij}) + \partial f^t/\partial \eta\ (\mu , t_{ij}) \times \eta_i + g(\mu, t_{ij})\ \varepsilon_{ij}$$

➔ Extended Least Square criterion

The FO method (2)

Advantages

- Better than Standard Two-Stage approach in many cases
  - STS neglects estimation error
    - Overstimation of inter-individual variability
    - OK for very rich design and small residual error
  - STS cannot be used for rather sparse designs

- Takes into account correlation within individuals
  - better than all naive approaches
  - Naive avering of data (NAD): "population average"
  - Naive pooling of data (NPD): one "giant" individual

More recent statistical developments in estimation methods for NLMEM: three periods

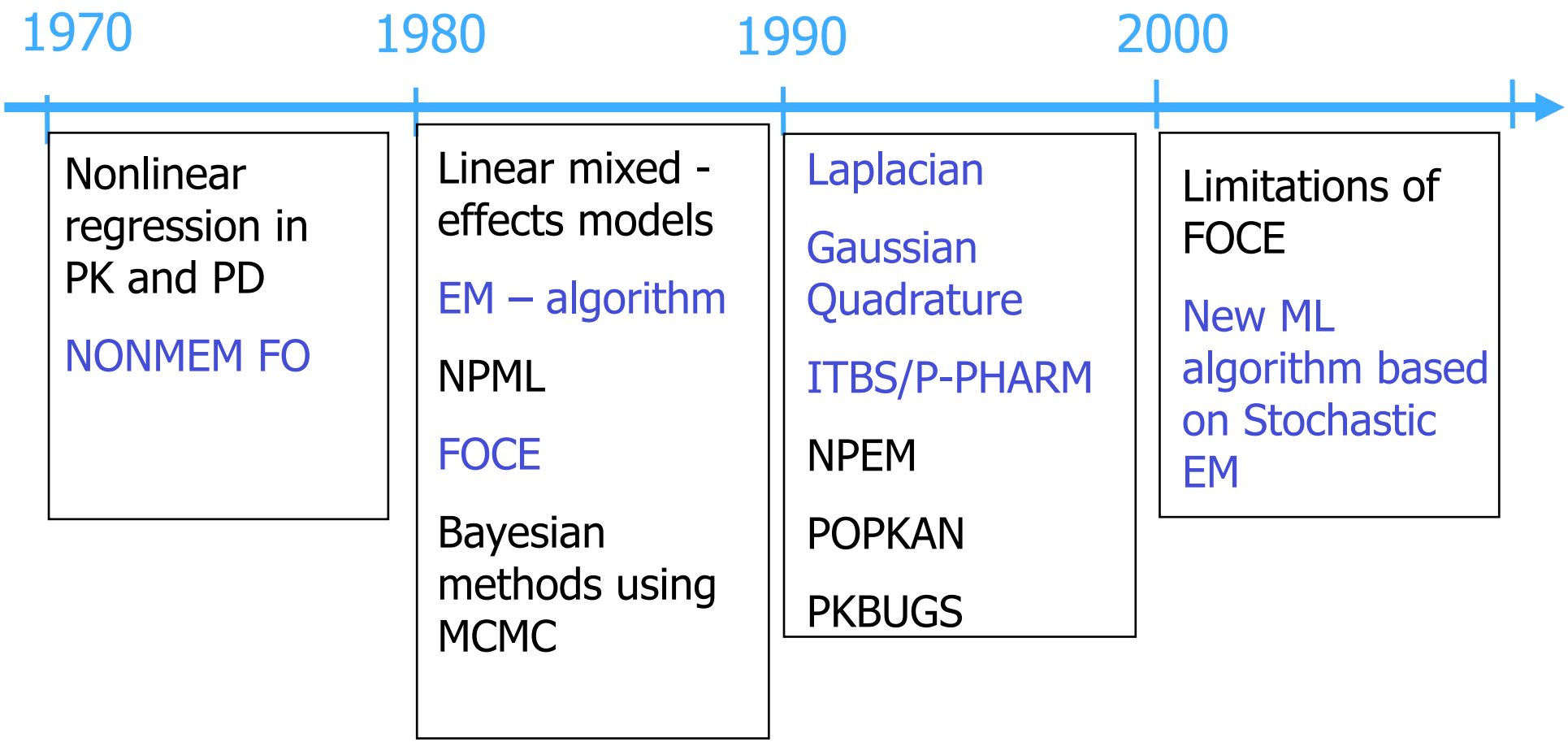
1. 85 – 90: FOCE + other approaches: nonparametric, Bayesian

2. The 90's: new software, growing interest, new statistical developments, limitations of FOCE

3. Since 00: Stochastic methods for parametric ML estimation + ...

Software for estimation in nonlinear mixed-effects models

	Maximum likelihood	Bayesian estimation
Parametric	NONMEM WinNonMix nlme (R and Splus) Proc NLMIXED (SAS) PPharm MONOLIX (SAEM) S-ADAPT (MCPEM) PDX-MCPEM	PK BUGS
Nonparametric	NPML NPEM (USC*PACK) NONMEM	Dirichlet process



Pillai, Mentré, Steimer (2005). Non-linear mixed effects modeling - from methodology and software development to driving implementation in drug development science. *J Pharmacokin Pharmacodyn*, 32:161-83.

