HelmholtzZentrum münchen
German Research Center for Environmental Health

Technische Universität München

Bayes Forum — August 4, 2017

Uncertainty analysis using profile likelihoods and profile posteriors

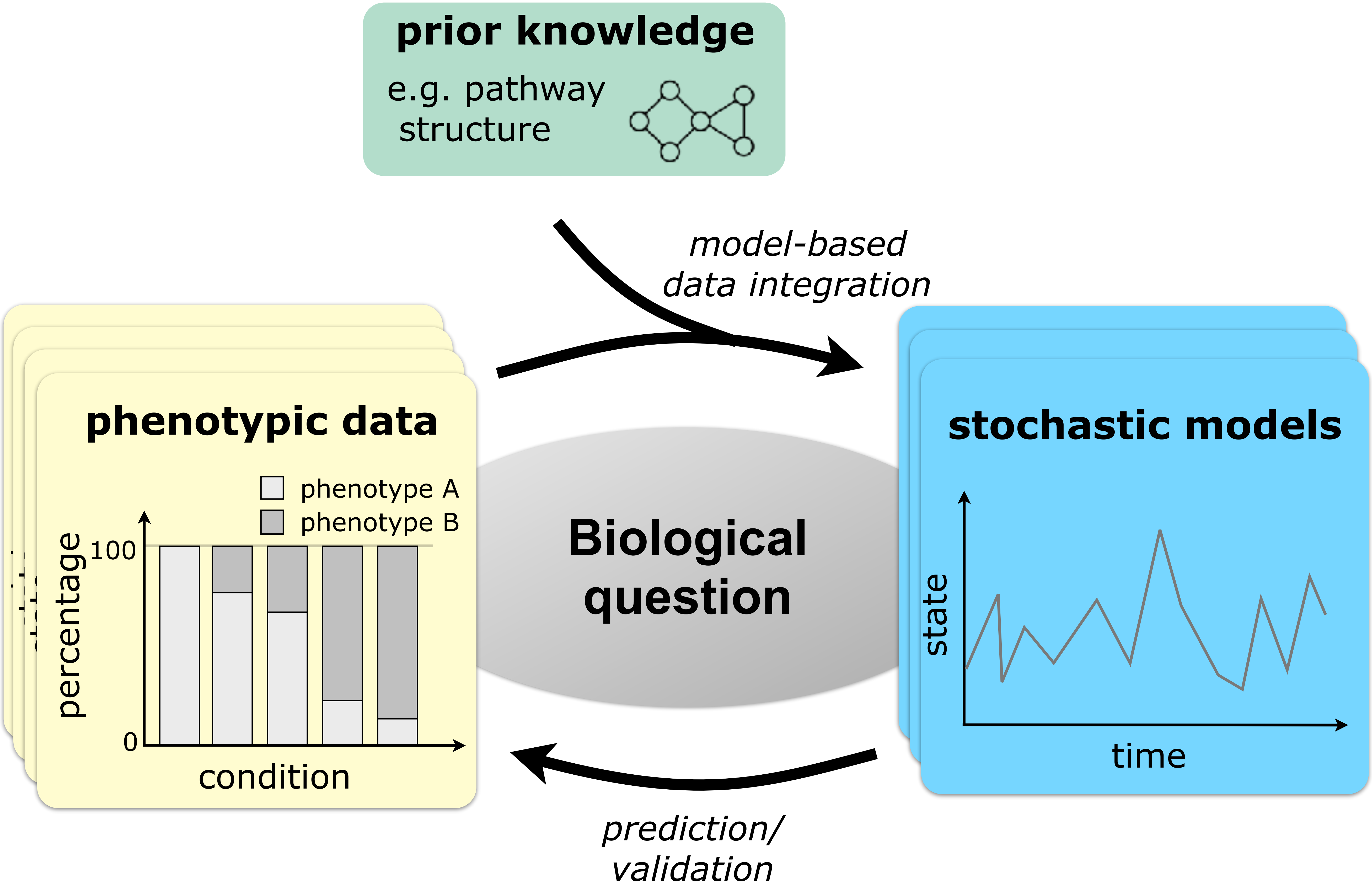
Jan Hasenauer

Research Group for Data-Driven Computational Modelling
Institute of Computational Biology
Helmholtz Center Munich

Department of Mathematics
Technical University of Munich

HELMHOLTZ
ASSOCIATION

Model-based approach



Modeling of biochemical reaction networks

Chemical reactions

$$\underbrace{\sum_{i=1}^{n_t} \nu_{i,j} X_i}_{\text{educts/reactants}} \quad \underset{k_{-j}}{\overset{k_{+j}}{\rightleftharpoons}} \quad \underbrace{\sum_{i=1}^{n_t} \eta_{i,j} X_i}_{\text{products}}$$

