Bayesian Estimation in mixed effects models using Monolix and mlxR

Marc Lavielle[1]

[1]Inria Saclay

Bayes 2015
Basel, Novartis Campus

Outline

What is a model?

A joint probability distribution!

$\mathrm{p}(y, \psi, \theta)$: joint probability distribution of y, ψ and θ.

- $y = (y_{ij}, 1 \leq i \leq N, 1 \leq j \leq n_i)$ are the observations,
- $\psi = (\psi_i, 1 \leq i \leq N)$ are the individual parameters,
- θ is the vector of population parameters.

What is a joint probability distribution?
A hierarchical model

$$\mathrm{p}(y, \psi, \theta) = \mathrm{p}(y|\psi, \theta)\mathrm{p}(\psi|\theta)\mathrm{p}(\theta)$$

- $\mathrm{p}(y|\psi, \theta)$: conditional distribution of the observations,
- $\mathrm{p}(\psi|\theta)$ population distribution of the individual parameters,
- $\mathrm{p}(\theta)$ the distribution of θ can represent
  - the *inter population variability*,
  - the *uncertainty* on θ,
  - the *prior distribution* of the population parameters.

Some tasks

$$\mathrm{p}(y, \psi, \theta) = \mathrm{p}(y|\psi, \theta)\mathrm{p}(\psi|\theta)\mathrm{p}(\theta)$$

1) **Estimation of the population parameters** (only y is given)

- *Maximum Likelihood approach*: maximize $\mathrm{p}(y|\theta) = \int \mathrm{p}(y, \psi|\theta)\, d\psi$
- *Bayesian approach*: compute/maximize

$$\mathrm{p}(\theta|y) = \frac{\mathrm{p}(y|\theta)\mathrm{p}(\theta)}{\mathrm{p}(y)}$$

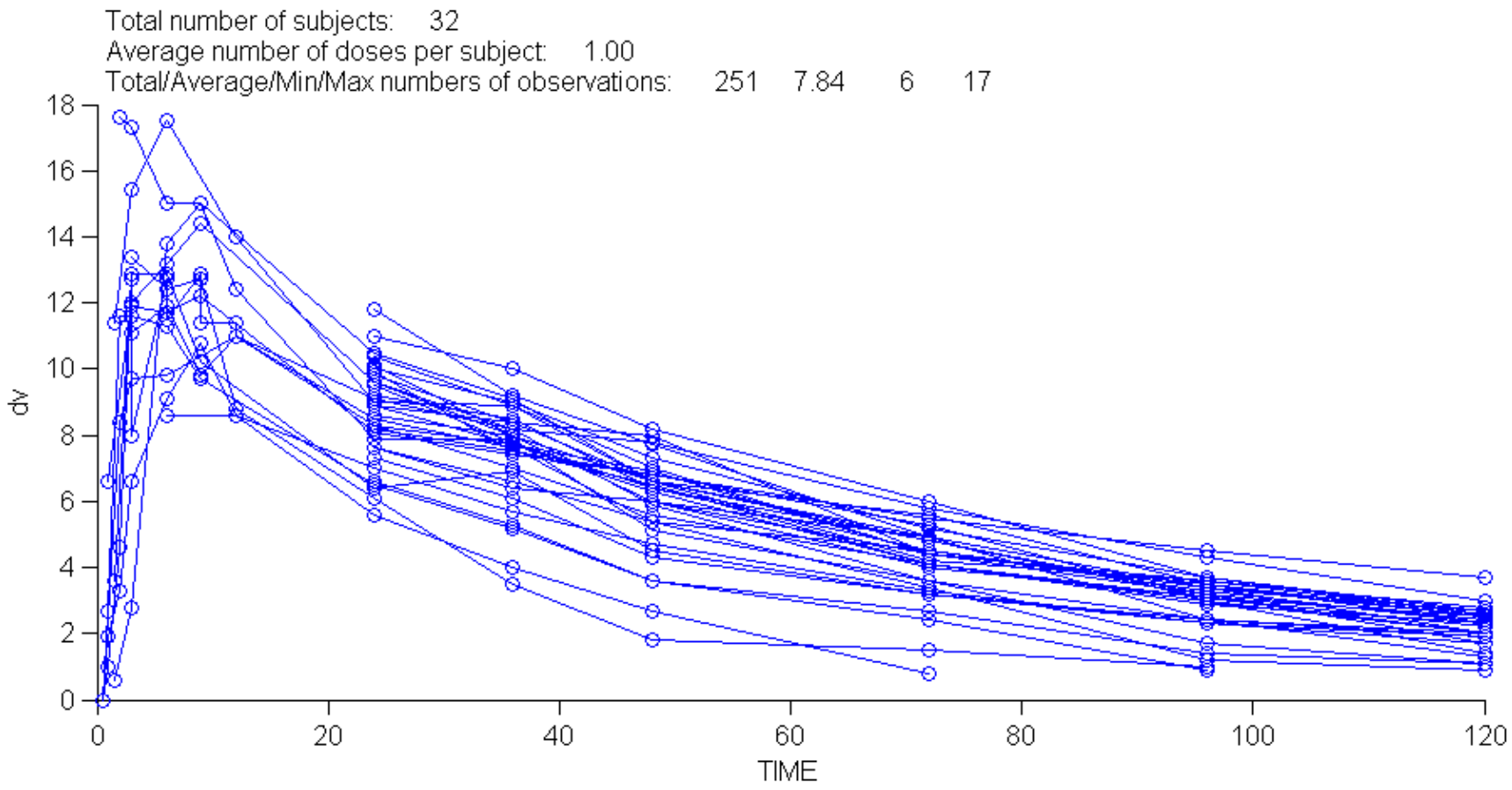
2) **Estimation of the individual parameters** (y and θ are given)

- *Individual approach*: maximize $\mathrm{p}(y|\psi, \theta)$
- *Population approach*: compute/maximize

$$\mathrm{p}(\psi|y, \theta) = \frac{\mathrm{p}(y|\psi, \theta)\mathrm{p}(\psi|\theta)}{\mathrm{p}(y, \theta)}$$

A PK example

The warfarin PK data:



A PK example

- The PK model:

$$C(t, \psi) = \frac{D\, k_a}{V\, k_a - Cl} \left(e^{-(Cl/V)\, t} - e^{-k_a\, t} \right).$$

- The model for the observed concentrations:

$$y_{ij} = C(t_{ij}, \psi_i) + a\varepsilon_{ij}$$

  i.e.

$$y_{ij} | \psi_i, a \sim \mathcal{N}(C(t_{ij}, \phi_i), a^2)$$

- The model for the individual PK parameters $\psi_i = (k_{a_i}, V_i, Cl_i)$:

$$\begin{aligned}
\log(k_{a_i}) &\sim \mathcal{N}(\log(k_{a_{\text{pop}}}), \omega^2_{k_a}) \\
\log(V_i) &\sim \mathcal{N}(\log(V_{\text{pop}}), \omega^2_V) \\
\log(Cl_i) &\sim \mathcal{N}(\log(Cl_{\text{pop}}), \omega^2_{Cl}).
\end{aligned}$$

II
Estimation of the population parameters using Monolix

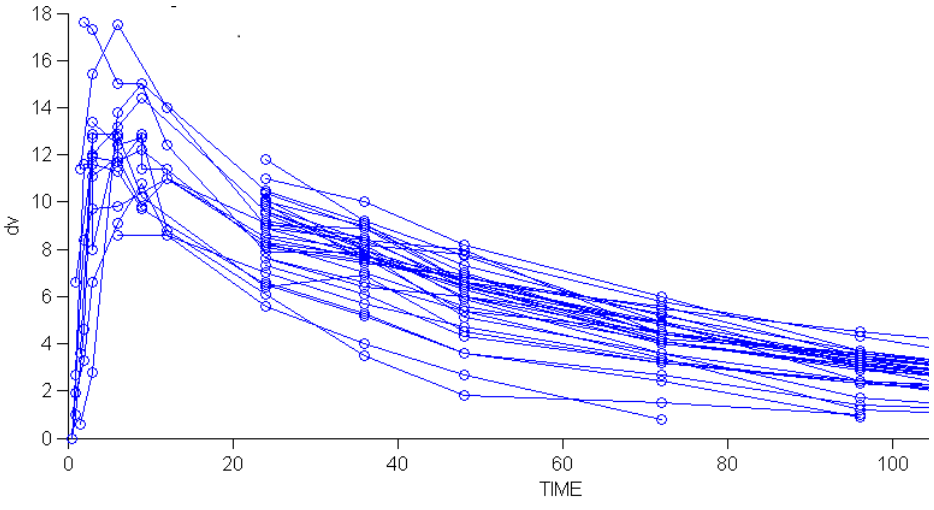
Estimation of the population parameters

With Monolix, a given parameter can be

- **a fixed constant** if we have absolute confidence in its value or the data does not allow it to be estimated, essentially due to lack of identifiability.
- **estimated by maximum likelihood**, either because we have great confidence in the data or no information on the parameter.
- **estimated by introducing a prior** and calculating the maximum a posteriori (MAP) estimate or estimating the posterior distribution.

Estimation of the population parameters
ML estimation

The warfarin PK data provides a limited information about *ka*



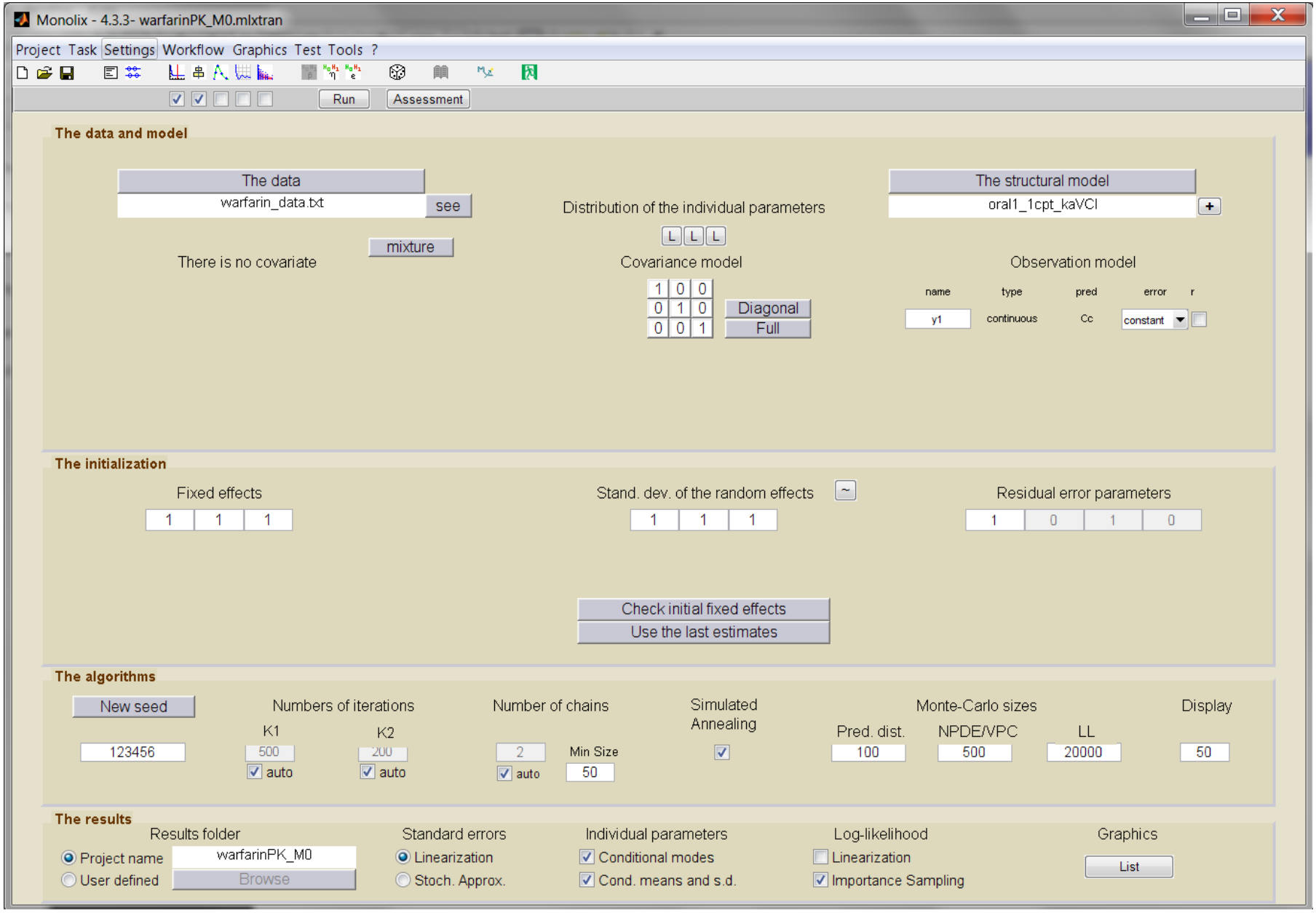
Estimation of the population parameters

		parameter	s.e. (lin)	r.s.e.(%)
ka_pop	:	0.587	0.12	20
V_pop	:	7.68	0.34	4
Cl_pop	:	0.135	0.0073	5
omega_ka	:	0.625	0.16	26
omega_V	:	0.221	0.036	16
omega_Cl	:	0.286	0.041	14
a	:	1.07	0.057	5