The FO and FOCE methods

- First and most popular methods for estimation of population parameters by maximum likelihood in NLMEM

- FO: First order linearisation of the model around random effects = 0

- FOCE: First order linearisation of the model around current estimates of random effects

Implemented in NONMEM, WinNonMix, nlme (R and Splus), Proc NLMIXED (SAS)

Limitations of FO and FOCE

- FO
  - assume that mean response = response for mean parameters
  - not true for nonlinear models!!
  - ➢ Bias if "not very small" inter-individual variability
- FOCE
  - not consistent for sparse designs
  - very sensitive to initial estimates:
  - ➢ Lot's of run failed to converge, waste of time for modellers
- Both: Not real Maximum Likelihood Estimates (MLE)
  - good properties of MLE not demonstrated (LRT, standard errors from Fisher Information matrix, …)

Other approaches for computation of likelihood

- With approximation: linearisation using Laplace (NONMEM)
  - ➢ Similar problems of initial values than FOCE

  Wolfinger (1993). Laplace's approximation for nonlinear mixed models. *Biometrika*, 80:791-5.

- Integration of the likelihood by Adaptative Gaussian Quadrature (Proc NLMIXED in SAS)
  - ➢ Limited to models with small number of random effects

  Pinheiro & Bates (1995). Approximations to the Log-Likelihood function in the nonlinear mixed-effects model. *J Comput Graph Stat*, 1:12-35.

  Guedj, Thiebaut & Commenges (2007). Maximum likelihood estimation in dynamical models of HIV. *Biometrics*, 63: 1198-1206

3. Stochastic EM algorithms

EM algorithm

- Developed for MLE in problems with missing data
- Two steps algorithm
  - E-step: expectation of the log-likelihood of the complete data
  - M-step: maximisation of the log-likelihood of the complete data
- Mixed-effects models
  - individual random-effects = missing data

Dempster, Laird & Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm, *JRSS B*, 1:1-38.
Lindstrom & Bates (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data, *JASA*, 83:1014-22

EM in NLMEM

- Problem in EM for NLMEM
  - no analytical solution for integral in E-step

1. Linearisation around current estimates of random effects (PPharm, ITS)
   - ➢ Similar problems for sparse design than FOCE

   Mentré & Gomeni (1995). A two-step algorithm for estimation on non-linear mixed-effects with an evaluation in population pharmacokinetics. *J Biopharm Stat*, 5:141-158.

2. Full stochastic E-step
   - ➢ Can be very time consuming, not in available software

   Walker (1986). An EM algorithm for nonlinear mixed effects models, *Biometrics*, 52:934-3944.

Stochastic EM in NLMEM

3. MCPEM (in S-ADAPT and PDX-MCPEM): Monte Carlo integration during the E step using importance sampling around current individual estimates

   Bauer & Guzy (2004). Monte Carlo Parametric Expectation Maximization Method for Analyzing Population PK/PD Data. In: D'Argenio DZ, ed. *Advanced Methods of PK and PD Systems Analysis*. pp: 135-163.

4. SAEM (in MONOLIX): Decomposition of E-step in 2 steps
   - S-step: simulation of individual parameters using MCMC
   - SA-step: stochastic approximation of expected likelihood

   Delyon, Lavielle & Moulines (1999). Convergence of a stochastic approximation version of the EM procedure. *Ann Stat*, 27: 94-128.

SAEM in NLMEM

- Do not compute integral of E-step at each iteration
  - less time consuming than MCPEM
- Good statistical properties clearly demonstrated

  Delyon, Lavielle & Moulines (1999). Convergence of a stochastic approximation version of the EM procedure. *Ann Stat*, 27: 94-128.

  Kuhn, Lavielle (2004). Coupling a stochastic approximation version of EM with a MCMC procedure. *ESAIM Prob & Stat*, 8: 115-131.

  Kuhn & Lavielle (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Comput Stat Data Analysis,* 49: 1020-1038.

  Samson, Mentré, Lavielle (2007). The SAEM algorithm for group comparison tests in longitudinal data analysis based on nonlinear mixed-effects model. *Stat Med*, 26: 4860-4875.
- Addition of a Simulated Annealing algorithm to converge more quickly around the MLE
  - robust with respect to choice of initial estimates
  - fast

Recent extensions of the SAEM algorithm

- Correct handling of BQL data

Samson, Mentré & Lavielle (2006). Extension of the SAEM algorithm to left-censored data in nonlinear mixed effects models: application to HIV dynamic data. *Comput Stat Data Analysis,* 51: 1562-1574,

- Models defined by ODE or SDE

Donnet, Samson (2007). Estimation of parameters in incomplete data models defined by dynamical systems. *J Stat Plan Infer*, 137:2815-2831

Donnet, Samson (2008). Parametric inference for mixed models defined with stochastic differential equations. *ESAIM Prob & Stat*, 12: 196-218

- REML Estimation

Meza, Jaffrézic, Foulley (2007). REML estimation of variance parameters in nonlinez miwed effects models using the SAEM algorithm. *Biometrical J*, 49:867-888