with

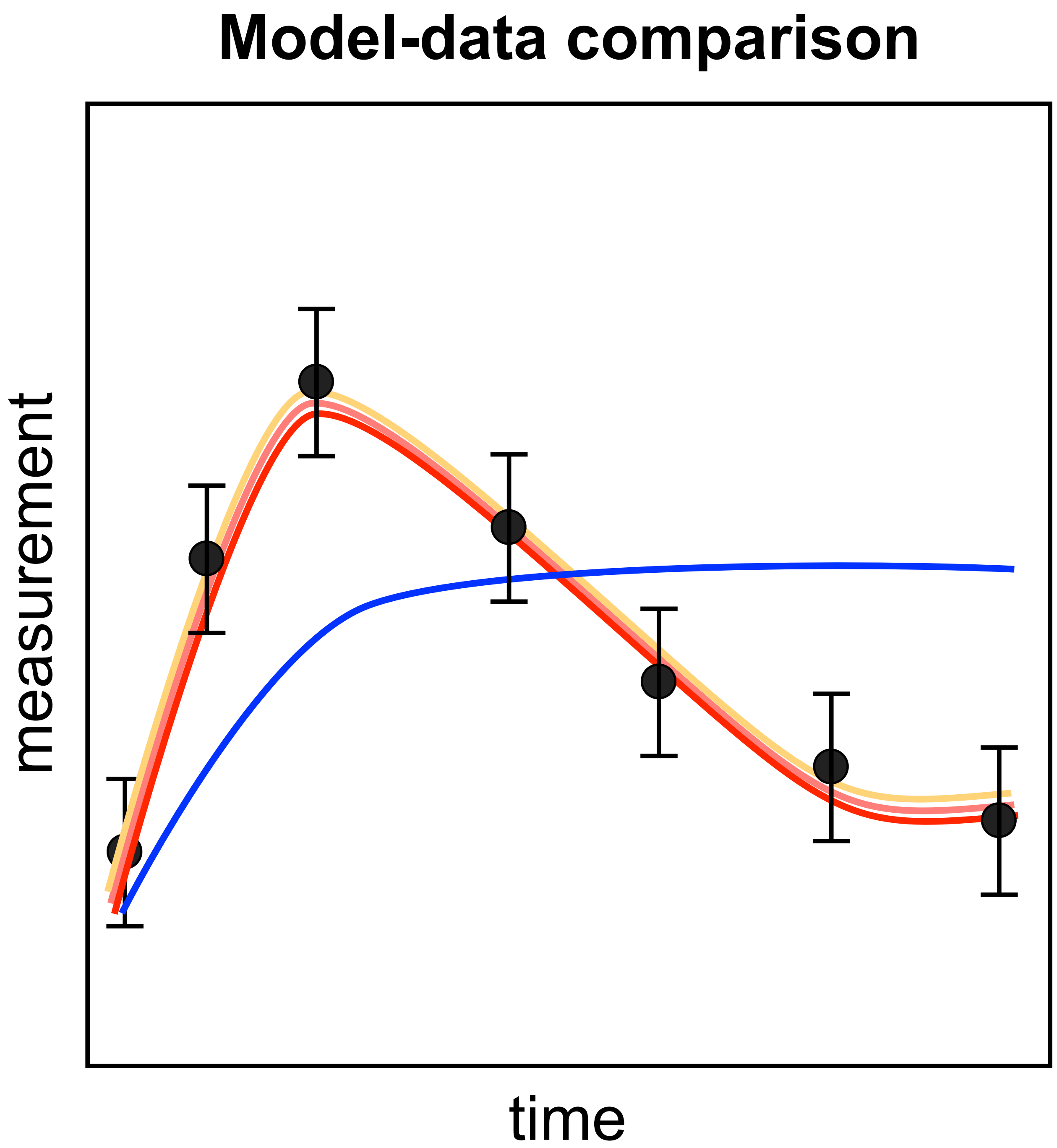
- biochemical species X_i
- stoichiometric coefficients $\nu_{i,j}, \eta_{i,j} \in \mathbb{N}_0$ and
- reaction rate constants $k_{+j}, k_{-j} \in \mathbb{R}_+$.

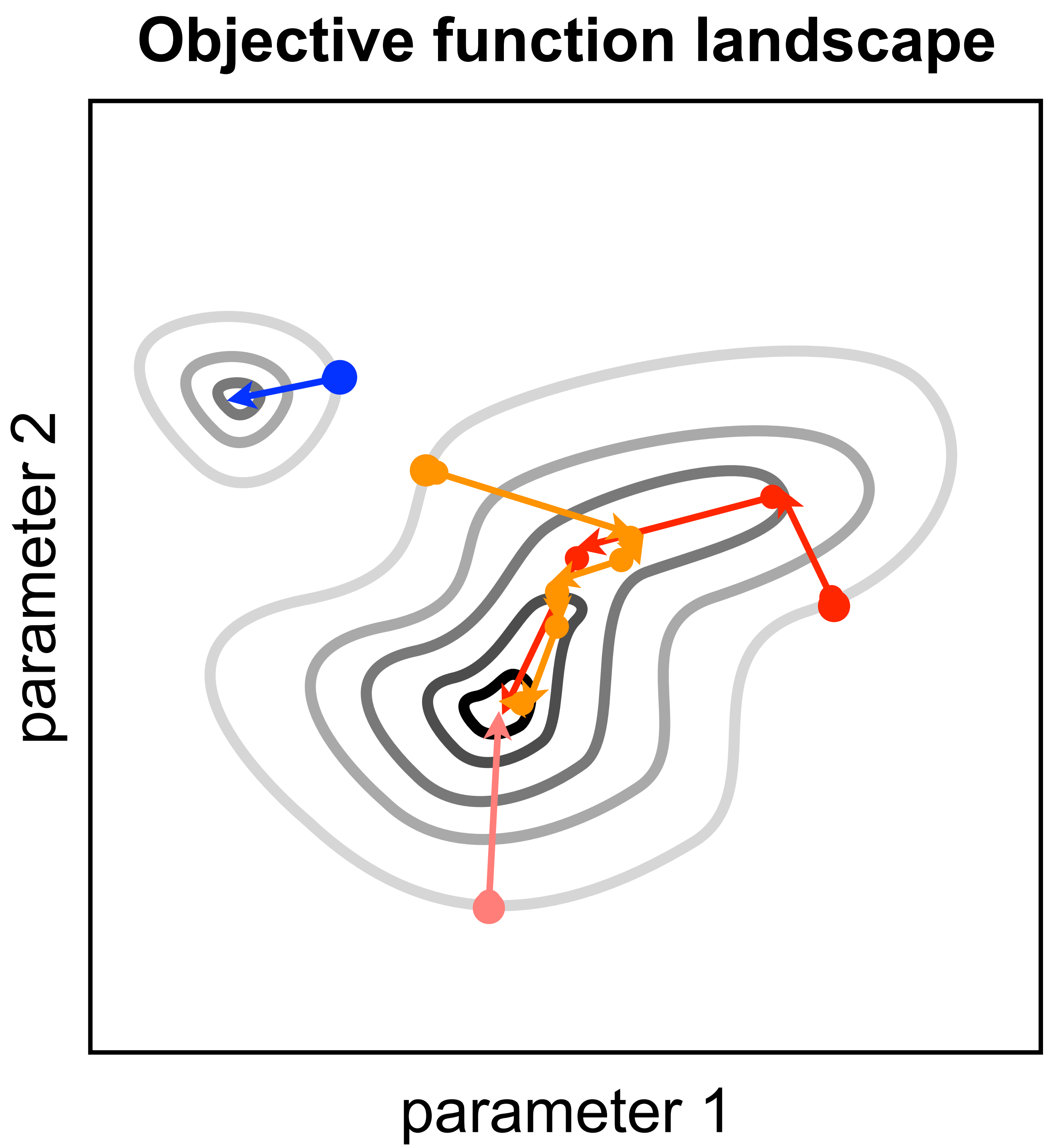
Reaction Rate Equation (RRE)