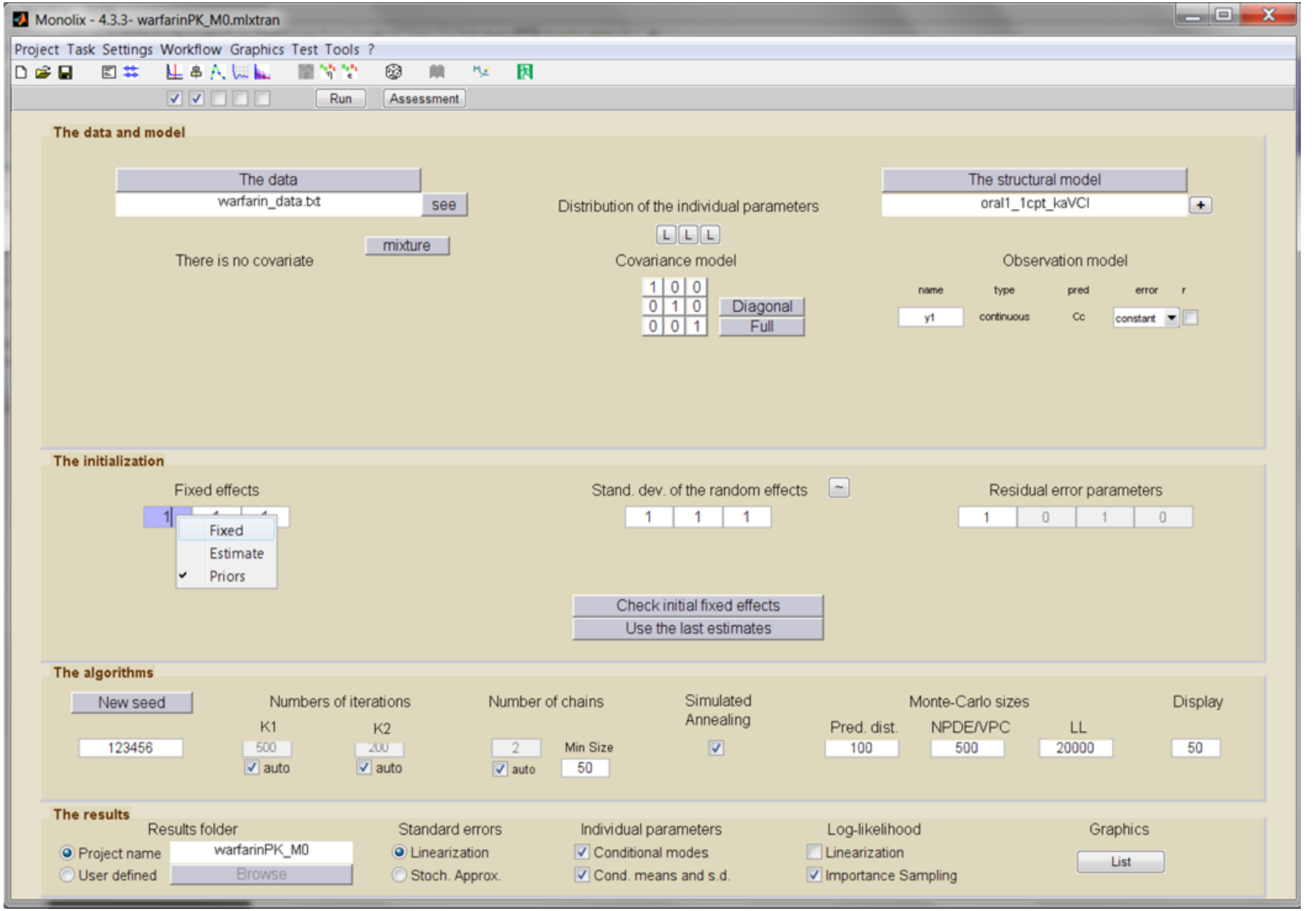
$\Rightarrow$ Let us introduce a prior for ka_{pop}:

$$\log(ka_{\text{pop}}) \sim \mathcal{N}(\log(ka^{\star}), \gamma^2)$$

Estimation of the population parameters

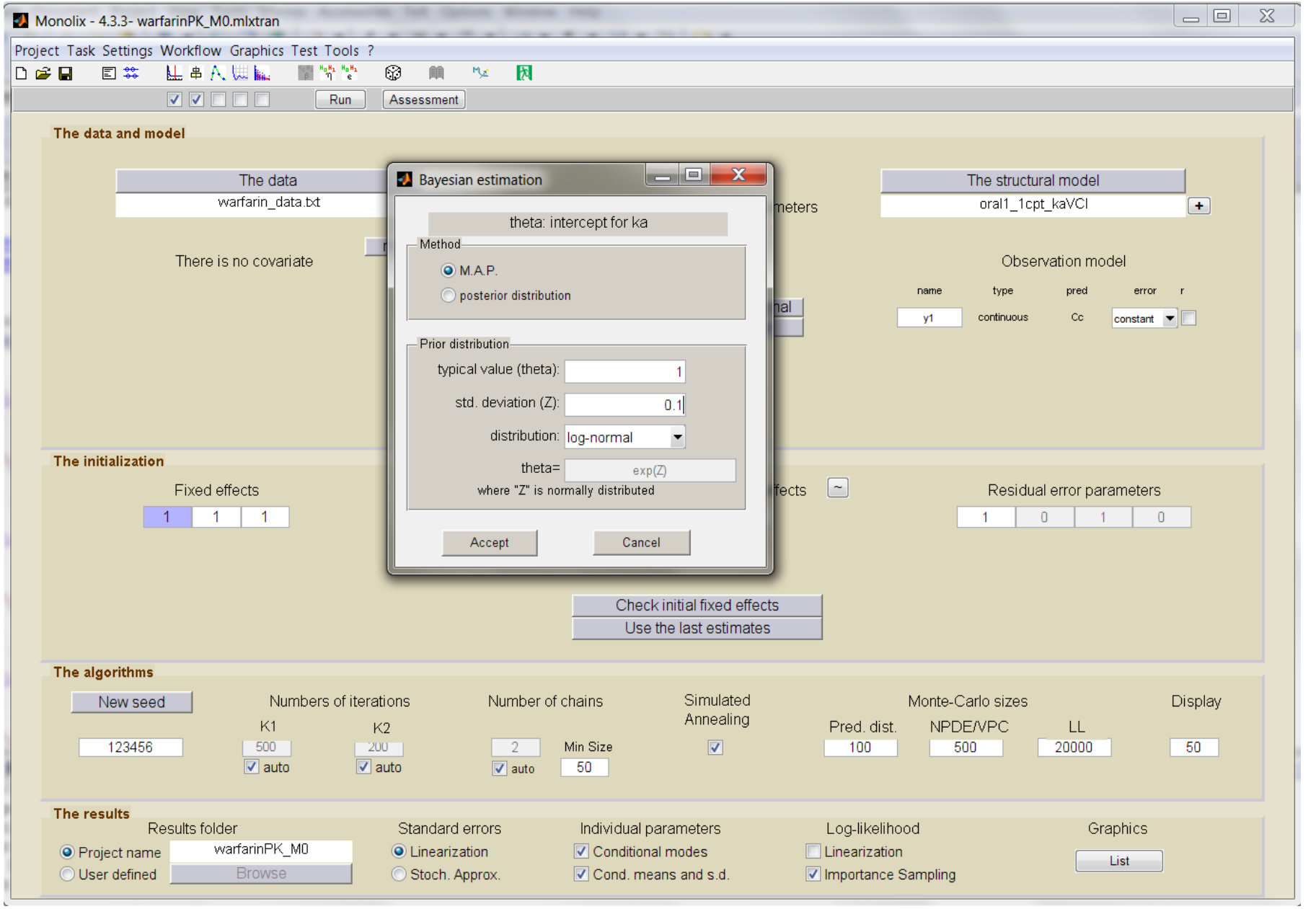


Estimation of the population parameters



Estimation of the population parameters

Combining ML and MAP estimations



Split θ into (θ_E, θ_M) where

- θ_E are the components of θ to be estimated with MLE,
- θ_M re the components of θ to be estimated with MAP.

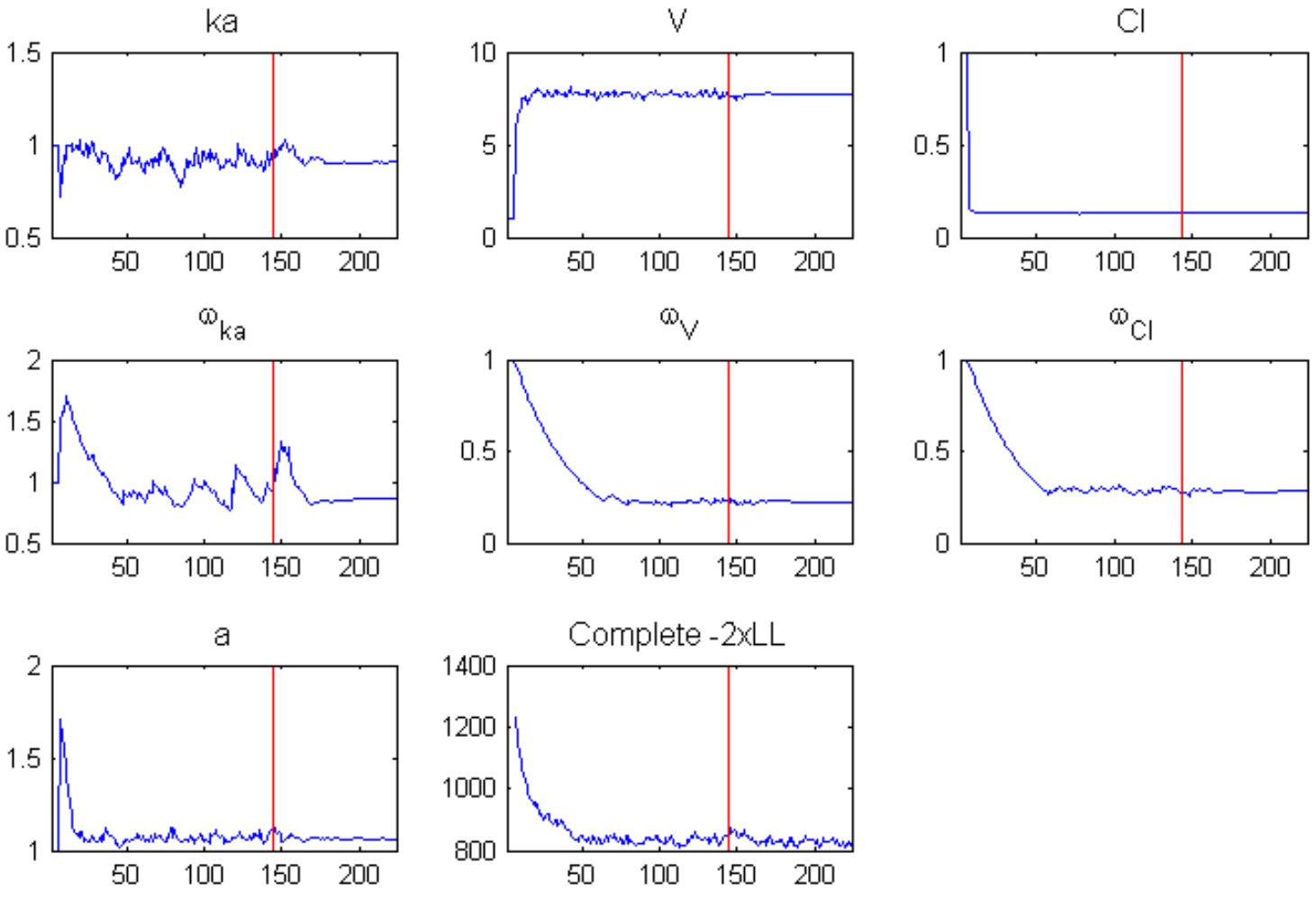
Then, $(\hat{\theta}_E, \hat{\theta}_M)$ maximizes the penalized likelihood of (θ_E, θ_M):

$$\begin{aligned}(\hat{\theta}_E, \hat{\theta}_M) &= \arg\max_{\theta_E, \theta_M} \log(\mathrm{p}(y, \theta_M; \theta_E)) \\ &= \arg\max_{\theta_E, \theta_M} \{\mathcal{LL}_y(\theta_E, \theta_M) + \log(\mathrm{p}(\theta_M))\},\end{aligned}$$

where $\mathcal{LL}_y(\theta_E, \theta_M) \stackrel{\text{def}}{=} \log(\mathrm{p}(y|\theta_M; \theta_E))$.