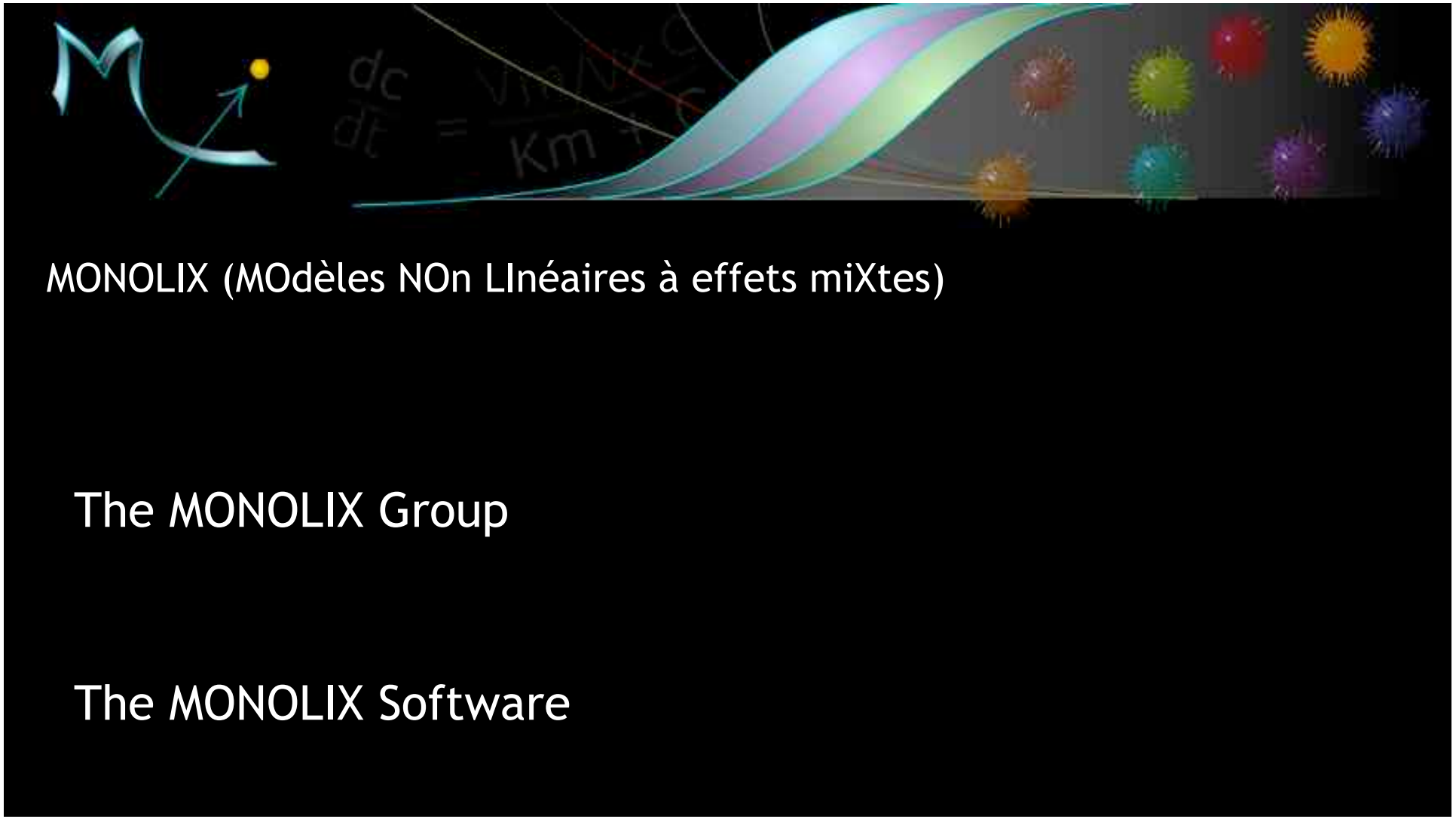
- Inter-occasion variability

Panhard, Samson (2008). Extension of the SAEM algorithm for nonlinear models with two levels of random effects. *Biostatistics* (in press)

- Binary data

Meza, Jaffrezic, Foulley (2008). Estimation in the probit normal model for binary outcomes using the SAEM alogorithm. (in revision)

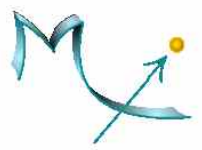
4. MONOLIX
www.monolix.org



The MONOLIX group

- Multi-disciplinary group (Pr M Lavielle & F Mentré)
  - created in october 2003
  - meets every month to exchange and develop activities in the field of mixed effect models
  - interest in both in the study and applications of these models
- Involves scientists with varied backgrounds
  - academic statisticians from several universities of Paris (theoretical developments),
  - researchers from INSERM (applications in pharmacology)
  - researchers from INRA (applications in agronomy, animal genetics)
  - Scientists from from the medical faculty of Lyon-Sud University (applications in oncology)

The MONOLIX software (1)



- SAEM algorithm for estimation in NLMEM
- Open-source software
  - MATLAB
  - Stand alone version for models in library
  - Rich PKPD library
  - ODE models in MATLAB or C++
  - MLXTRAN for complex models definition
  - Correct handling of LOQ data
  - Estimation of inter-occasion variability
  - Friendly Graphical User Interface
  - Graphical outputs (GOF, VPC, …)
- Supported by several ingeneers from INRIA (Institut National de la Recherche en Informatique et Automatique, France)

The MONOLIX software (2)

- MONOLIX 1: v1.1 Feb 2005
- MONOLIX 2: v2.1 April 2007
  - v2.3 November 2007 – May 2008
    - C++ package for ODE models
    - Categorical covariates
    - Several distribution for random effects
    - Inter-occasion variability
  - v2.4 September 2008 (beta version in June 2008)
    - MLXTRAN
    - Extension of PKPD library (3 cpt PK, effect compartment)
- MONOLIX 3: new MONOLIX project
  - Support from INRIA and several drug companies

Algorithms in MONOLIX

Intensive use of powerful and well-known algorithms in the MONOLIX software:

- **Estimation of the population parameters:** Maximum likelihood estimation with the SAEM (Stochastic Approximation of EM) algorithm, combined with MCMC (Markov Chain Monte Carlo) and Simulated Annealing,
- **Estimation of the individual parameters:** Estimation/Maximization of the conditional distributions with MCMC,
- **Estimation of the objective (likelihood) function:** Monte Carlo and minimum variance Importance Sampling,
- **Model selection and assessment:** Information criteria (AIC, BIC), Statistical Tests (LRT, Wald test), Goodness of fit plots (Individual fits, Weighted residuals, NPDE, VPC,...).