The temporal evaluation of the concentration $x_i = [X_i]$ is captured by

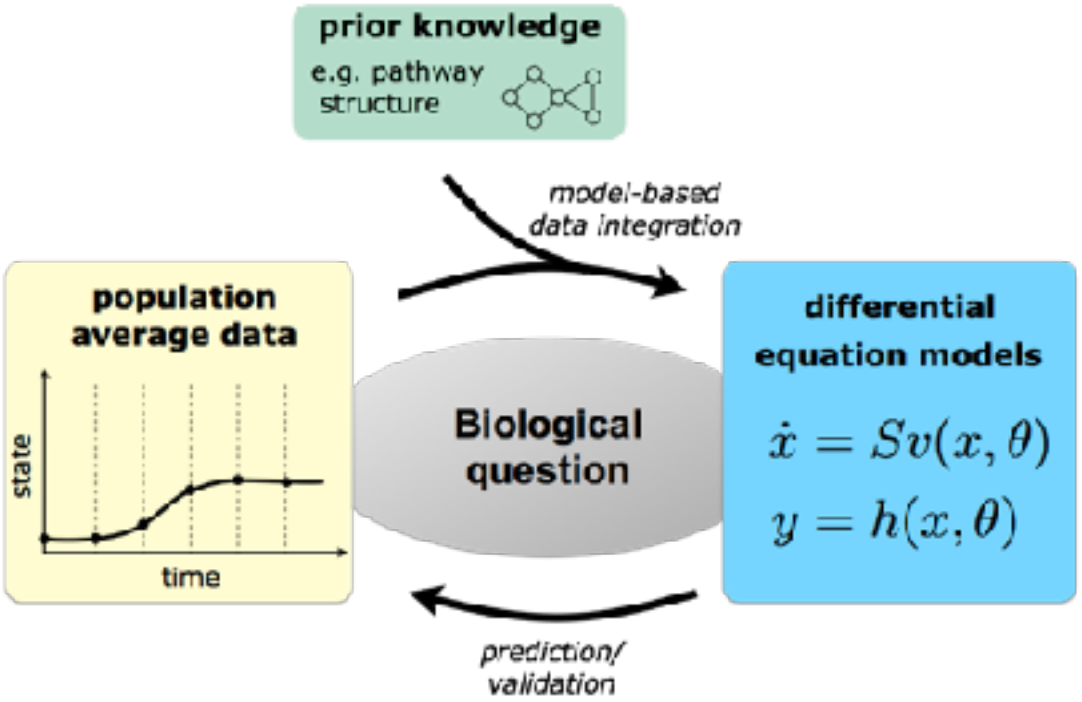
$$\begin{aligned} \frac{dx(t,\theta)}{dt} &= f(x(t,\theta),\theta,t), \qquad x(0) = x_0(\theta) \\ &= S \cdot v(x(t,\theta),\theta,t), \end{aligned}$$

Parameter optimisation





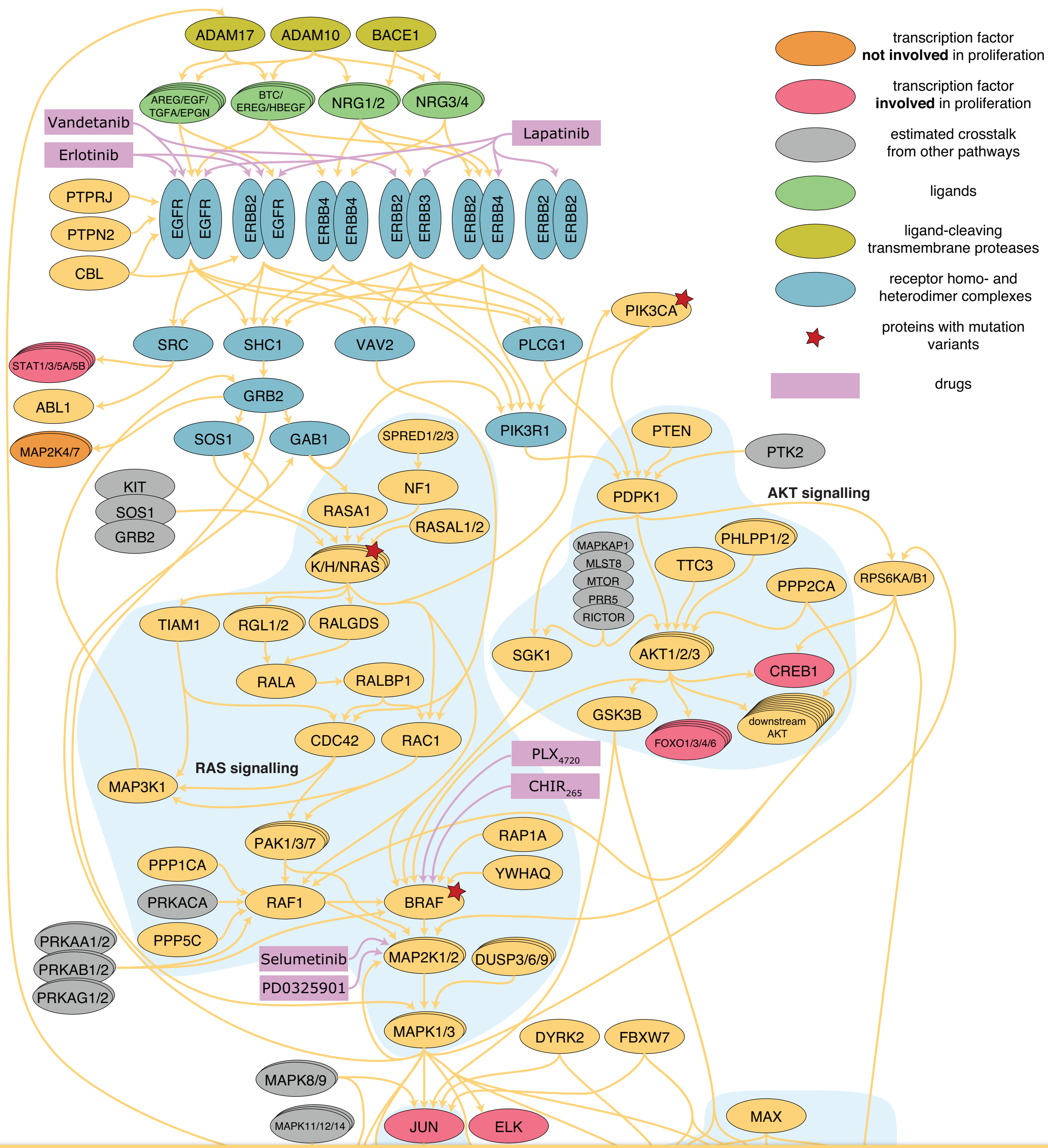
Scalability of conventional methods?



Scalable inference for differential equations

F. Fröhlich, B. Kaltenbacher, F. J. Theis and J. Hasenauer. Scalable parameter estimation for genome-scale biochemical reaction networks. *PLoS Computational Biology*, 13(1):e1005331, 2017.