Estimation of the population parameters

Combining ML and MAP estimations



The MAP estimate of $\theta = (ka_{\text{pop}}, V_{\text{pop}}, \ldots)$ is the solution of the following minimization problem:

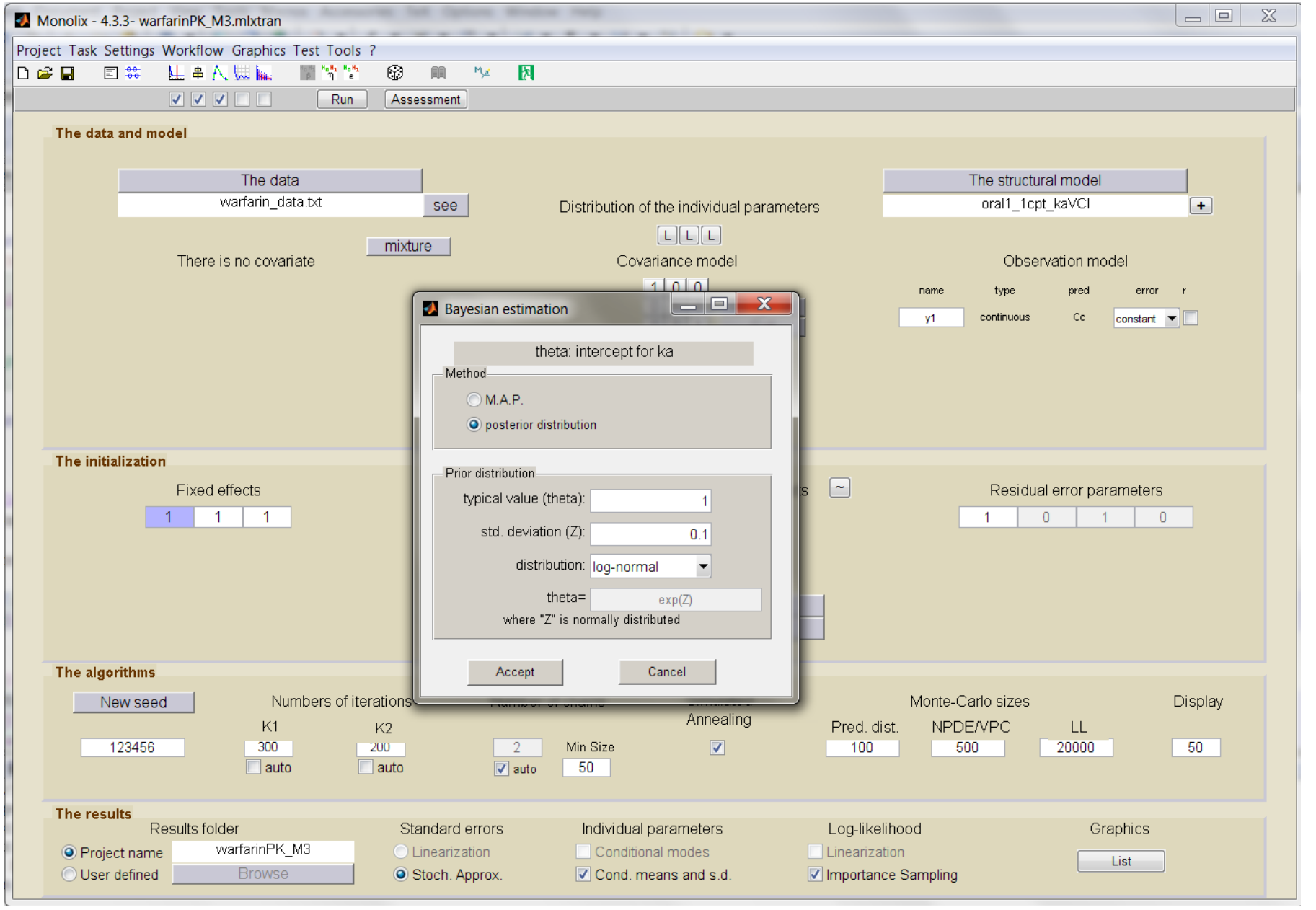
$$\hat{\theta}^{\text{MAP}} = \arg\min_{\theta} \left\{ -2\mathcal{LL}_y(\theta) + \frac{1}{\gamma^2}(\log(ka_{\text{pop}}) - \log(ka^\star)^2 \right\}.$$

We can compute the MAP estimate of ka_{pop} for various values of γ and $ka^\star = 1$:

γ	0	0.02	0.03	0.2	0.3	$+\infty$
$\hat{ka}_{\text{pop}}^{\text{MAP}}$	1	0.991	0.963	0.809	0.623	0.587

Estimation of the population parameters

Combining ML and posterior distribution estimation



Split θ into (θ_E, θ_R) where

- θ_E are the components of θ to be estimated with MLE,
- θ_R are the components of θ whose posterior distribution is estimated.

Then,

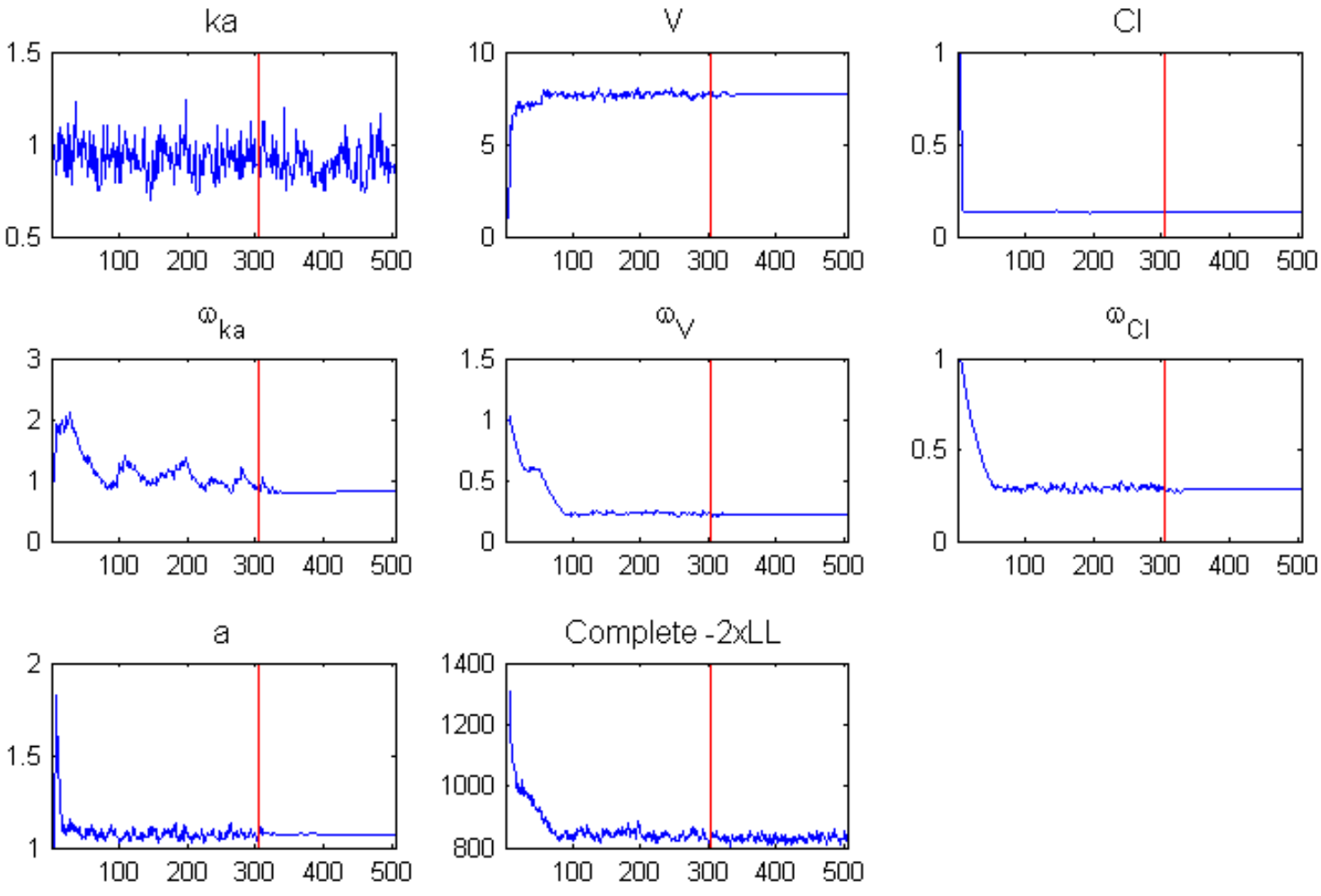
i) Compute the maximum likelihood of θ_E:

$$\begin{aligned} \hat{\theta}_E &= \arg\max_{\theta_E} \mathrm{p}(y; \theta_E) \\ &= \arg\max_{\theta_E} \int \mathrm{p}(y, \theta_R; \theta_E) d\theta_R. \end{aligned}$$

ii) Estimate the conditional distribution $\mathrm{p}(\theta_R | y; \hat{\theta}_E)$.