Estimation & outputs: without linearisation

- Estimation of all components of variability (even small) and their standard errors
- Estimation of individual random effects
  - From simulated posterior/conditional distribution
  - Mean, Var and Mode without approximation
- Likelihood estimated by importance sampling
- Population and Individual residuals
  - From simulated marginal and posterior distributions
- No real importance of shrinkage

5. Comparison of Stochastic EM algorithms to NONMEM

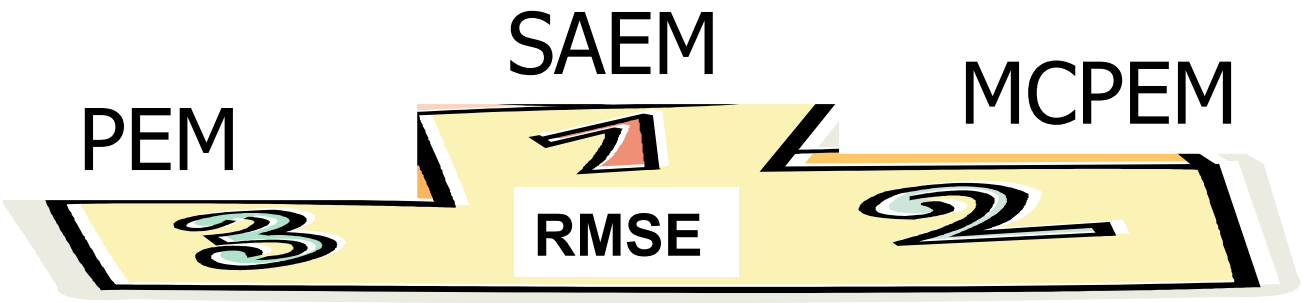
- Girard & Mentré PAGE 2005
  - 100 replicated simulated data sets: one PK and one PD
  - NONMEM V & VI, MCPEM, SAEM (blind evaluation)
- Bauer, Guzy & Chee AAPS J 2007
  - Four simulated examples: PK or PKPD
  - NONMEM VI, PDx-MCPEM, S-ADAPT, MONOLIX 1.1
- Bazzoli, Retout & Mentré PAGE 2007
  - 1000 replicated simulated data sets: one PKPD model
  - NONMEM V, MONOLIX 2.1
- Laveille, Lavielle, Chatel & Jacqmin PAGE 2008
  - 150 simulated examples: PK (linear & nonlinear)
  - NONMEM V & VI, MONOLIX 2.2 & 2.3
- …

Main conclusions from comparisons

S-ADAPT (MCPEM) and MONOLIX (SAEM)

- Never failed to provide results whatever the models
  - in replicated simulated data sets, often NONMEM results on several data sets are missing
- Much faster than NONMEM FOCE for complex ODE models
- Better results (bias, RMSE) than NONMEM FOCE in sparse designs
- Applied successfully to real complex PKPD data sets (where NONMEM failed to converge)

Simulated sparse design PK example
Classification of methods based on RMSE of the 10 population parameters (%)