Large-scale model for personalised medicine



Model properties

Genes: 112

Mutant genes: 24

Reactions: 2704

⇒ State variables: 1230

⇒ Parameters: 4256

Dataset

Cell lines: 120

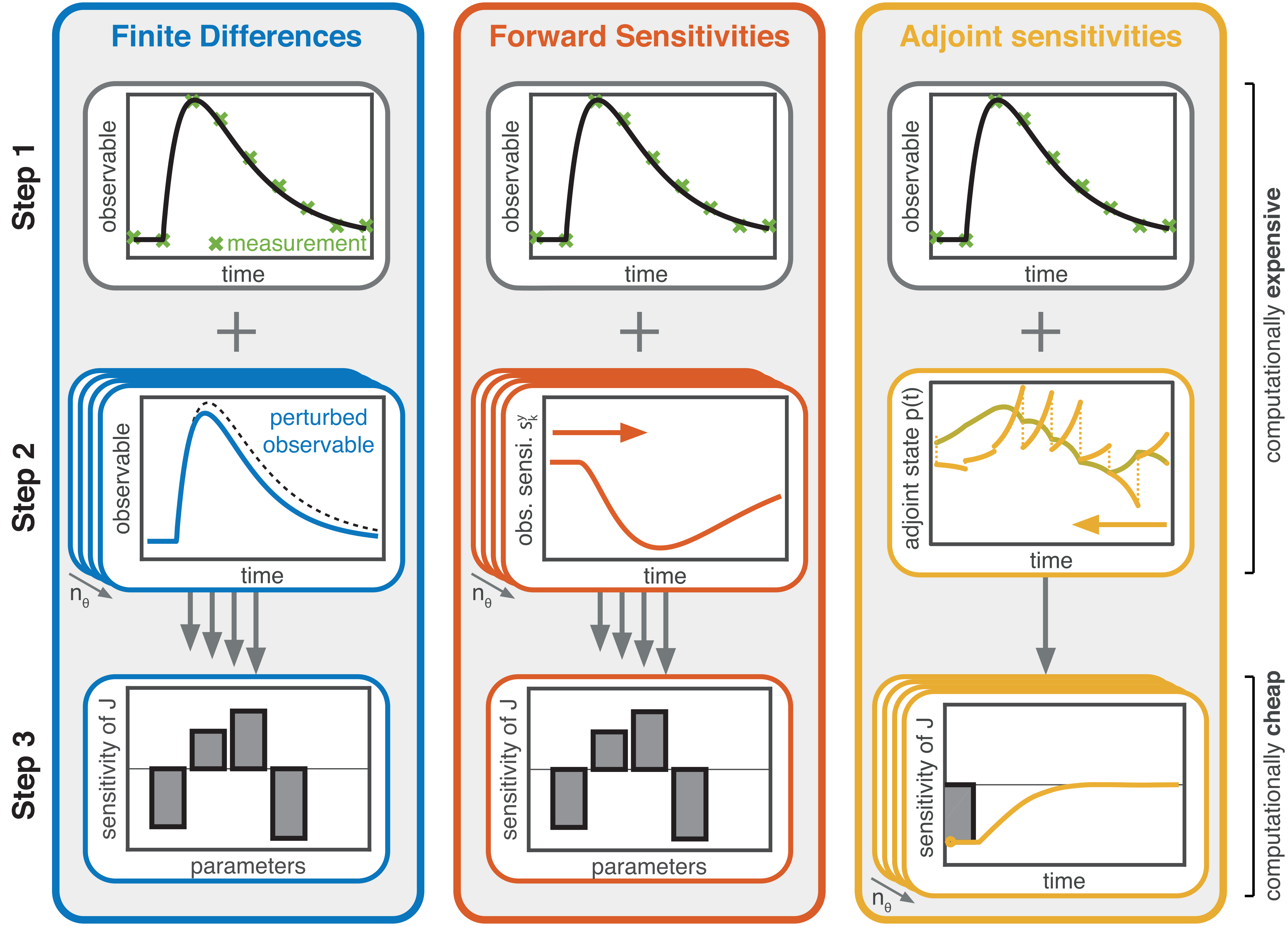
Drugs: 7

Drug concentrations: 7

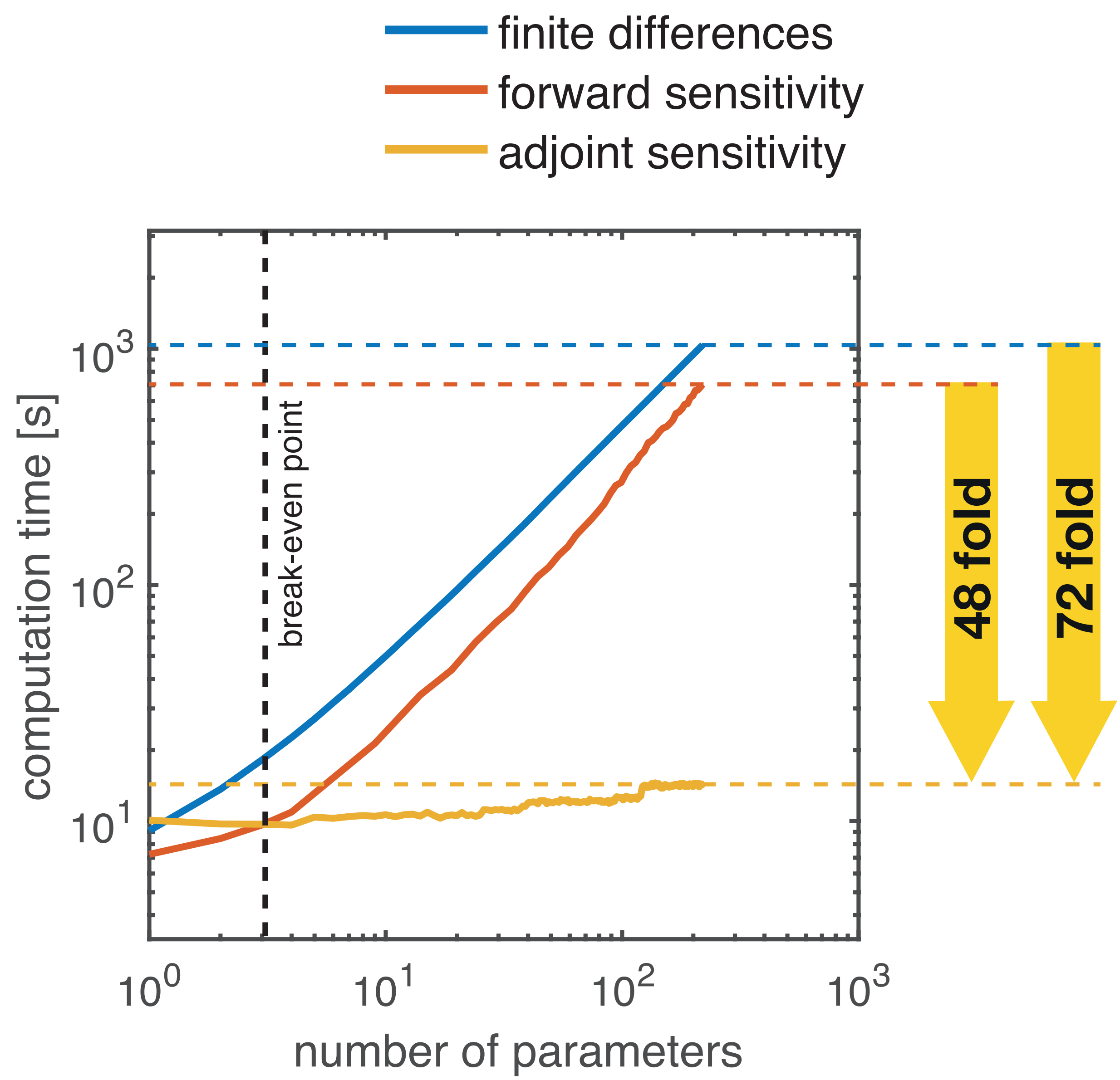
⇒ ~6000 conditions

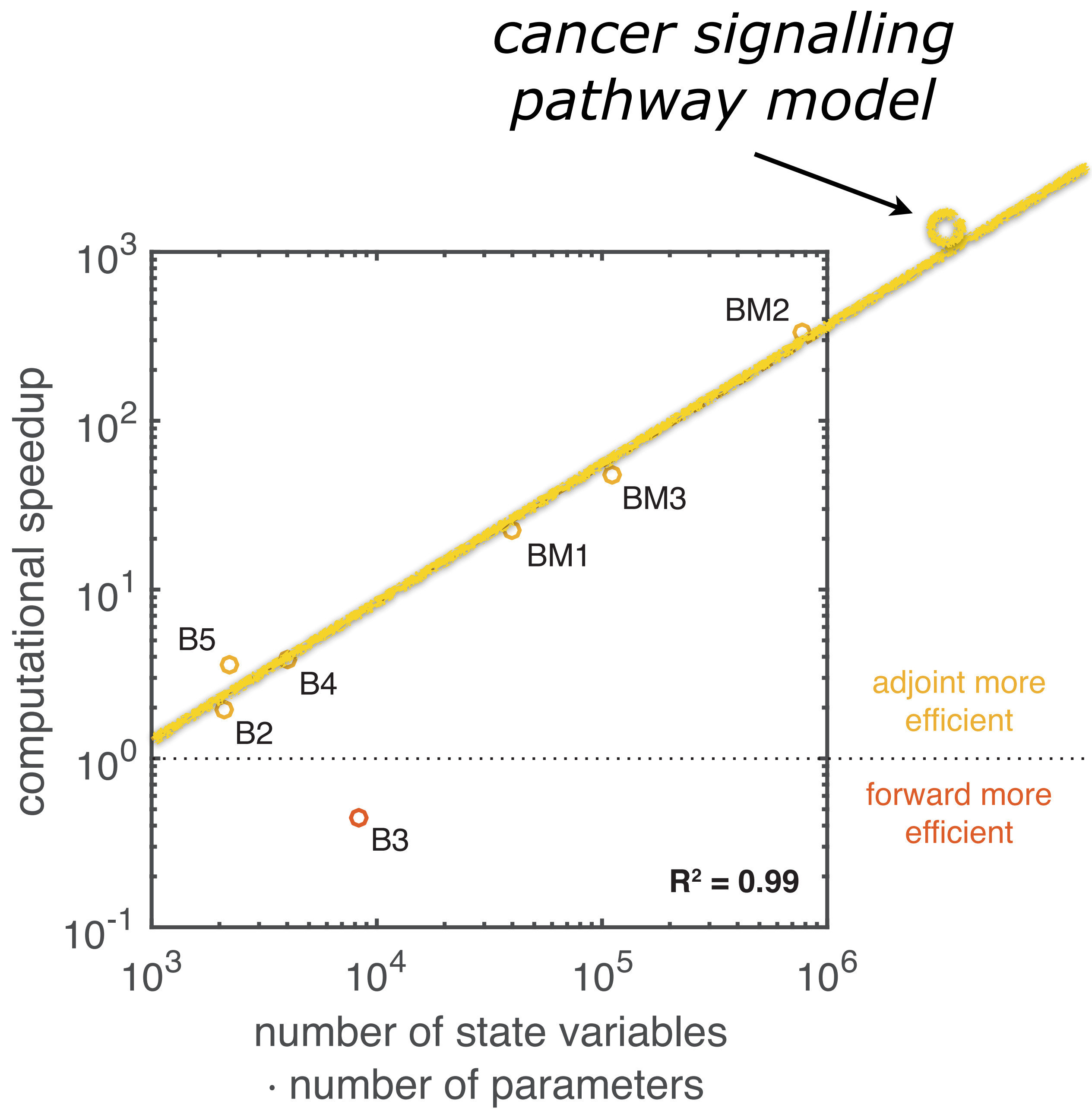
Parameter estimation for models and datasets of this size?

Computation of objective function gradient



Sensitivity analysis





Adjoint methods facilitate scalable gradient evaluation.

Adjoint method for gradient evaluation

1) Calculation of **state** via simulation

$$\dot{x}(t) = f(x(t), \theta), \quad x(0) = x_0(\theta)$$
$$y(t) = h(x(t), \theta)$$

2) Calculation of **adjoint state** as solution to backward differential equation

$$p(t) = 0, \quad t \in (t_N, t_{N+1})$$

for $k = N : -1 : 1$

$$\dot{p}(t) = - \left.\frac{\partial f}{\partial x}\right|_{x(t),\theta}^{T} p(t), \quad t \in (t_{k-1}, t_k)$$

$$\text{with } p(t_k) = \lim_{t \to t_k^+} p(t) + \sum_{j=1}^{m} \frac{1}{\sigma_{j,k}^2} \frac{\partial h_j}{\partial x}(x(t_k), \theta)^T (\bar{y}_{j,k} - h_j(x(t_k), \theta))$$