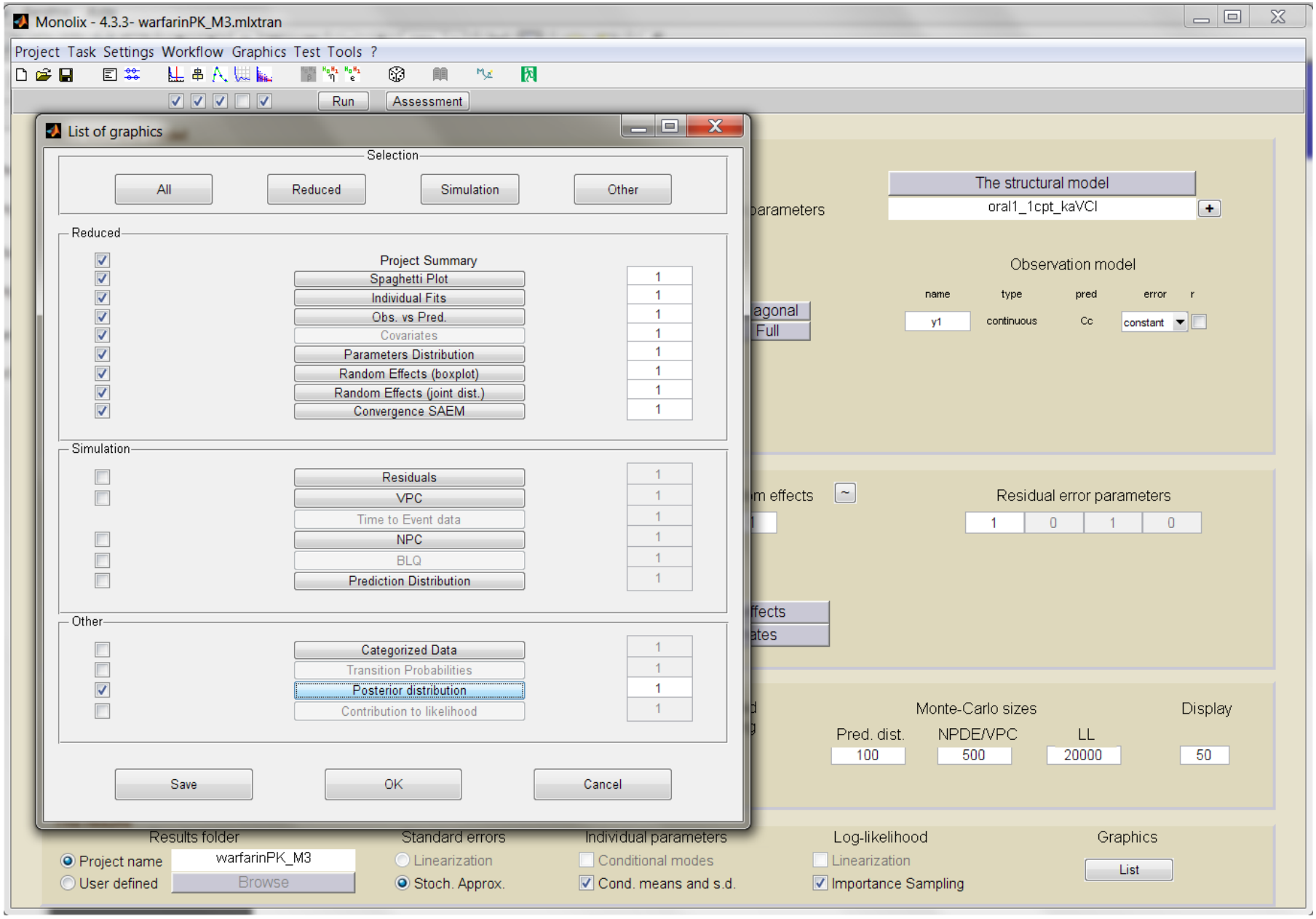
Estimation of the population parameters

Combining ML and posterior distribution estimation



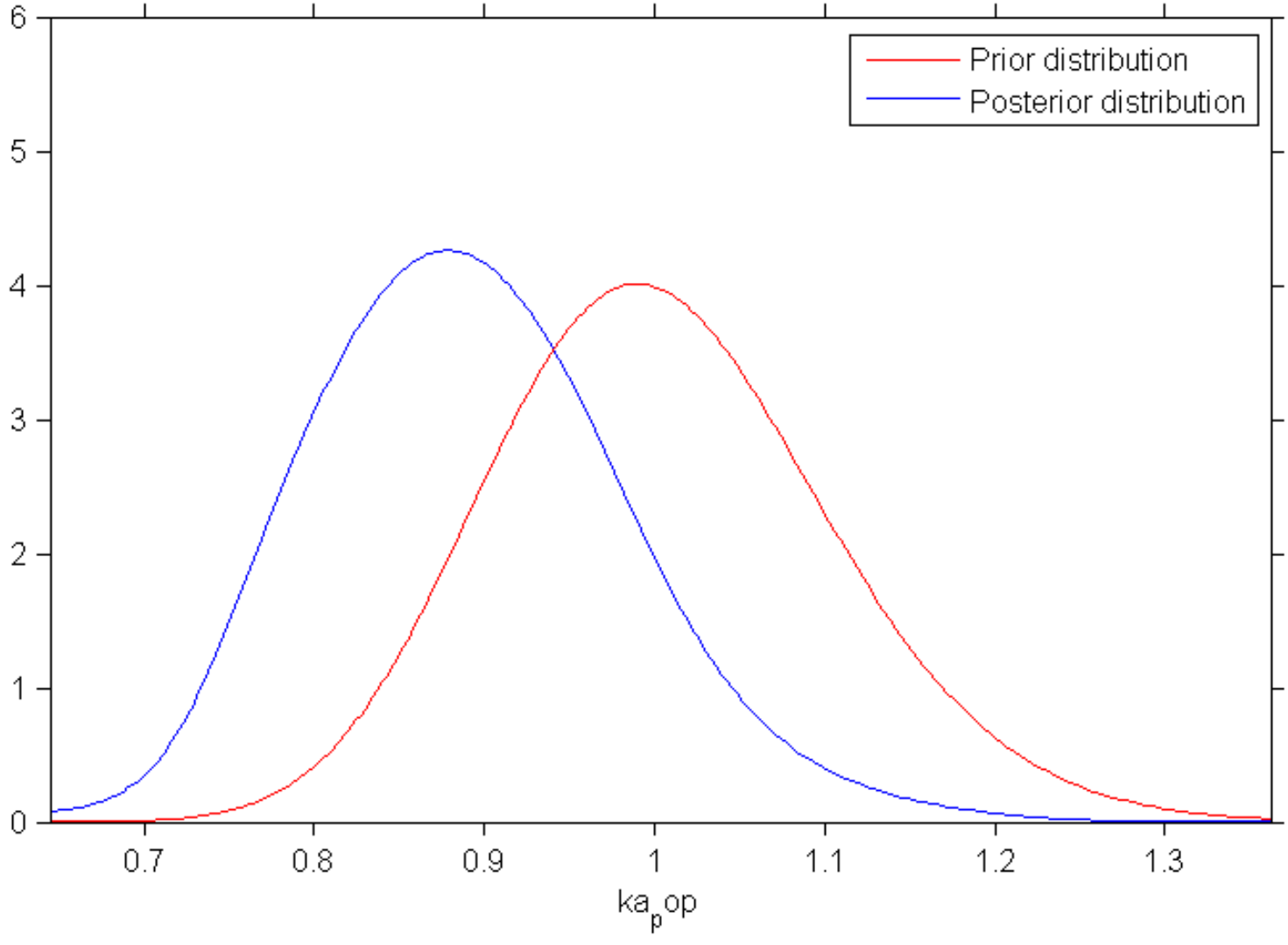
Estimation of the population parameters

Combining ML and posterior distribution estimation



Estimation of the population parameters

Combining ML and posterior distribution estimation



III
Estimation of the individual parameters using Monolix

Estimation of the individual parameters

Once θ has been estimated, Monolix allows one:

- to compute the maximum a posteriori (MAP) estimate of ψ_i, for $i = 1, 2, \ldots, N$

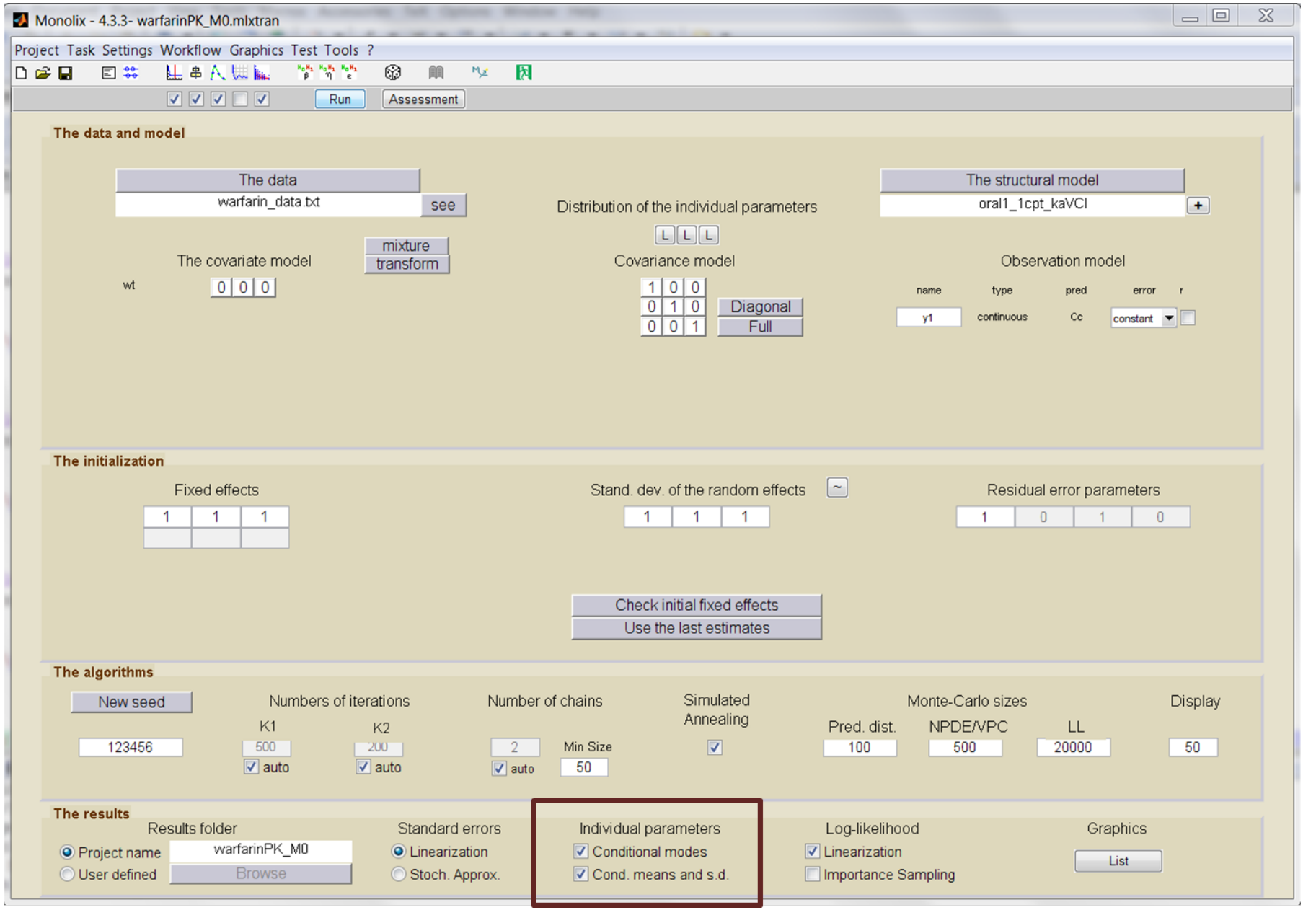
$$\hat{\psi}_i^{\text{mode}} = \arg\max_{\psi_i} \text{p}(\psi_i | y_i; \hat{\theta})$$

- to estimate the conditional distribution $\text{p}(\psi_i | y_i; \hat{\theta})$, for $i = 1, 2, \ldots, N$ using a Metropolis Hastings (MH) algorithm and estimate the conditional mean

$$\begin{aligned}\hat{\psi}_i^{\text{mean}} &= \widehat{\mathbb{E}(\psi_i | y_i; \hat{\theta})} \\ &= \frac{1}{K} \sum_{k=1}^{K} \psi_i^{(k)}.\end{aligned}$$