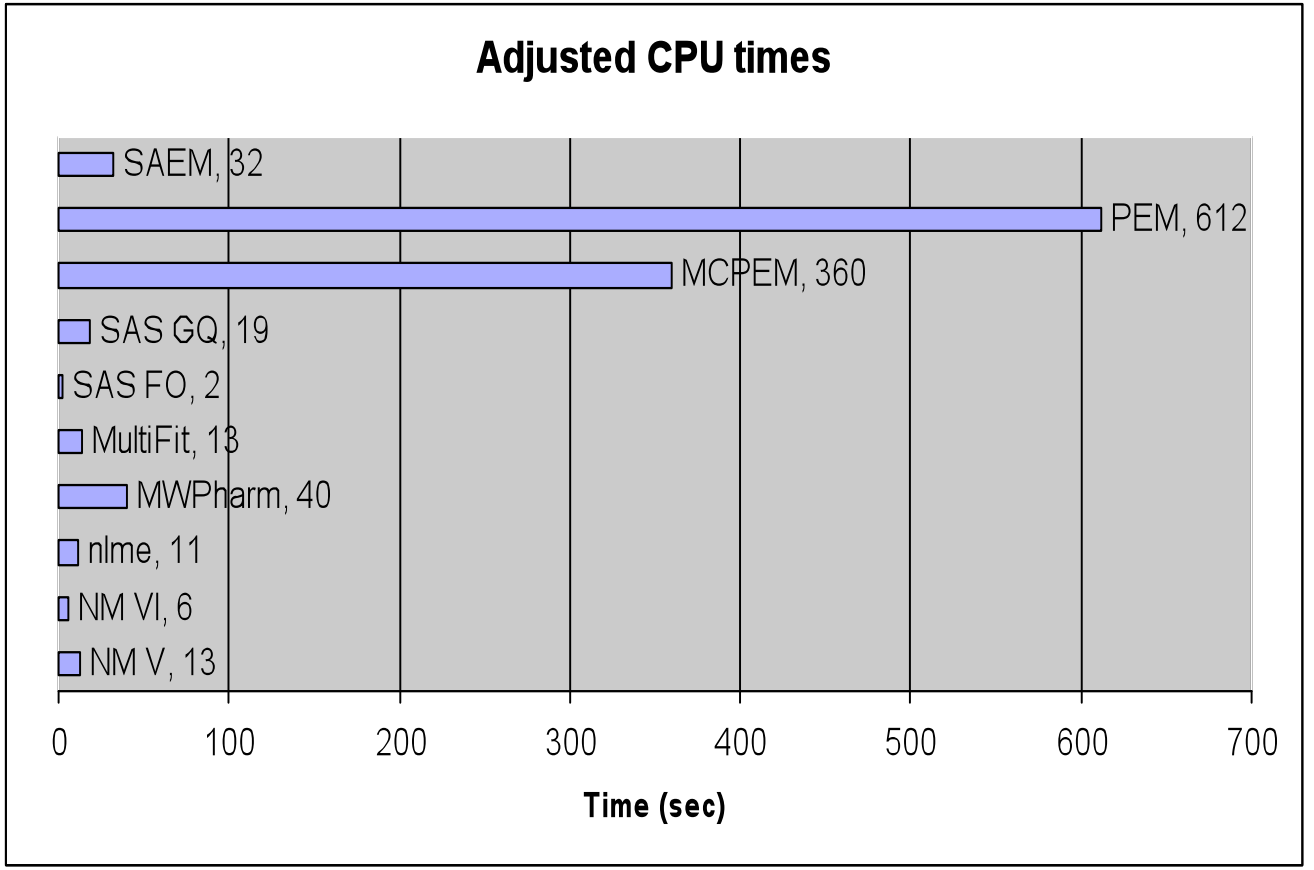


	1	2	3	4	5	6	9	10
	SAEM	MC PEM	PEM	SAS AGQ	NM V	NM VI	nlme	SAS FO
Pts	89	78	73	70	64	59	44	17
Run	100	100	100	100	100*	49	88	100

Girard & Mentré, PAGE 2005

Simulated sparse design PK example
Adjusted CPU times for one dataset

	GHz	CPU time for 1 run (sec)
SAEM[1]	1.6	20
PEM	1.7	360
MCPEM	1	360
SAS GQ	3.1	6
SAS FO	3.1	1
Nlme [2]	1.6	7
NM VI [3]	2.13	3
NM V [3]	2.13	6



1. A compiled version would probably be considerably faster than actual implementation in Matlab for PEM and SAEM
2. For a successful nls + nlme run
3. For a successful NONMEM run (EST & COV)

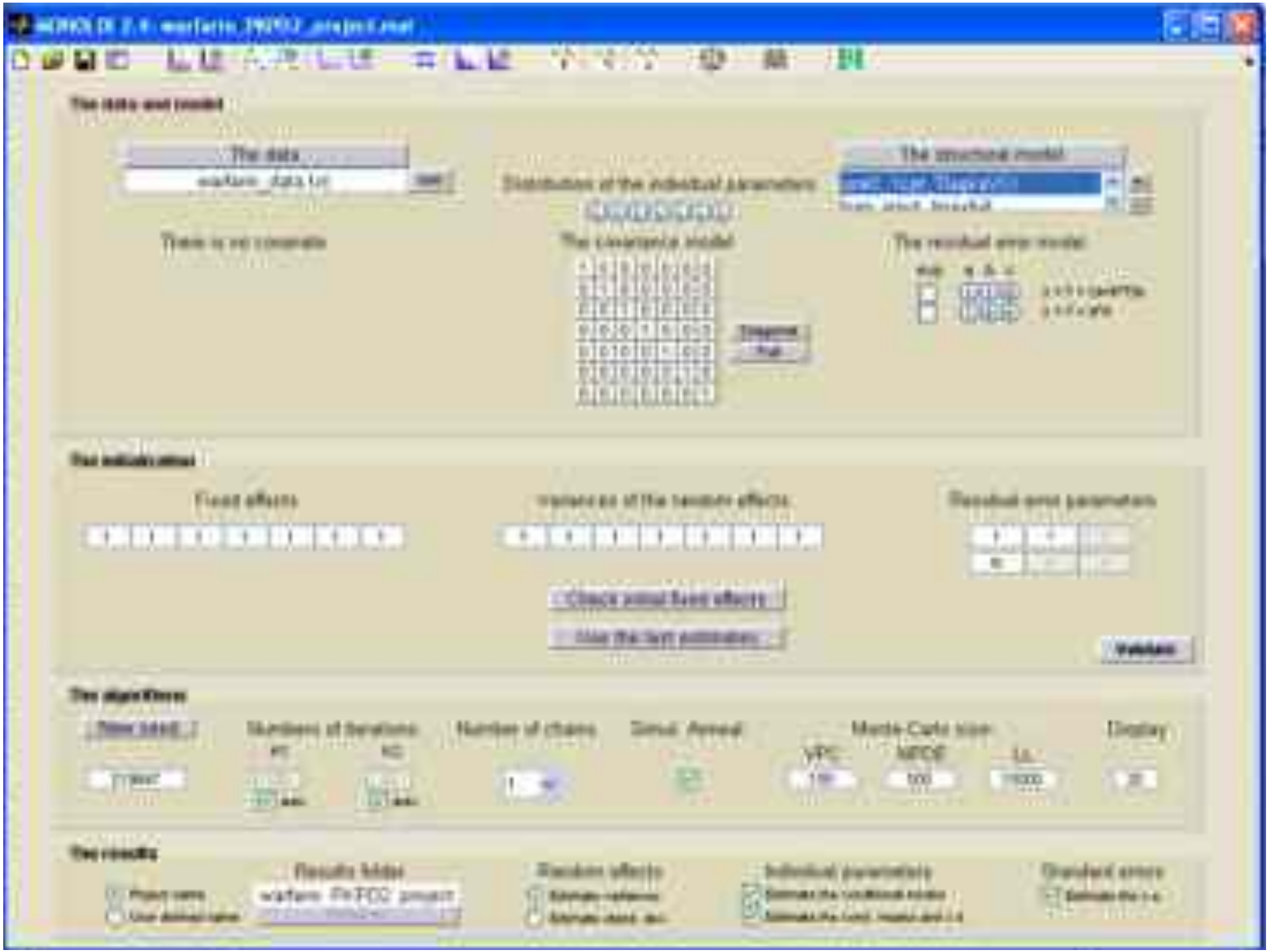
Evaluation of PK library in MONOLIX 2.3
CPU times