3) Calculation of **gradient** using one-dimensional integral

$$\frac{\partial J}{\partial \theta_i} = - \int_0^T p(t)^T \left.\frac{\partial f}{\partial \theta_i}\right|_{x(t),\theta} dt - \sum_{k=1}^{N} \sum_{j=1}^{m} 1 \left.\frac{\partial h_j}{\partial \theta_i}\right|_{x(t_k),\theta}^{T} \left(\frac{\bar{y}_{j,k} - h_j(x(t_k), \theta)}{\sigma_{j,k}^2} \right) - p(0)^T \frac{\partial x_0}{\partial \theta_i}$$

Optimisation of large-scale signalling pathway model using CCLE data

Standard optimisation methods

#cross validations (5)
x #local optimisations(10)
x #iterations(100)
x #parameters(~4000)
x #conditions(~5500)
= **~10^{11} ODE solves**

x ~2min = **~200k years**

large dataset

Acceleration method

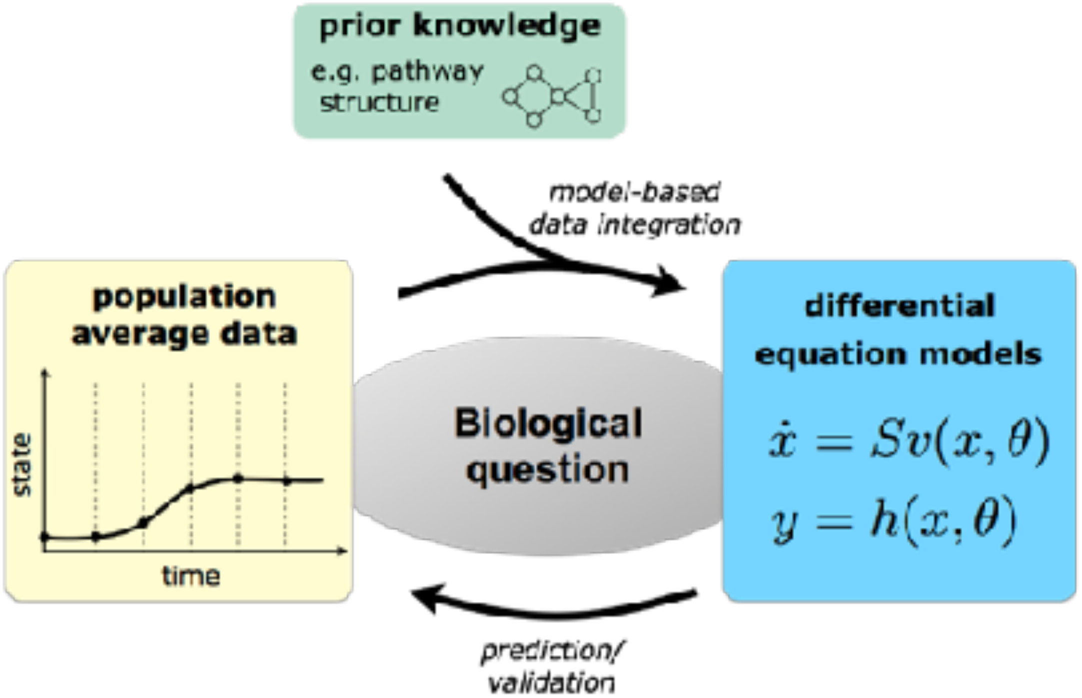
exploit sparsity (~60x)

Tailored optimisation methods

#cross validations (**1**)
x #local optimisations(**1**)
x #iterations(100)
x #parameters(**~3**)
x #conditions(**~800**)
= **~2×10^{5} ODE solves**

x ~2s = **~1 week**

BUT: How uncertain are the parameters?



Profile likelihoods and profile posteriors for uncertainty analysis

W. Q. Meeker and L. A. Escobar. Teaching about approximate confidence regions based on maximum likelihood estimation. *Am. Stat.,* 49(1):48-53, 1995.

J.-S. Chen and R. I. Jennrich. Simple accurate approximation of likelihood profiles. *J. Comput. Graphical Statist.,* 11(3):714-732, 2002.

Frequentist and Bayesian methods

Maximum Likelihood (ML) estimator

The ML estimate $\theta^{ml} \in \Omega \subseteq \mathbb{R}_+^{n_\theta}$ maximises the likelihood,

$$\theta^{ml} = \arg\max_{\theta \in \Omega} p(\mathcal{D}|\theta), \quad \text{subject to } \mathcal{M}(\theta).$$

Bayes's theorem:

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}$$

with

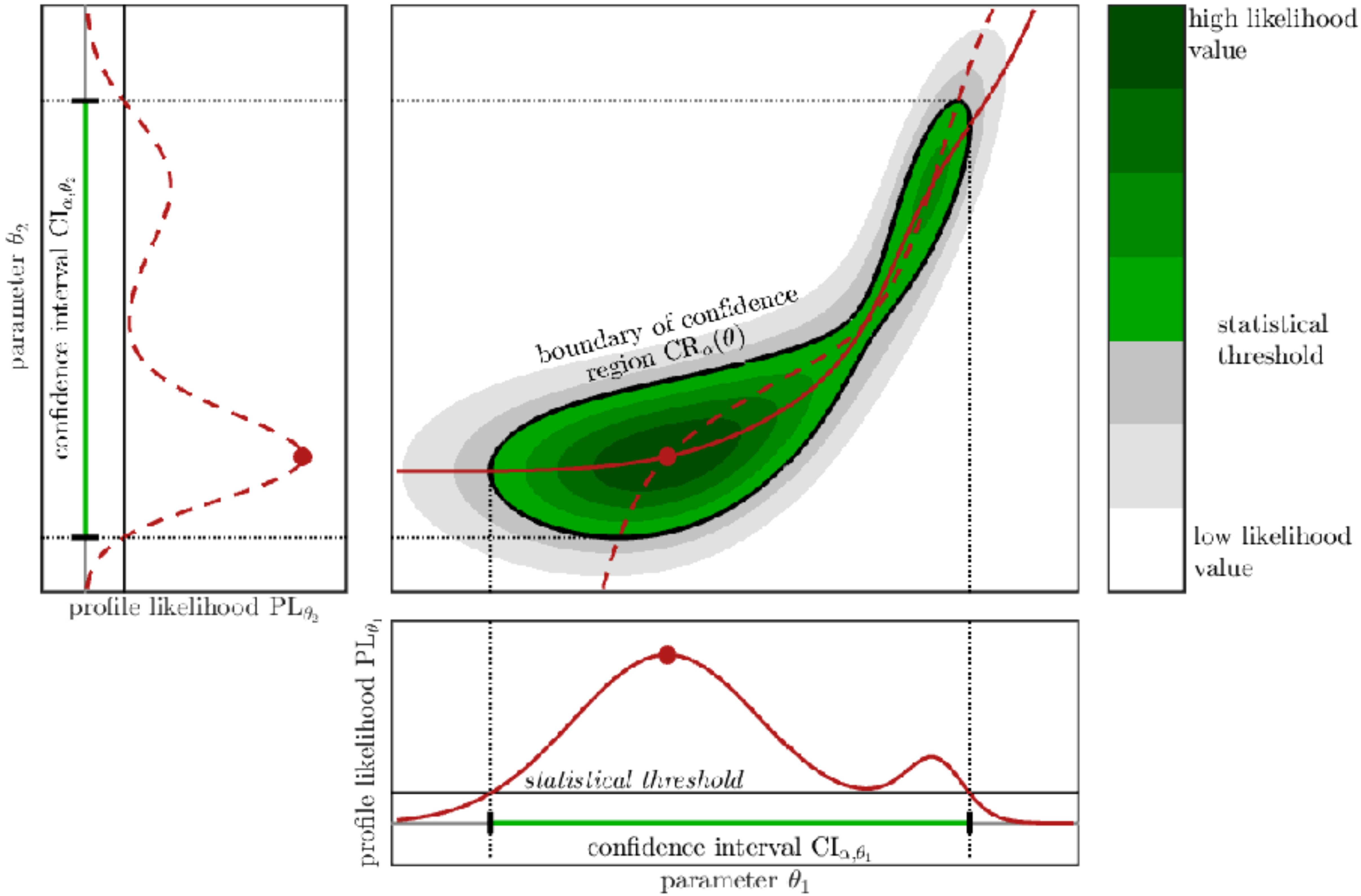
- $p(\theta|\mathcal{D})$: posterior probability of parameters given data
- $p(\mathcal{D}|\theta)$: conditional probability of data given model / likelihood
- $p(\theta)$: prior probability
- $p(\mathcal{D})$: marginal probability of data