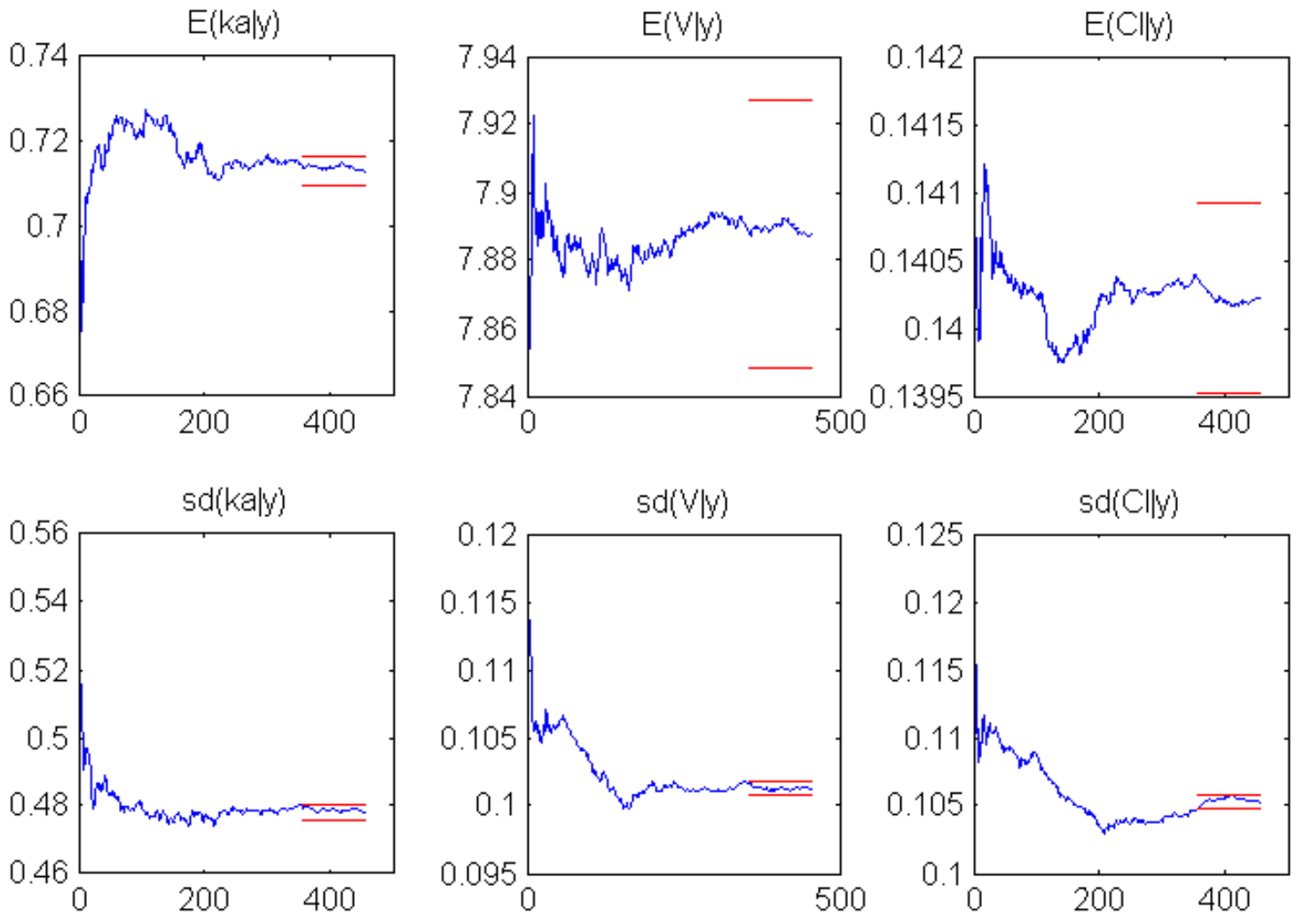
Estimation of the individual parameters

Conditional mode and conditional distribution



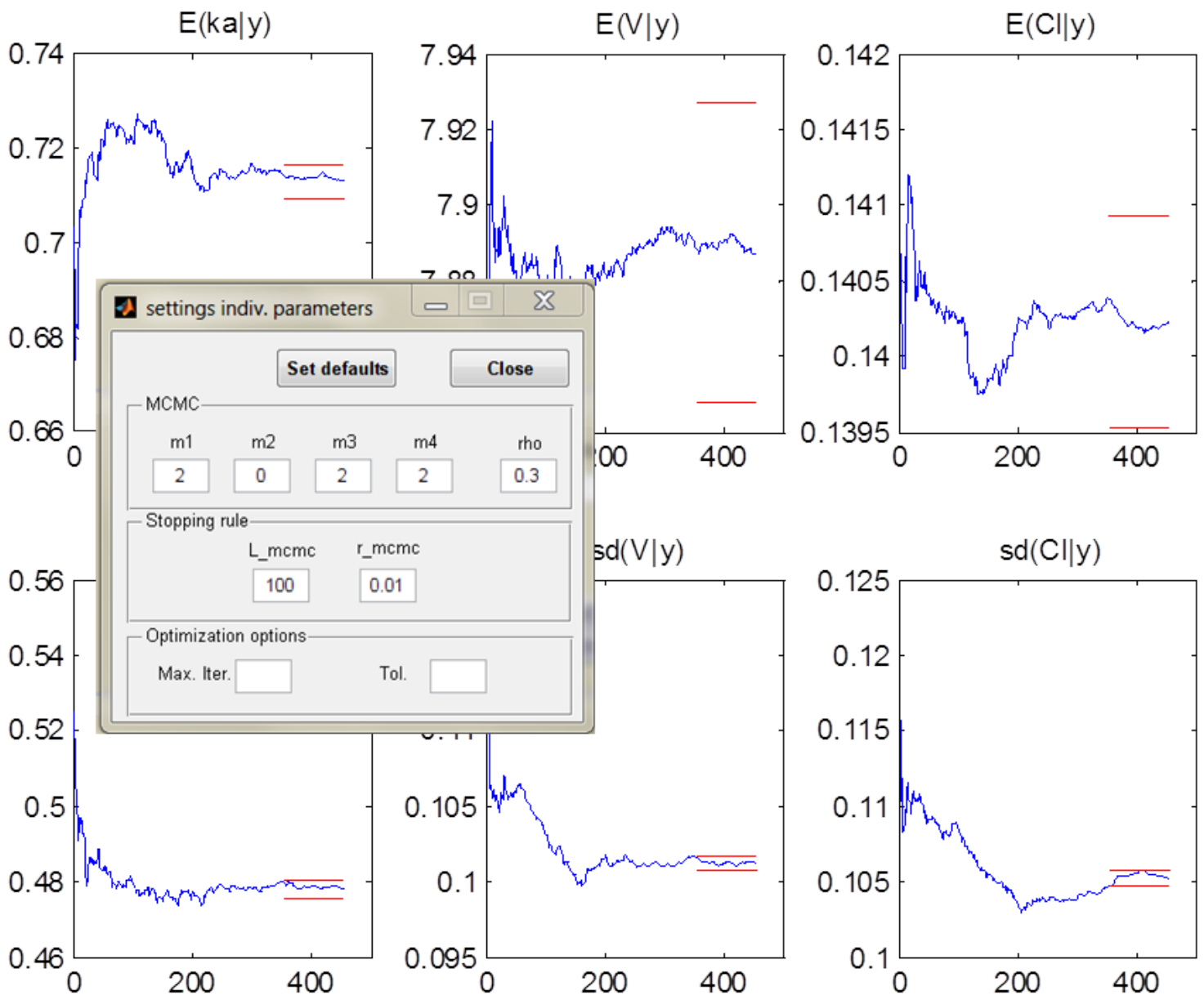
The Metropolis Hastings algorithm

Transitions kernels and stopping rule



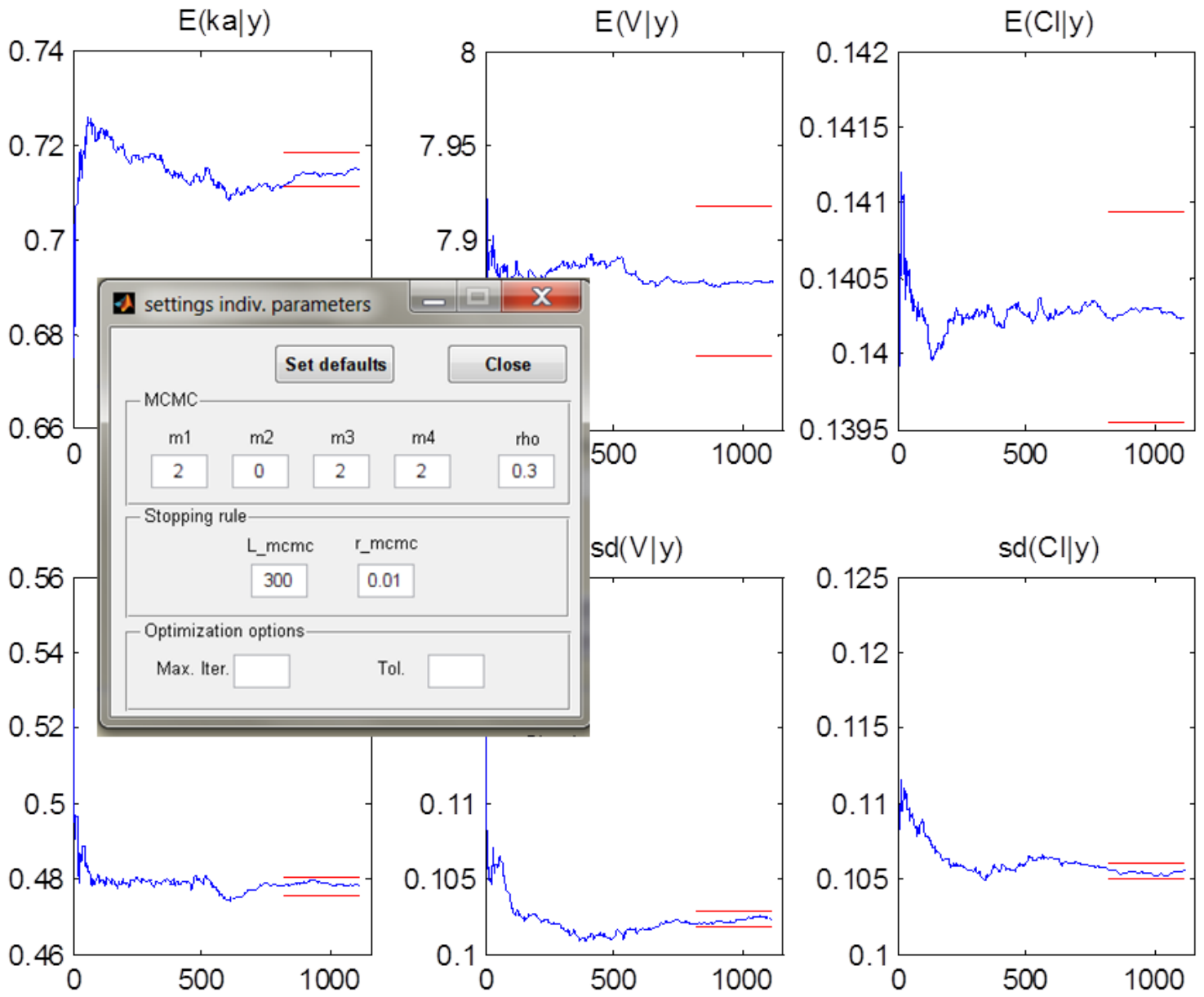
The Metropolis Hastings algorithm

Transitions kernels and stopping rule



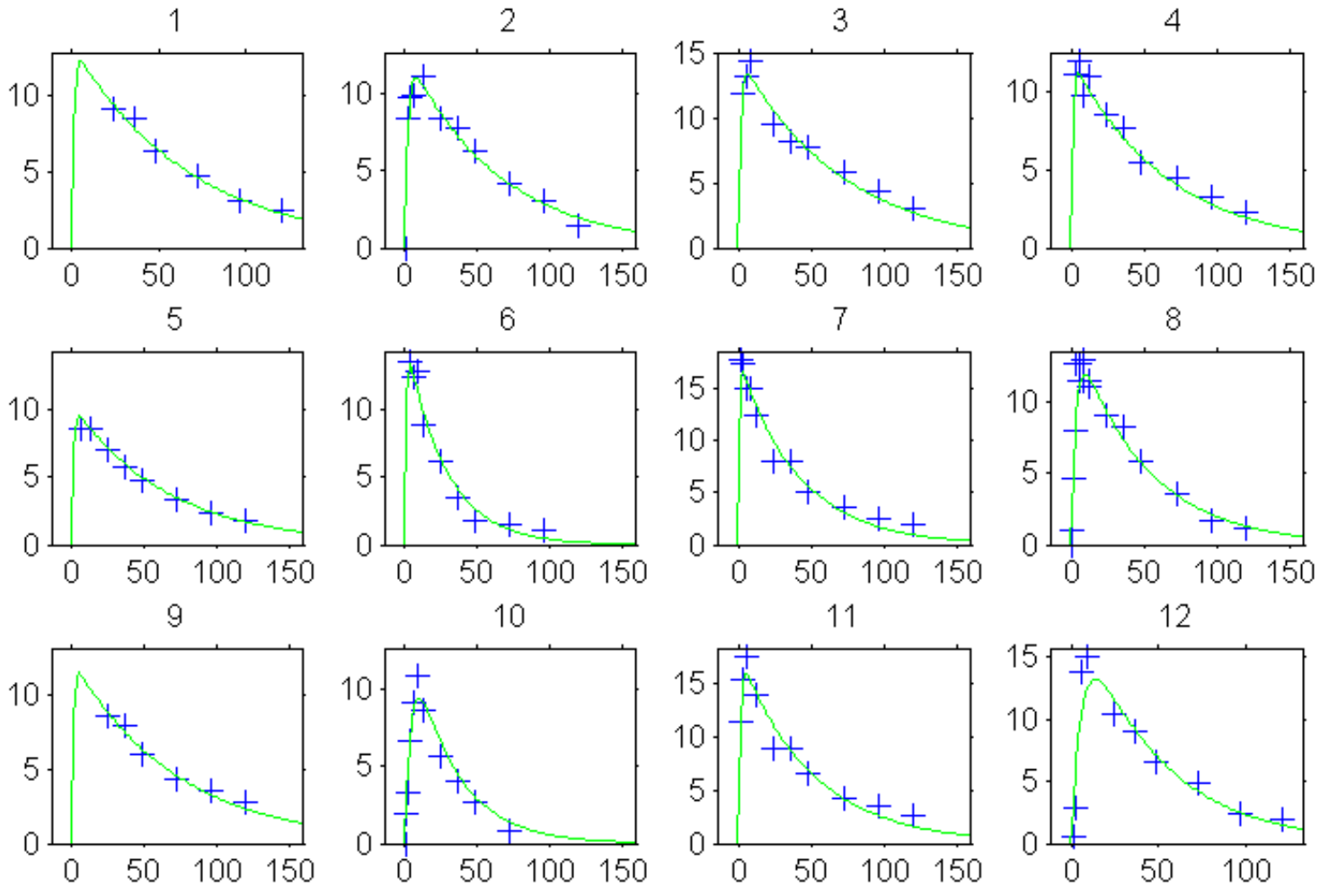
The Metropolis Hastings algorithm

Transitions kernels and stopping rule

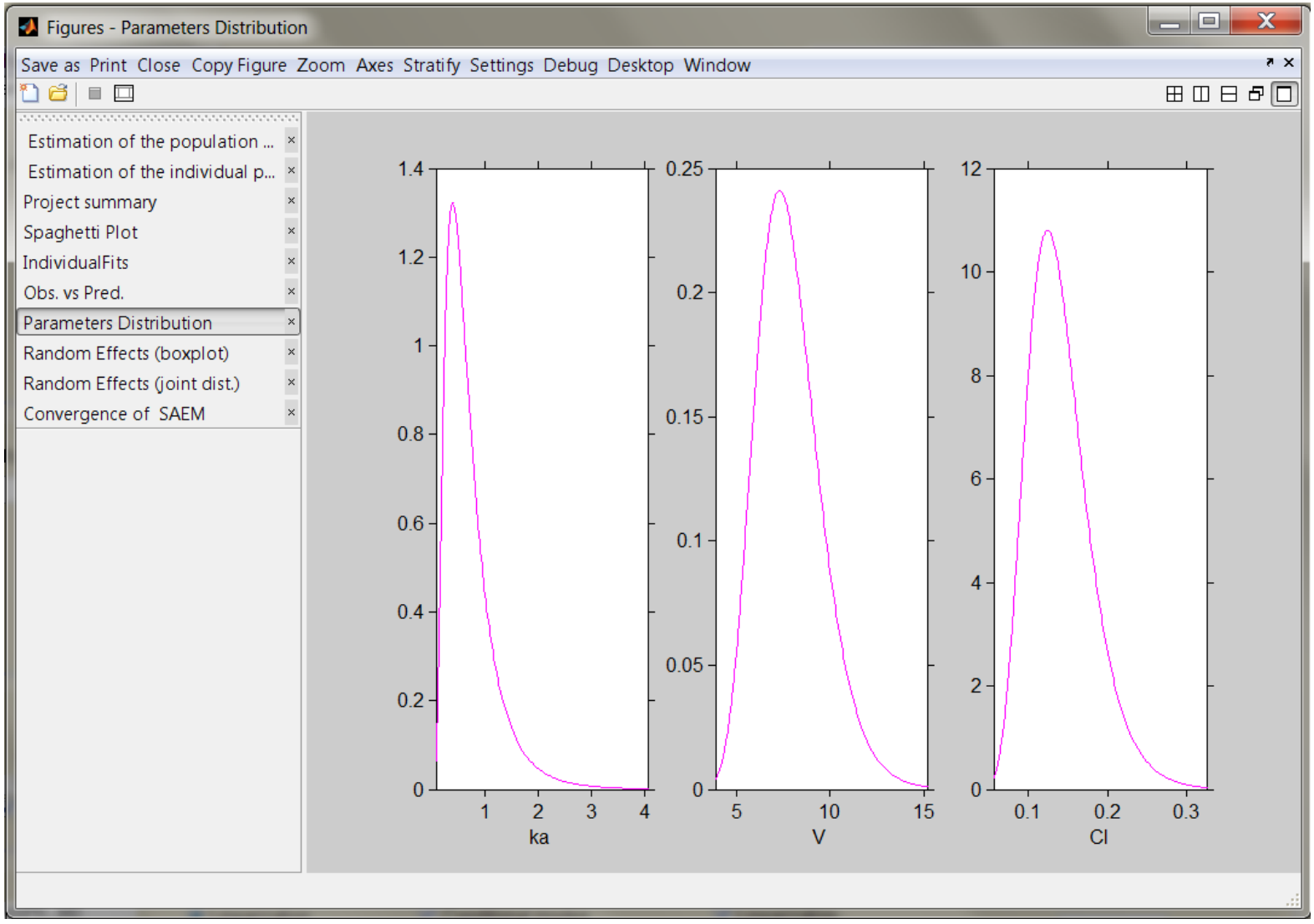


Some diagnostic plots

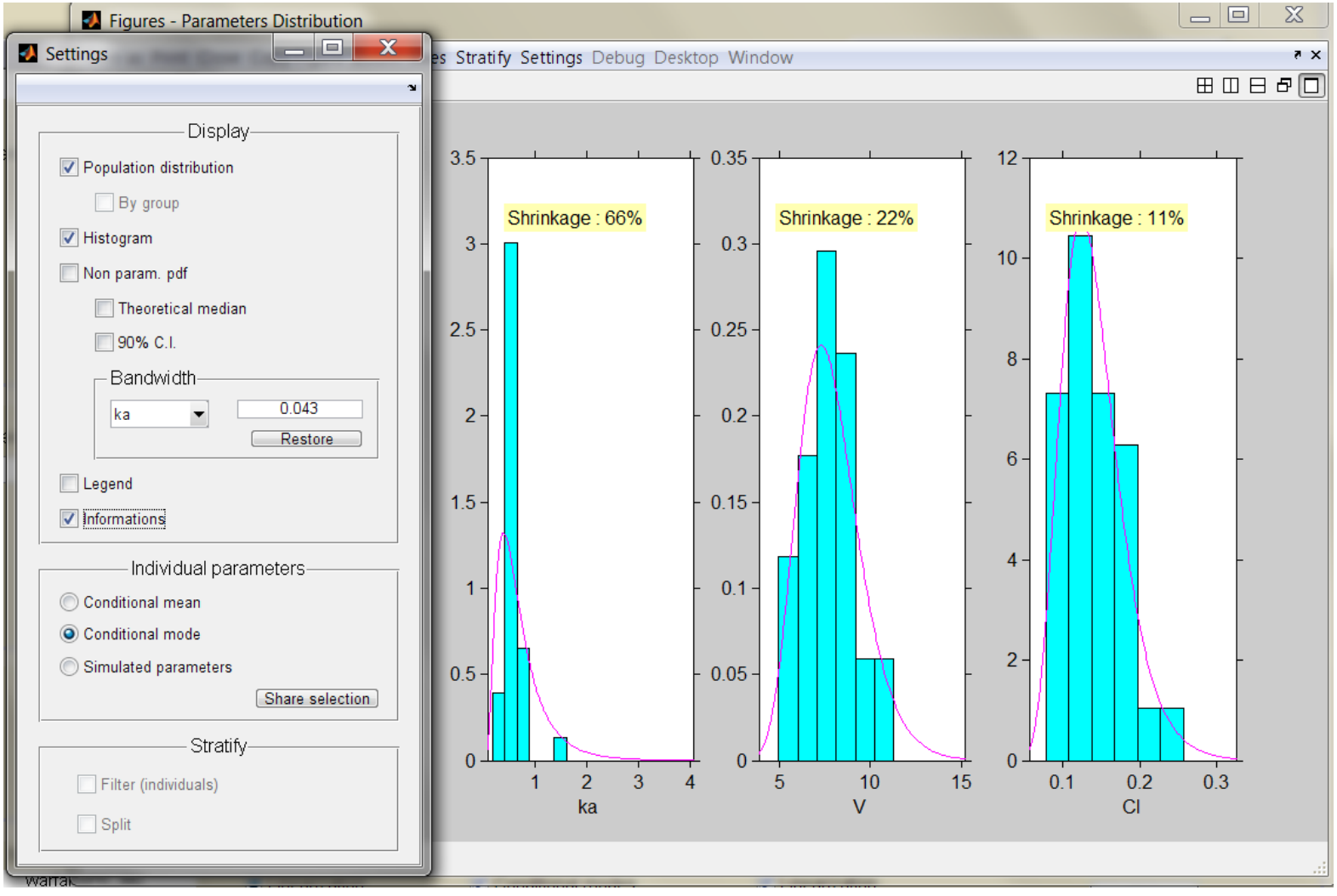
Individual fits



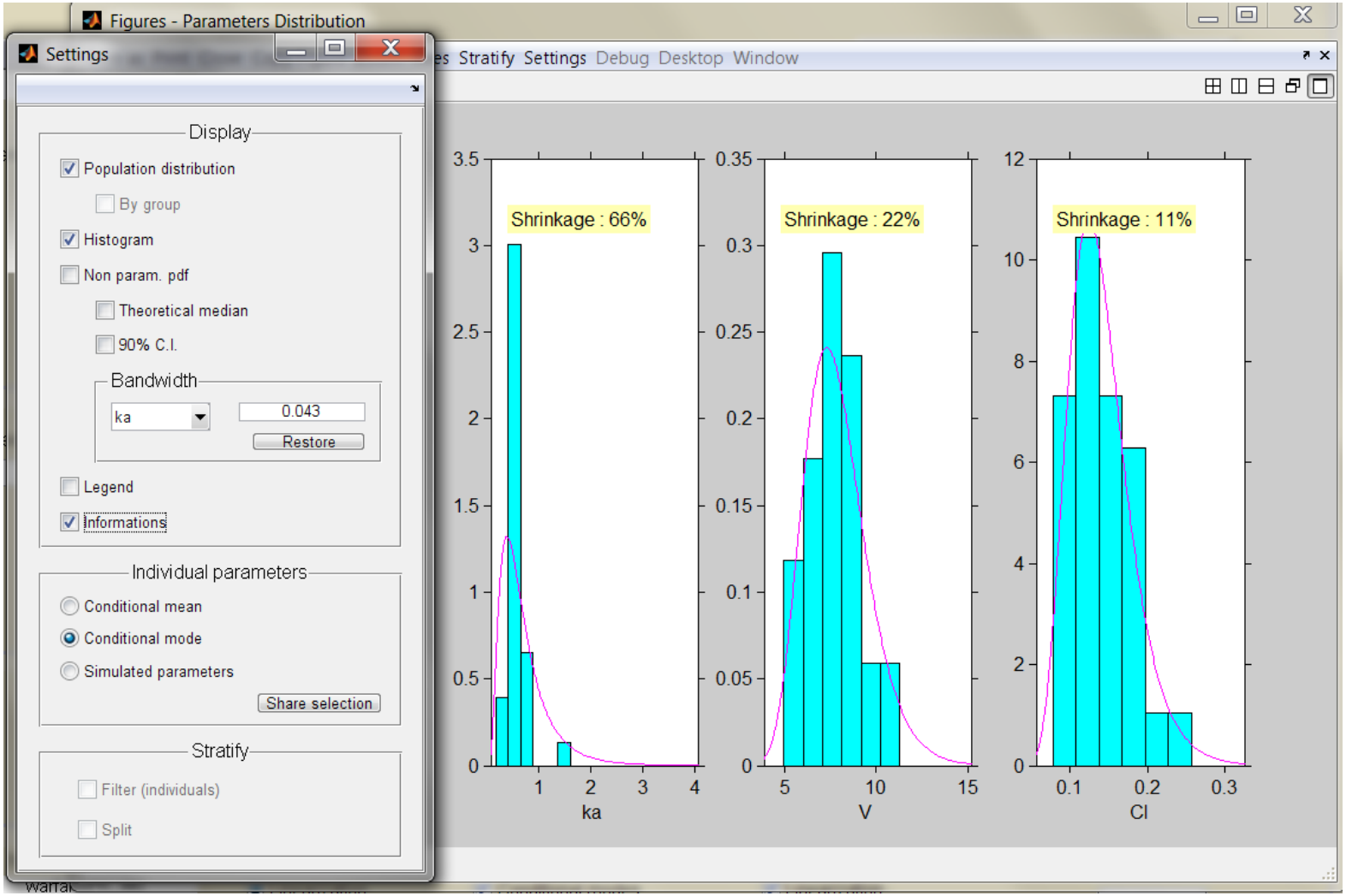