	Single Dose 120 subjects 2280 observations	Multiple Doses (7 doses) 120 subjects 5520 observations
1 compartment model IV bolus *linear elimination* *non linear elimination*	 7" 34"	 25" 1' 36"
2 compartments model 1st order oral absorption **lag time** *linear elimination* *non linear elimination*	 15" 2' 18"	 51" 6'52"

Laveille & Lavielle, MONOLIX website, PAGE 2008

6. PKPD example with MONOLIX (v 2.4)

- PKPD analysis of warfarine
  - 32 healthy volunteers
  - 1.5 mg/kg single dose
  - PK: total racemic warfarin plasma concentration
  - PD: prothrombin complex activity (PCA)
  - 250 concentrations, 232 PCA
- Models
  - PK: 1 cp, first order absorption with lag time
    Tlag, ka, V, CL
  - PD: turnover model with inhibition of input by warfarine
    Rin, kout, Imax, C50



Estimation of the population parameters

	parameter	s.e.
tlag :	0.683	0.238
ka :	0.772	0.115
V :	7.91	0.211
Cl :	0.132	0.00591
Imax :	1.11	0.0364
C50 :	1.72	0.225
Rin :	4.47	0.201
kout :	0.0465	0.00196
a_1 :	0.303	0.0479
b_1 :	0.0529	0.00934
c_1 :	1	-
a_2 :	3.84	0.226
b_2 :	0	-
c_2 :	1	-

Estimation of the population parameters

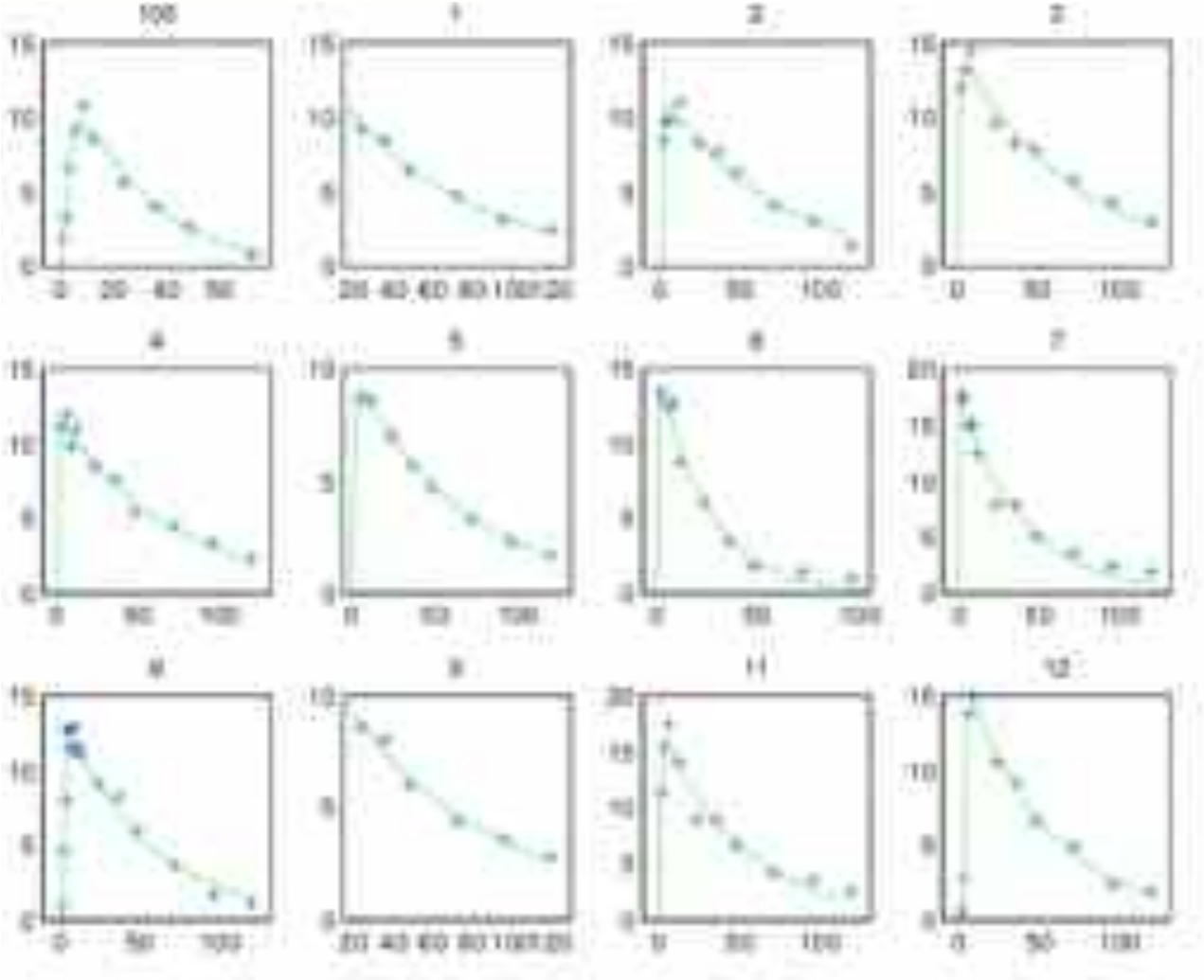
	parameter	s.e.
omega_tlag :	0.707	0.23
omega_ka :	0.885	0.249
omega_V :	0.225	0.031
omega_Cl :	0.285	0.0369
omega_Imax :	0.0374	0.018
omega_C50 :	0.392	0.0621
omega_Rin :	0.0487	0.018
omega_kout :	0.0316	0.0204