Maximum A Posterior (MAP) estimator

The MAP estimate $\theta^{ml} \in \Omega \subseteq \mathbb{R}_+^{n_\theta}$ maximises the posterior probability,

$$\theta^{map} = \arg\max_{\theta \in \Omega} \left\{ p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta) \right\}, \quad \text{subject to } \mathcal{M}(\theta).$$

Illustration of profile likelihood



Confidence regions and intervals

Confidence region

For the parameter vector $\theta \in \Theta$ we define the confidence region to the confidence level α as

$$\mathrm{CR}_\alpha = \left\{ \theta \in \Theta \,\middle|\, \frac{\mathcal{L}_\mathcal{D}(\theta)}{\mathcal{L}_\mathcal{D}(\hat{\theta})} \geq \exp\left(-\frac{\Delta_\alpha}{2}\right) \right\},$$
$$= \left\{ \theta \in \Theta \,\middle|\, 2\left(J(\theta) - J(\hat{\theta})\right) \leq \Delta_\alpha \right\},$$

with Δ_α denoting the αth-percentile of the χ^2 distribution with one degree of freedom.

Model property $g(\theta)$, e.g.

- individual parameter: $g(\theta) = \theta_j$
- state x_j a time point T: $g(\theta) = x_j(T, \theta)$

Confidence interval

The confidence interval for a model property $g(\theta)$ is the projection of CR_α onto $g(\theta)$,

$$\mathrm{CI}_{\alpha,g(\theta)} = P_{g(\theta)}\mathrm{CR}_\alpha = \{c \,|\, \exists\theta \in \mathrm{CR}_\alpha \wedge g(\theta) = c\}.$$

Profile likelihood and confidence interval

Profile likelihood

For the model property $g(\theta)$ we define the profile likelihood as

$$\mathrm{PL}_{g(\theta)}(c) = \max_{\theta \in \Theta} \mathcal{L}_{\mathcal{D}}(\theta) \text{ subject to } g(\theta) = c.$$

For values c outside the range of $g(\theta)$, $\mathrm{PL}_{g(\theta)}(c) = 0$.

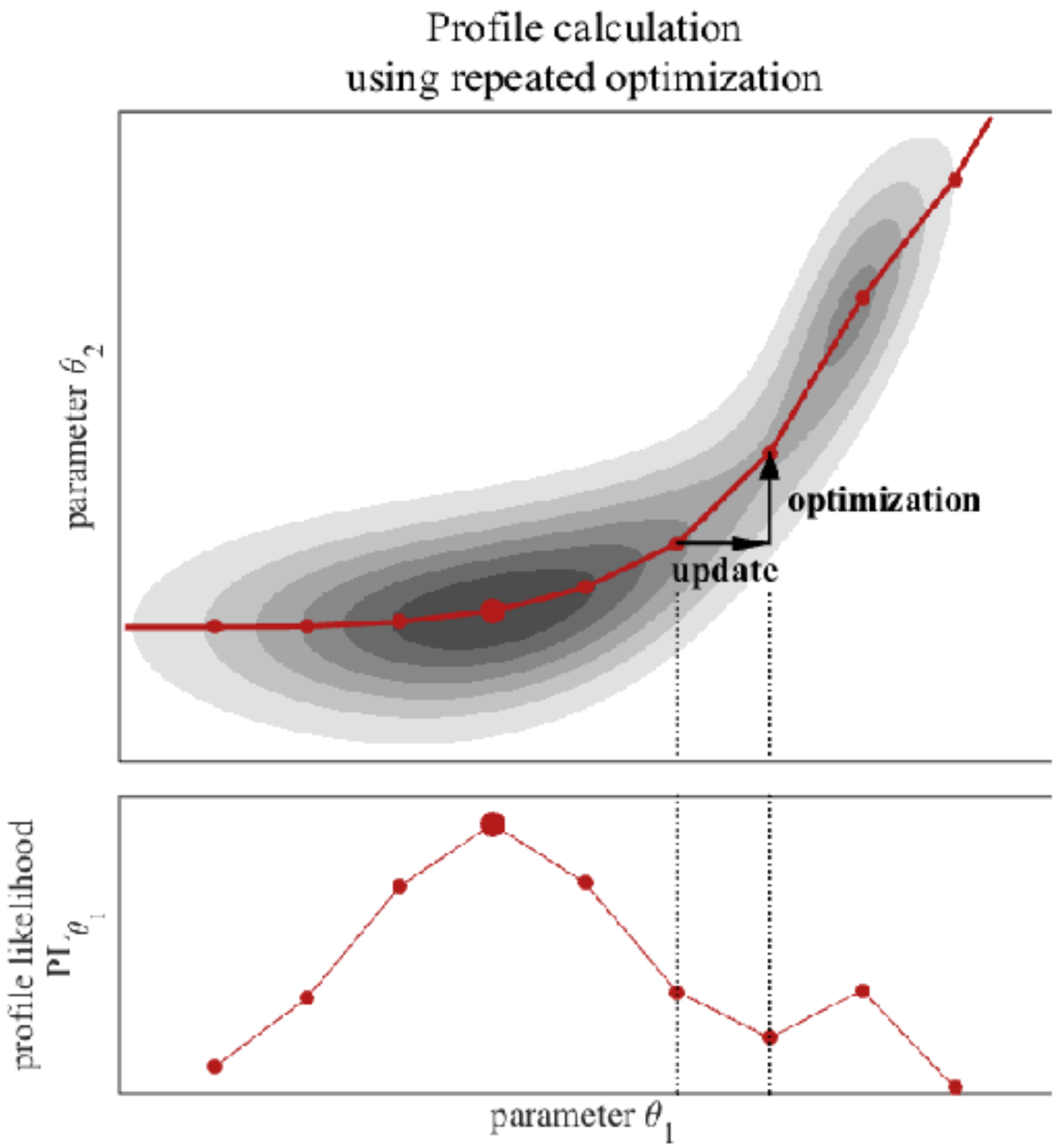
From the profile likelihood the confidence interval for $g(\theta)$ follows as