Some diagnostic plots
Distribution of the individual parameters



Some diagnostic plots
Distribution of the individual parameters

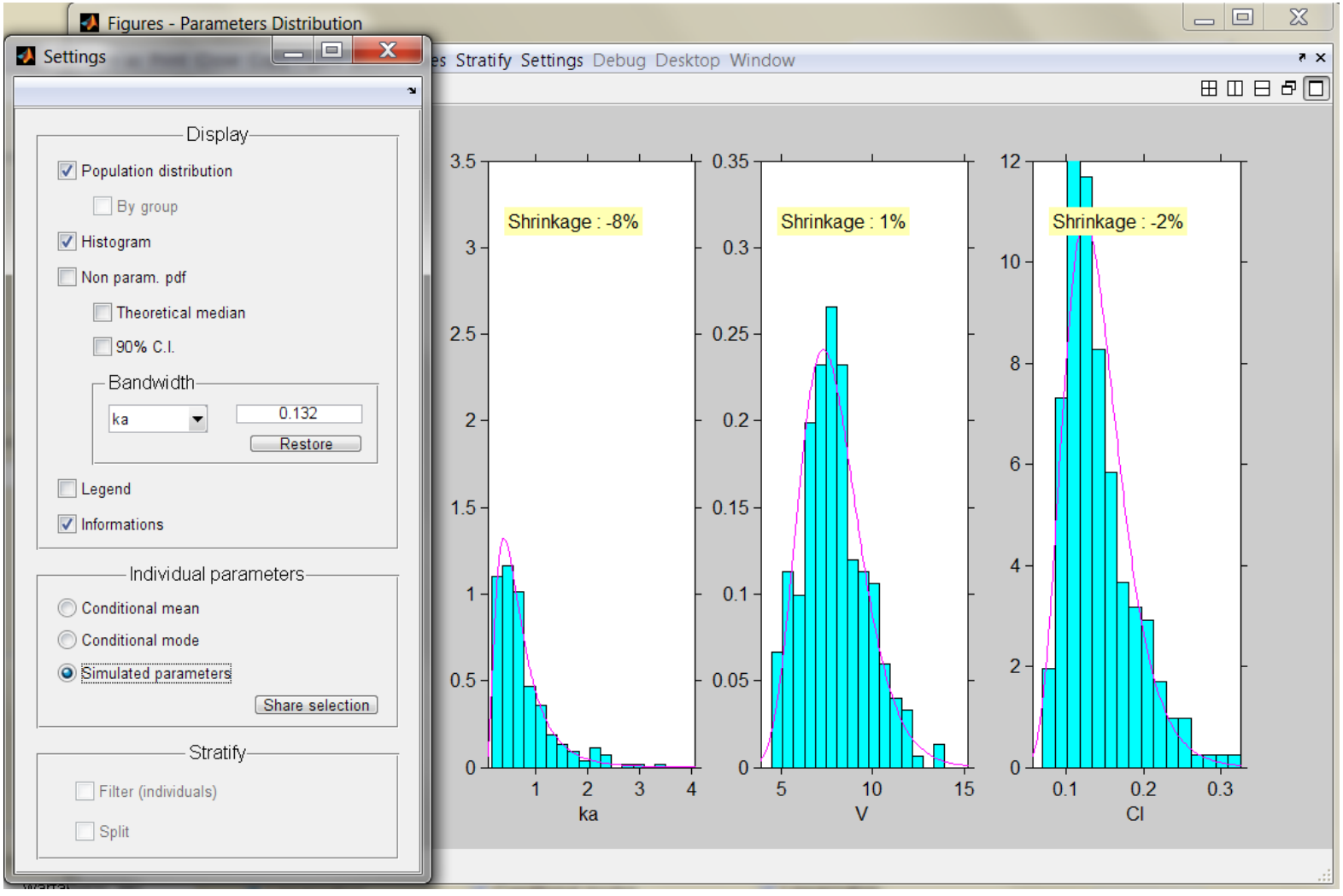


Some diagnostic plots
Distribution of the individual parameters



=> Don't use the MAP/conditional mode/posthoc/EBEs for diagnostic!

Some diagnostic plots
Distribution of the individual parameters



$=>$ Use individual parameters simulated with the conditional distribution $\mathrm{p}(\psi_i|y_i,\hat{\theta})$!

Indeed, whenever data becomes sparse or uninformative, the EBE distribution will shrink towards the population value.