Estimation by linearization

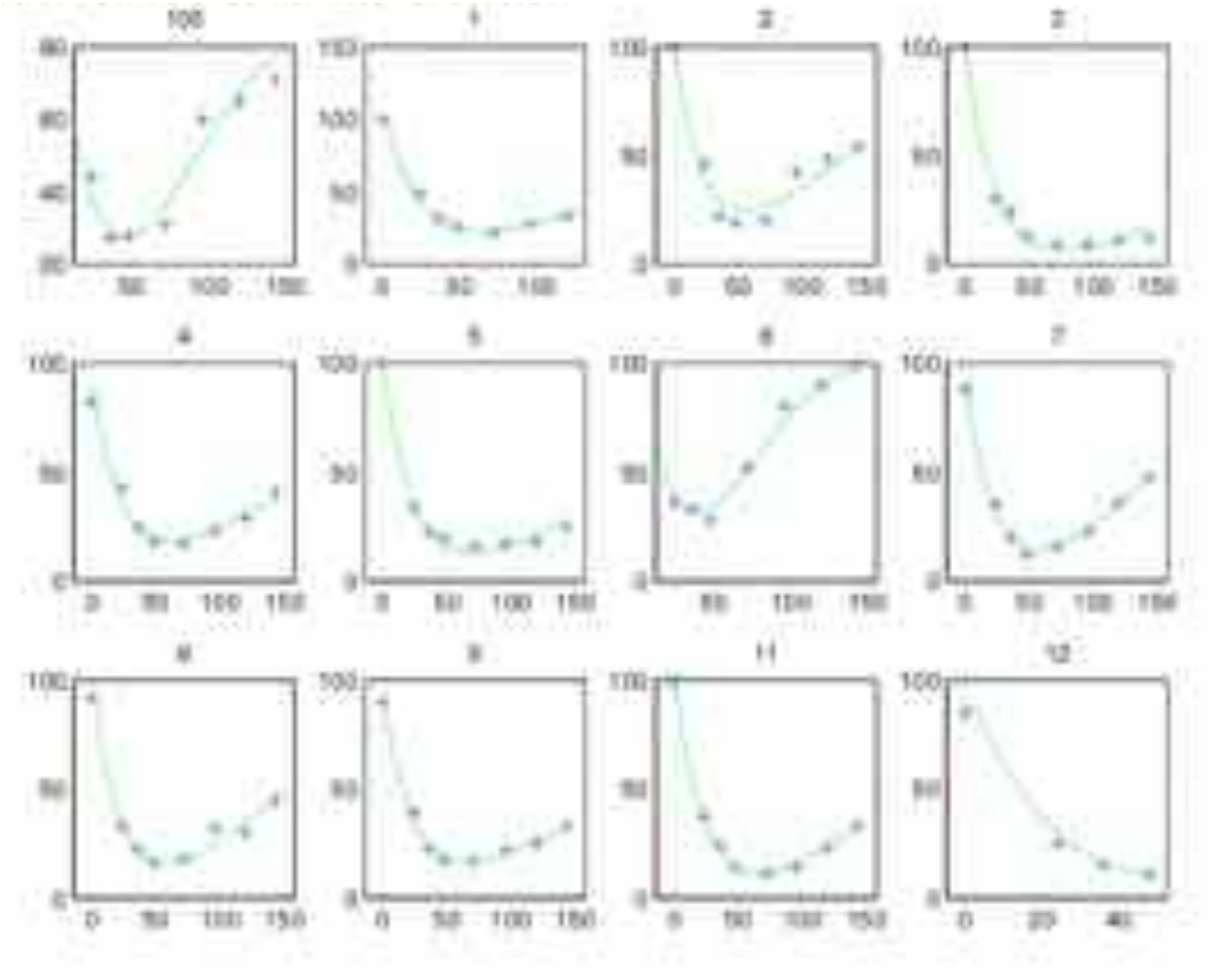
-2 x log-likelihood:	2143.77
Akaike Information Criteria:	2173.77
Bayesian Information Criteria:	2195.76

Elapsed time is 485.8 seconds.
CPU time is 464 seconds.