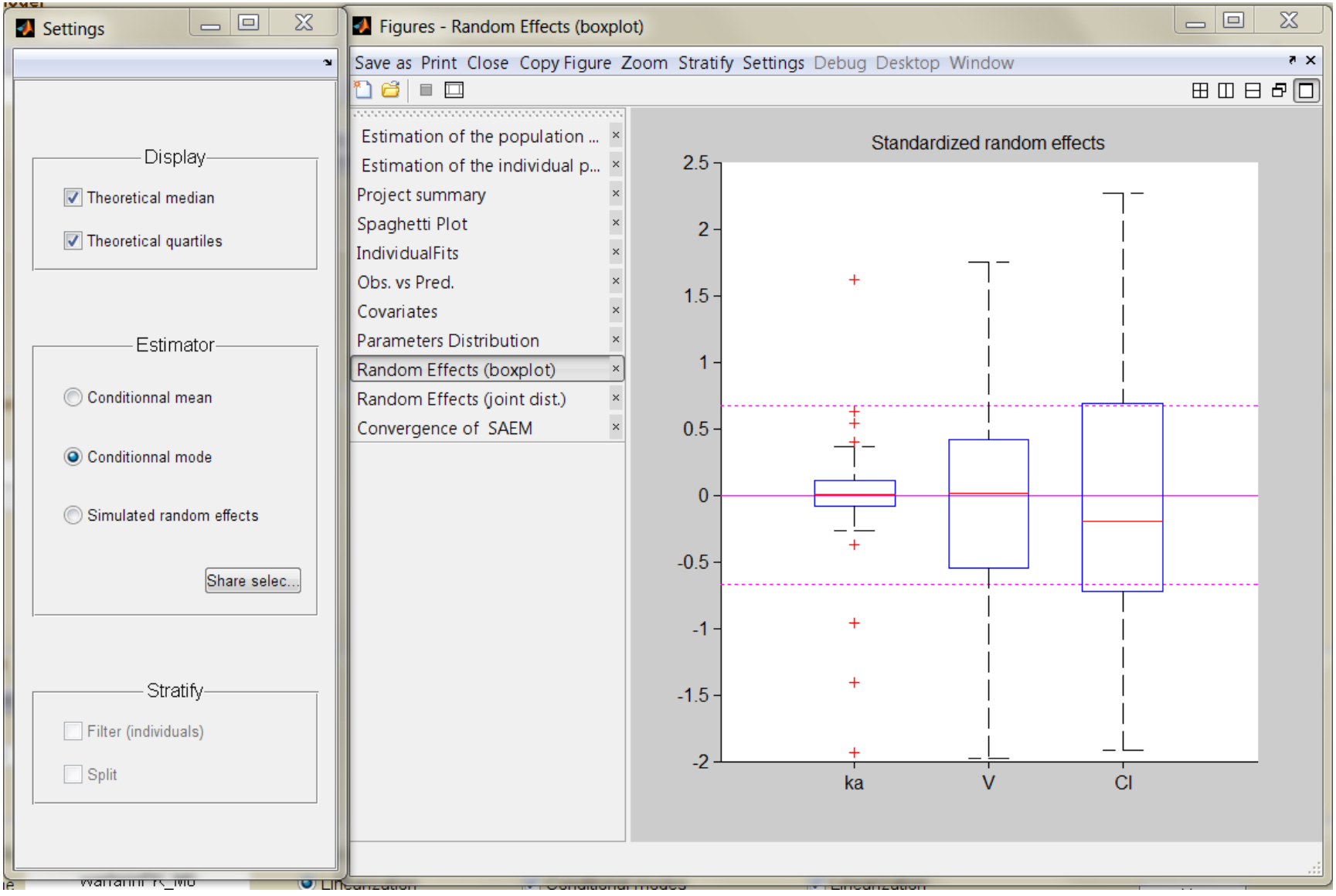
A particularly effective solution is to draw individual parameters $(\psi_i^{(k)}, 1 \leq k \leq K)$ with the conditional distribution $\mathrm{p}(\psi_i|y_i, \hat{\theta})$ rather than taking the mode.

The resulting estimator is unbiased under H_0 in the following sense:

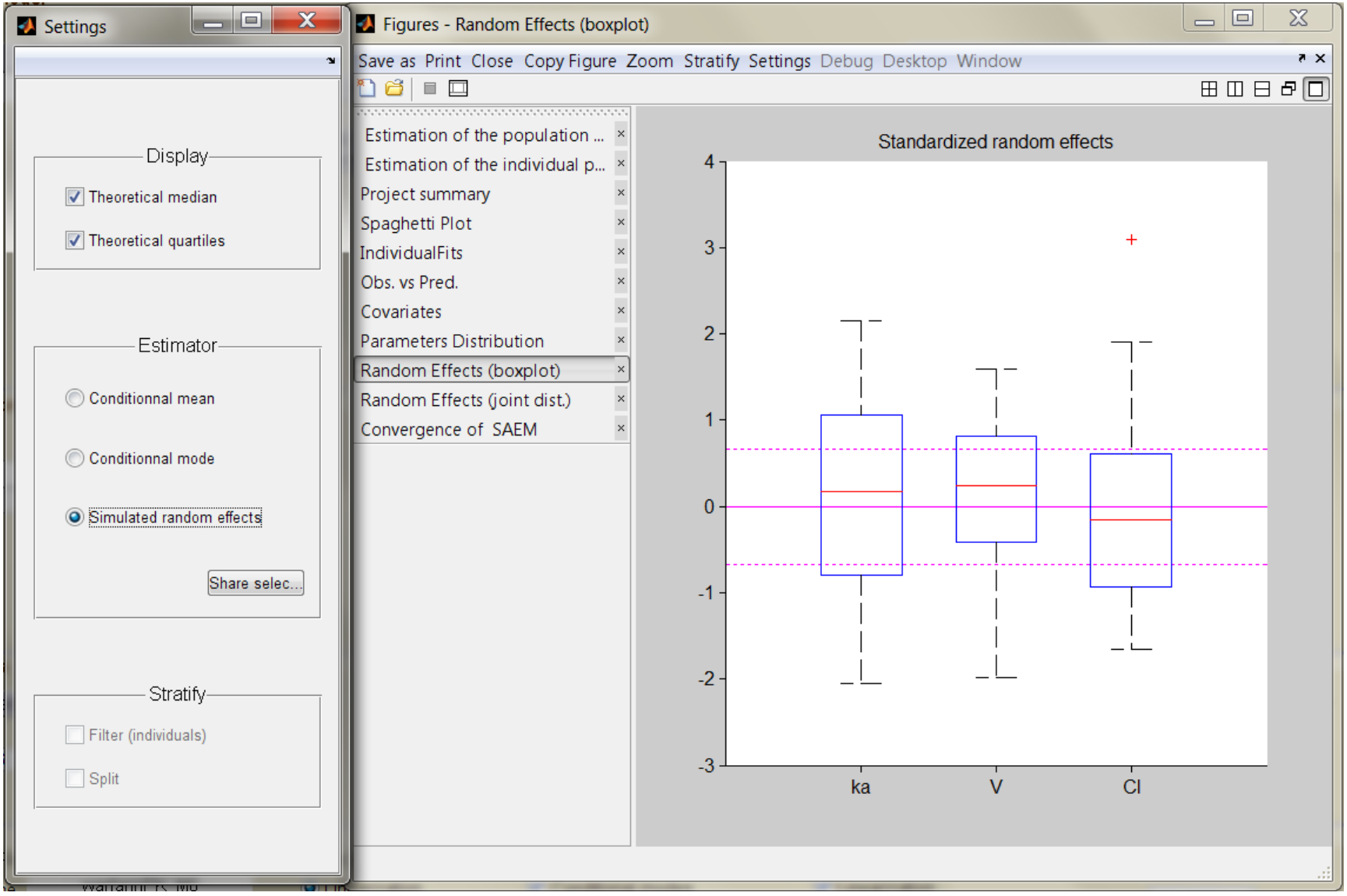
$$\begin{aligned}\mathrm{p}(\psi_i) &= \int \mathrm{p}(\psi_i|y_i)\mathrm{p}(y_i)d\, y_i \\ &= \mathbb{E}_{y_i}(\mathrm{p}(\psi_i|y_i)).\end{aligned}$$

This relationship is a fundamental one when considering inverse problems, incomplete data models, mixed effects models, etc.

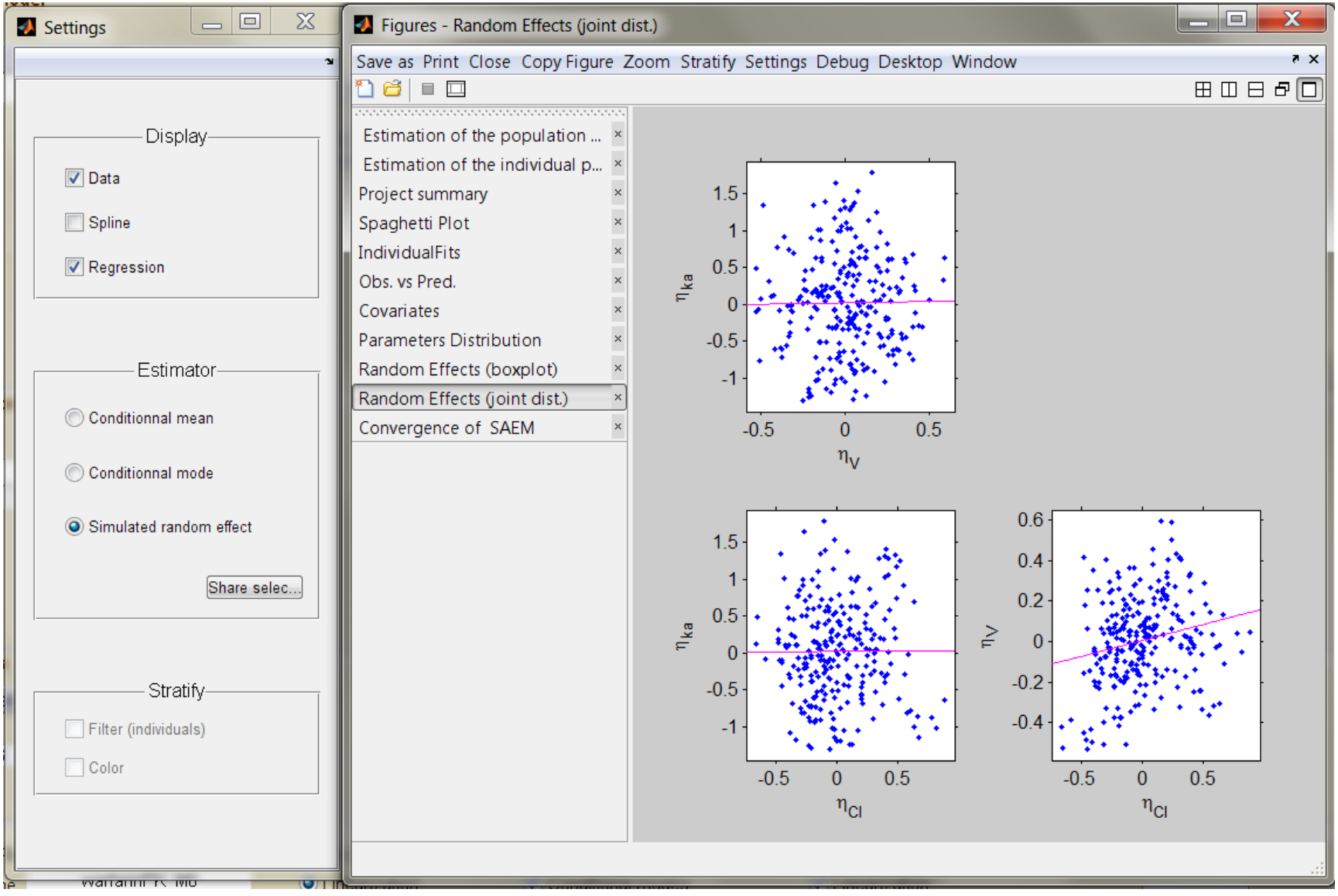
Diagnostic plots
Marginal distributions of the random effects



Diagnostic plots
Marginal distributions of the random effects

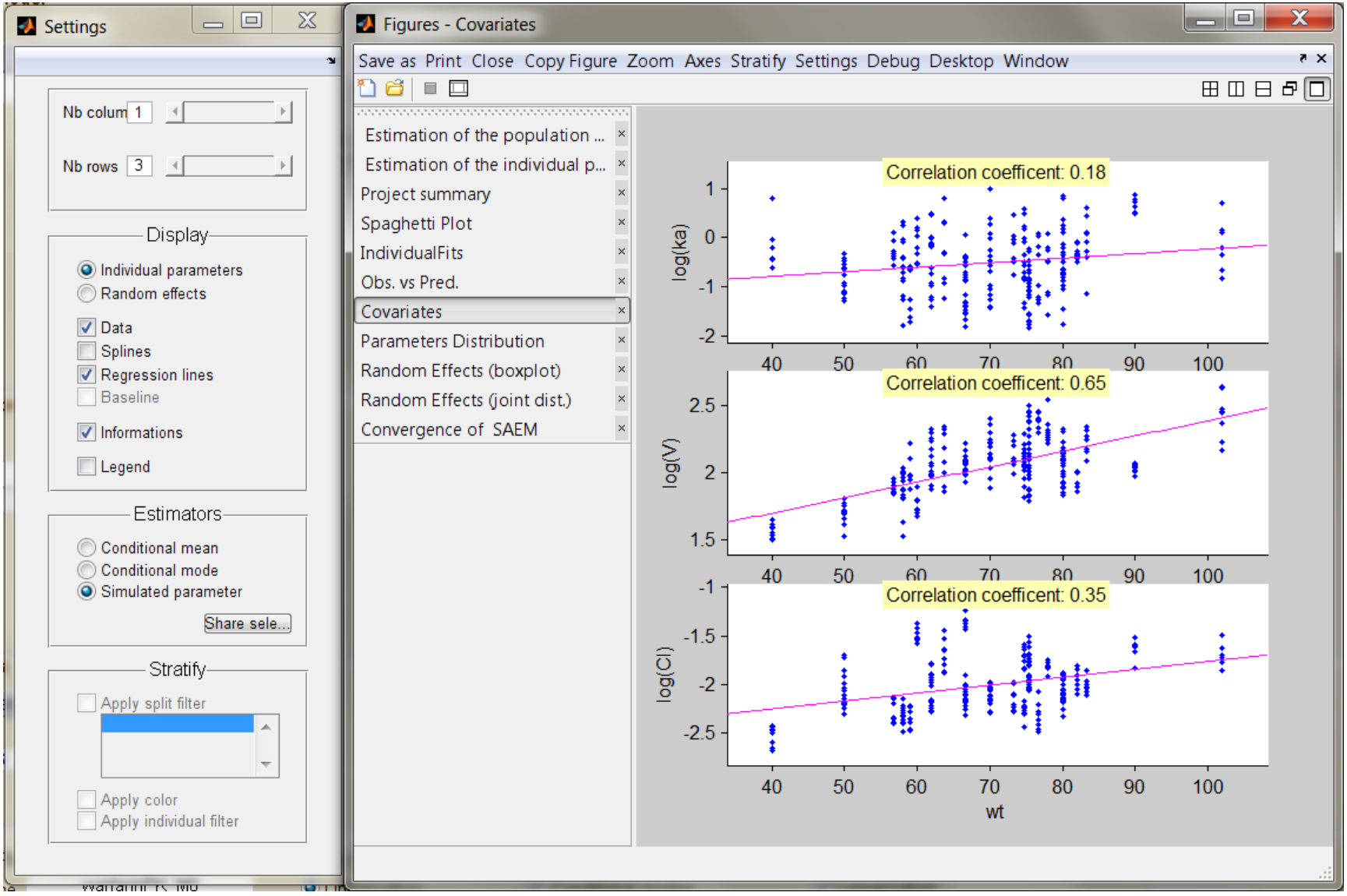


Diagnostic plots
Joint distributions of the random effects



Diagnostic plots

Relationship between parameters and covariates



IV
Mlxtran: a powerful language for complex hierarchical models

A hierarchical model

- The PK model:

$$C(t,\psi) = \frac{D\, ka}{V\, ka - Cl}\left(e^{-(Cl/V)\, t} - e^{-ka\, t}\right).$$