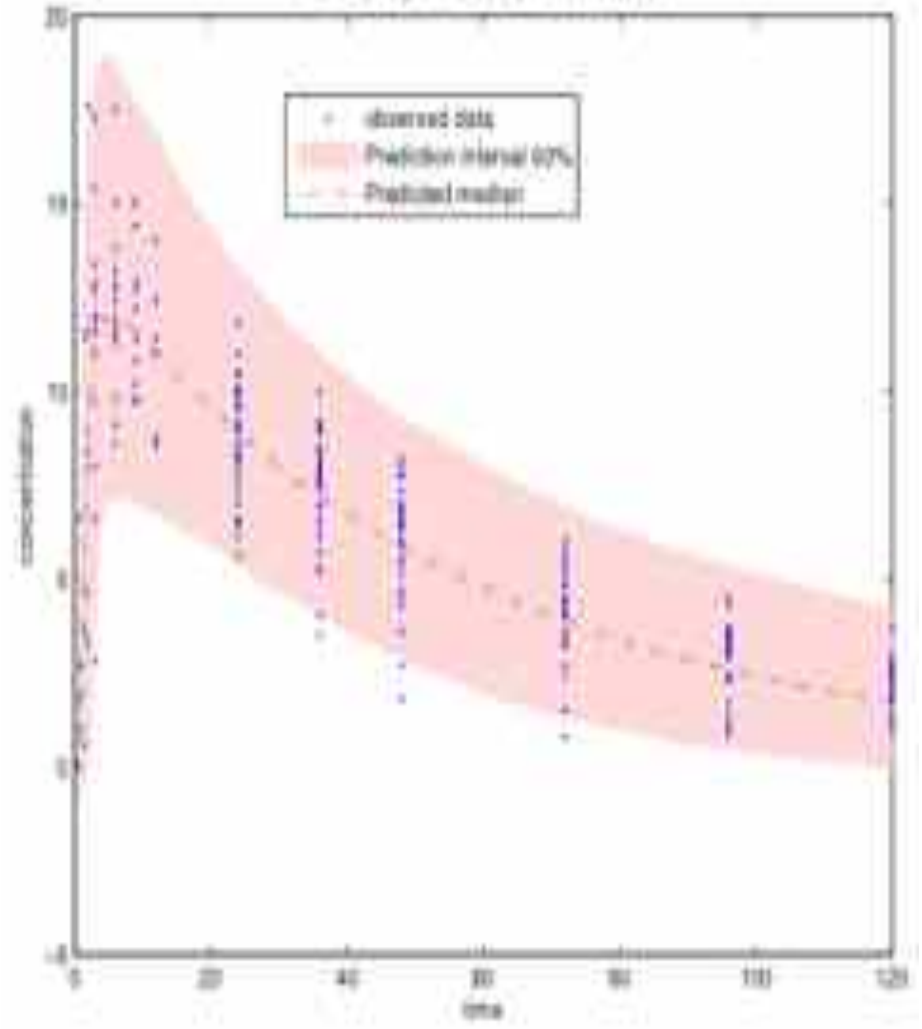
Some individual PK fits

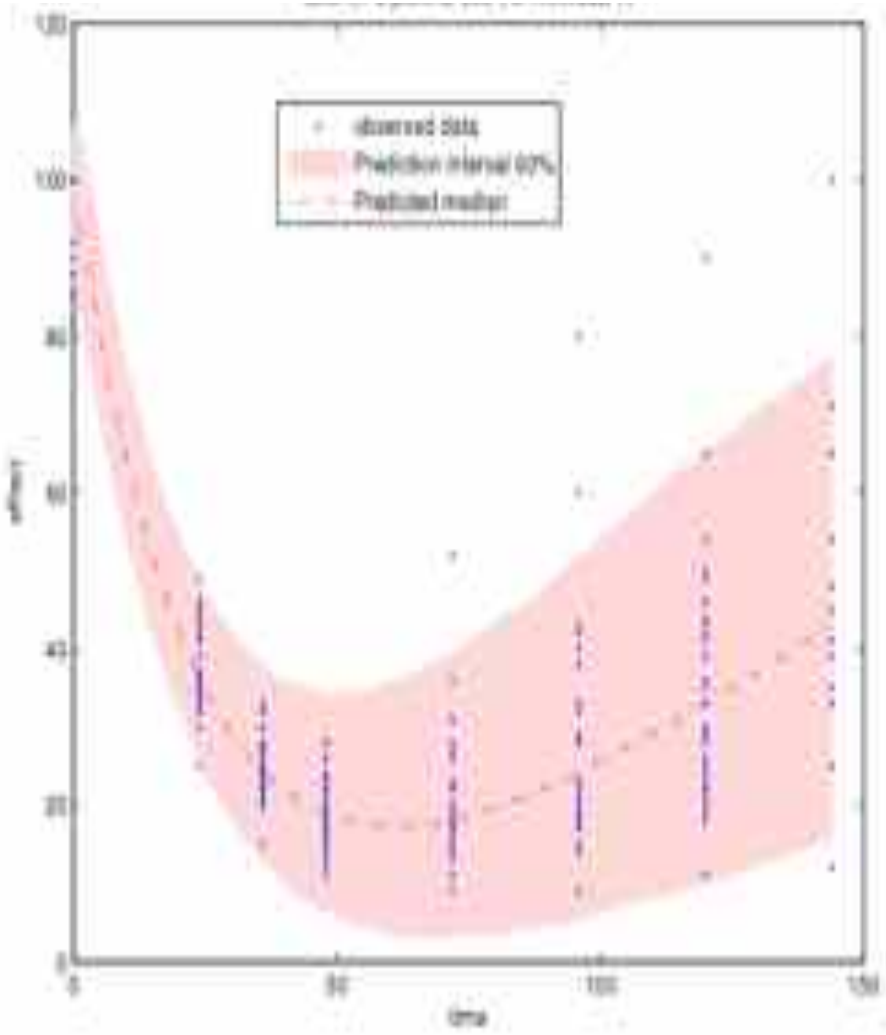


Some individual PD fits



VPC provided by MONOLIX
Warfarin PKPD data





MLXTRAN specification

```
$PROBLEM Turnover model PKPD model
$PSI Tlag ka V Cl Imax C50 Rin kout

$PK
ALAG1 = Tlag
KA1 = ka

$ODE
A_0(1) = 0
A_0(2) = Rin/kout
k = Cl/V
Cc = A(1)/V
DADT(1) = -k*A(1)
DADT(2) = Rin*(1-Imax*Cc/(Cc+C50))-kout*A(2)

$OUTPUT
OUPUT1 = Cc
OUTPUT2= A(2)
```

7. Conclusion (1)

- NLMEM: good approach for PKPD analysis of preclinical studies with rich or sparse designs
  - accurate estimation of all components of variability
- NLMEM applied for increasingly complex dynamic models
- Drug companies mostly used NONMEM
  - FOCE developed 15 years ago: several drawbacks
- New method based on AGQ (in SAS)
  - limited to problems of small dimension
- New MLE methods based on stochastic EM developed by statisticians
  - fast, consistent, no linearization, …
  - SAEM cleverly used the iteration process

7. Conclusion (2)

- *New software / algorithms should be used*
- *Extensions of MONOLIX, NONMEM… are ongoing*
  - MONOLIX
    - User friendly, graphical outputs
    - Open source
    - Several statistical extensions planed in next releases
    - Based on thorough and published statistical methods
  - NONMEM
    - MCPEM & SAEM available in NONMEM VII
    - Implemented by B. Bauer (S-ADAPT)