- The model for the observed concentrations:

$$y_{ij} = C(t_{ij},\psi_i) + a\varepsilon_{ij}$$

  i.e.

$$y_{ij}|\psi_i, a \sim \mathcal{N}(C(t_{ij},\phi_i), a^2)$$

- The model for the individual PK parameters $\psi_i = (ka_i, V_i, Cl_i)$:

$$\begin{aligned}
\log(ka_i) &\sim \mathcal{N}(\log(ka_{\text{pop}}), \omega^2_{ka}) \\
\log(V_i) &\sim \mathcal{N}(\log(V_{\text{pop}}), \omega^2_{V}) \\
\log(Cl_i) &\sim \mathcal{N}(\log(Cl_{\text{pop}}), \omega^2_{Cl}).
\end{aligned}$$

- The distribution for ka_{pop}: $\log(ka_{\text{pop}}) \sim \mathcal{N}(\log(ka^\star), \gamma^2)$

Mlxtran code for hierarchical model

```
[LONGITUDINAL]
input = {ka, V, Cl, a}
EQUATION:
C = pkmodel(ka, V, Cl)
DEFINITION:
y = {distribution = normal, prediction = C, sd = a}
;-------------------------------------------

[INDIVIDUAL]
input = {ka_pop, omega_ka, V_pop, omega_V, Cl_pop, omega_Cl}
DEFINITION:
ka = {distribution = lognormal, mean = log(ka_pop), sd = omega_ka}
V  = {distribution = lognormal, mean = log(V_pop) , sd = omega_V }
Cl = {distribution = lognormal, mean = log(Cl_pop), sd = omega_Cl}
;-------------------------------------------

[POPULATION]
input = {ka_star, gamma_ka}
DEFINITION:
ka_pop = {distribution = lognormal, mean = log(ka_star), sd = gamma_ka}
```

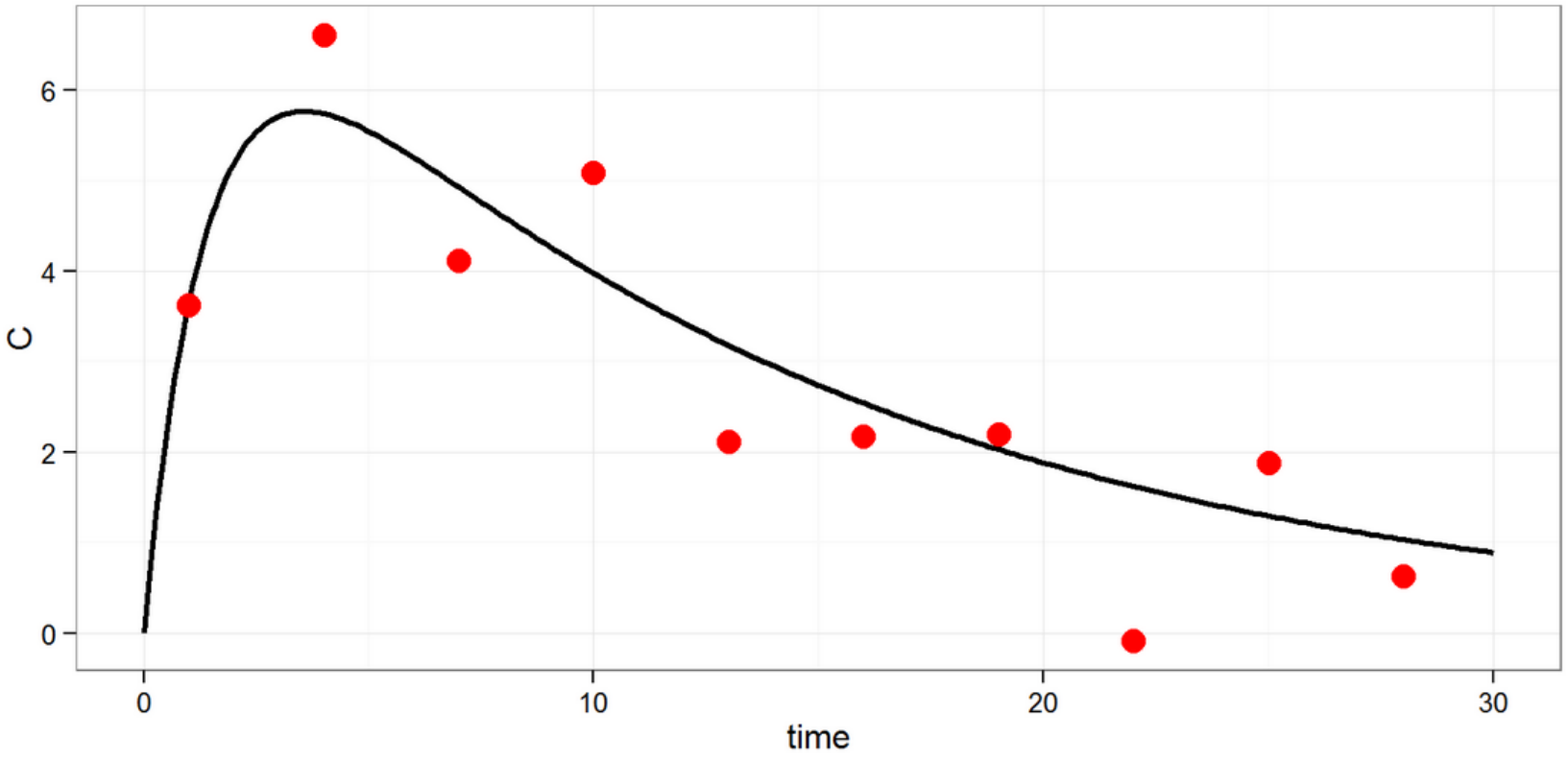
R code using Simulx

```
p <- c(ka_star=0.5, gamma_ka=0.1, omega_ka=0.2,
       V_pop=10,    omega_V=0.2,
       Cl_pop=1,    omega_Cl=0.15,
       a=0.5)

f   <- list(name='C', time=seq(0, 30, by=0.1))
obs <- list(name='y', time=seq(1, 30, by=3))
ind <- list(name=c('ka','V', 'Cl'))
pop <- list(name='ka_pop')

res <- simulx(model     = 'mlxtran1.txt',
              parameter = p,
              treatment = list(time=0, amount=100),
              output    = list(pop, ind, f, obs))

print(res$parameter)
```

```
##      ka_pop        ka         V       Cl
## 1 0.6039348 0.7340071 10.46294 1.327486
```

R code using Simulx

```
print(ggplotmlx() + geom_line(data=res$C, aes(x=time, y=C), size=1) +
        geom_point(data=res$y, aes(x=time, y=y), colour='red', size=4))
```



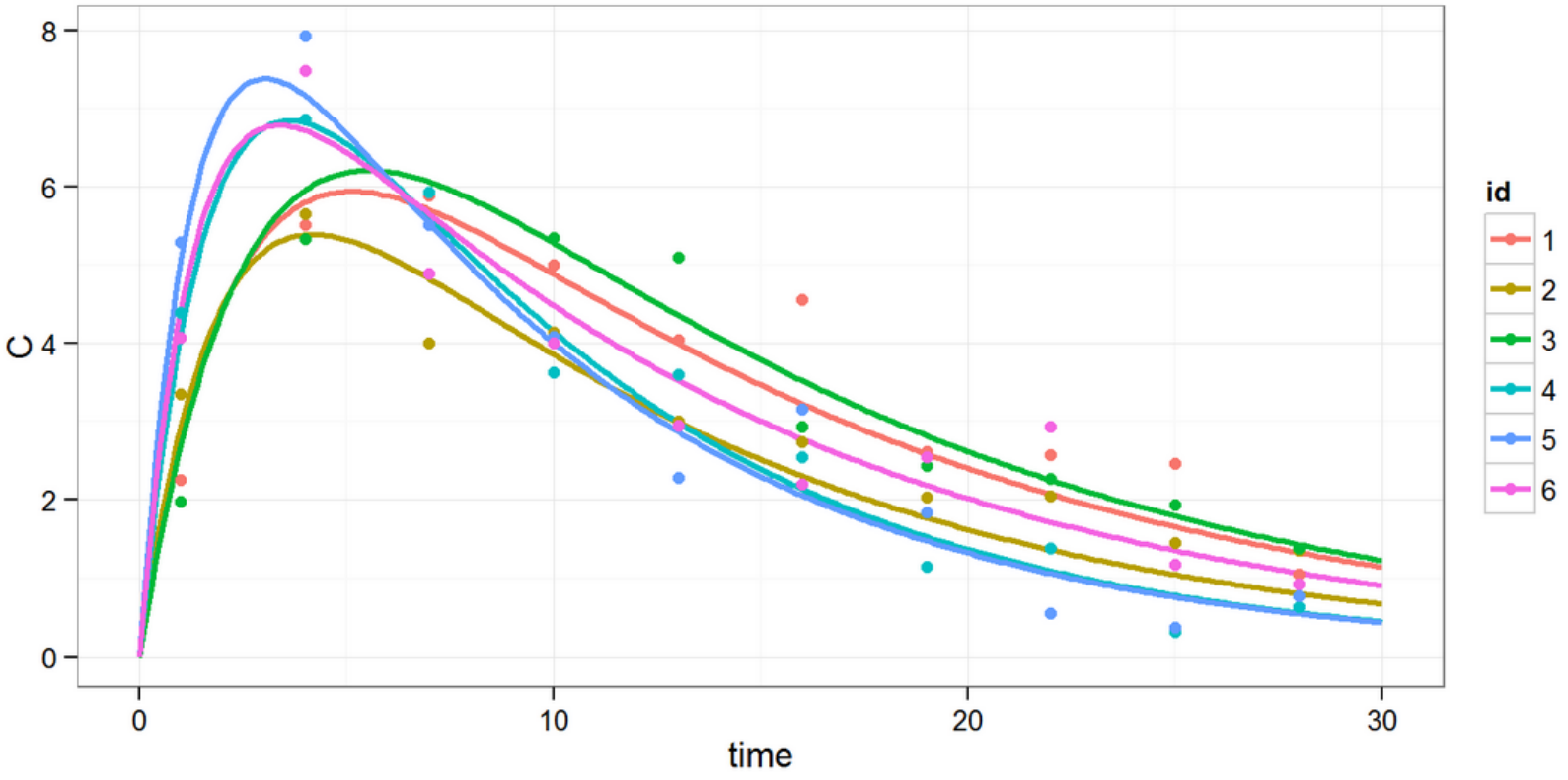
R code using Simulx

```
res <- simulx(model     = 'mlxtran1.txt',
              parameter = p,
              treatment = list(time=0, amount=100),
              output    = list(pop, ind, f, obs),
              group     = list(size=c(2,3), level=c('population','individual')))

print(res$parameter)
```

```
##   id    ka_pop        ka         V        Cl
## 1  1 0.4505751 0.4021775 11.465519 0.8554521
## 2  2 0.4505751 0.5028199 12.809041 1.1290427
## 3  3 0.4505751 0.3519262 10.550282 0.8050163
## 4  4 0.5421385 0.5509877  9.718898 1.0911209
## 5  5 0.5421385 0.7338368  9.671072 1.0744846
## 6  6 0.5421385 0.7349852 11.236472 0.8981349
```

R code using Simulx

```
print(ggplotmlx() +
        geom_line( data=res$C, aes(x=time, y=C, colour=id), size=1) +
        geom_point(data=res$y, aes(x=time, y=y, colour=id), size=2))
```



simulx.webpopix.org