$$\mathrm{CI}_{\alpha,g(\theta)} = \left\{ c \,\middle|\, \frac{\mathrm{PL}_{g(\theta)}(c)}{\mathcal{L}_{\mathcal{D}}(\hat{\theta})} \geq \exp\left(-\frac{\Delta_\alpha}{2}\right) \right\}.$$

Remark:

- Profile likelihoods facilitate the calculation of confidence intervals without the evaluation of the confidence region or its projection.
- Profile likelihood based confidence intervals are also called “finite sample confidence intervals”.

Optimisation-based profile likelihood calculation



Optimisation-based profile likelihood calculation

Profile likelihood

Sequence of constraint optimisation problems,

$$\min_{\theta \in \Theta} J(\theta) \text{ subject to } g(\theta) = c,$$

for values c which are either on a grid or chosen adaptively.

Implementation as sequence of local optimisation problems with starting point

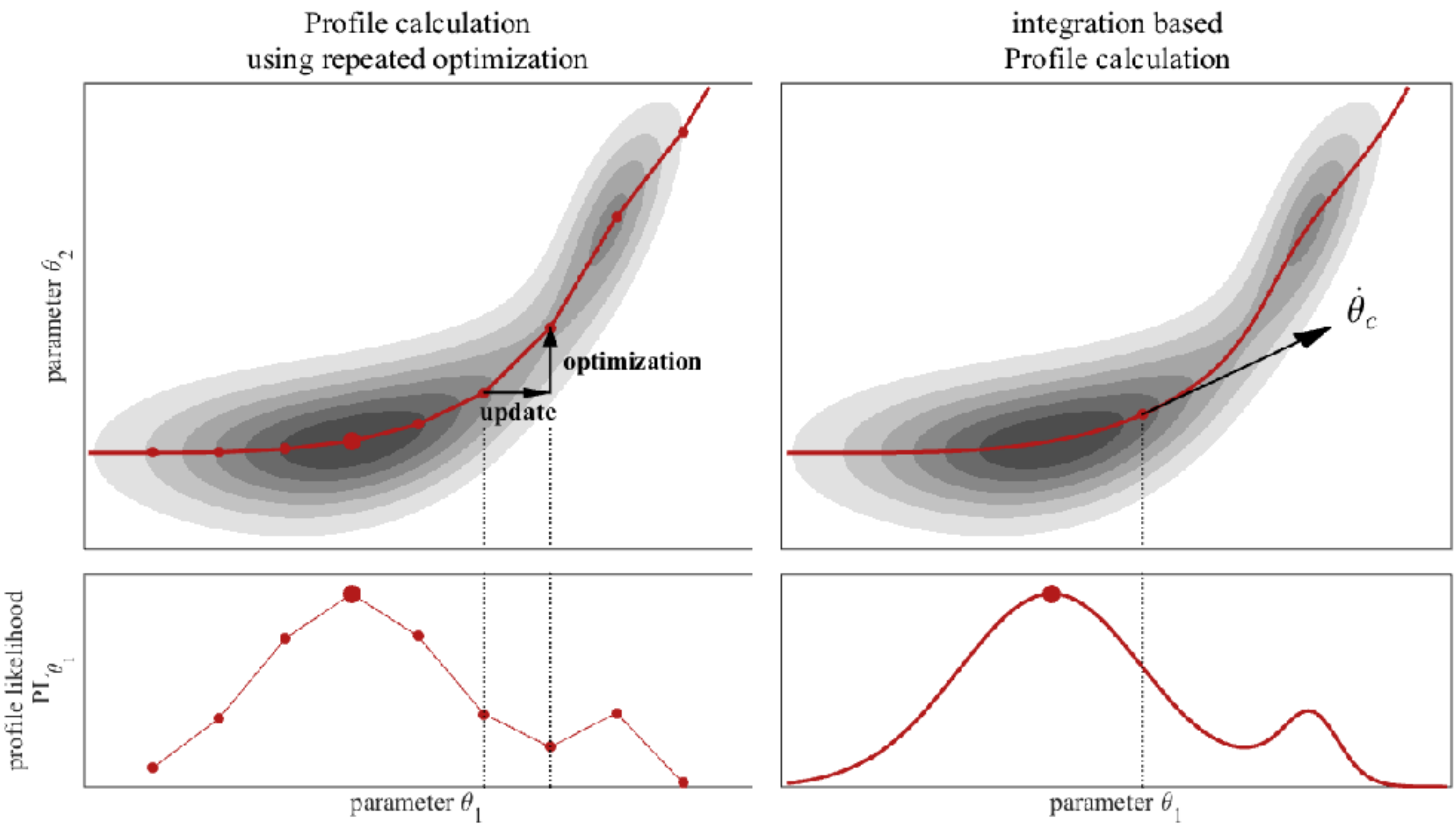
1. **0th order proposal:** the optimal point for c_{l-1}, $\theta_{c_l}^{(0)} = \theta_{c_{l-1}}$, or
2. **1st order proposal:** the linear extrapolation based on the optimal points for c_{l-1} and c_{l-2},

$$\theta_{c_l}^{(0)} = \theta_{c_{l-1}} + \frac{c_l - c_{l-1}}{c_{l-1} - c_{l-2}} (\theta_{c_{l-1}} - \theta_{c_{l-2}}).$$

Properties:

- Large number of local optimisations.
- (Relatively) efficient and robust implementation. (see D2D and PESTO)
- Potentially initialisation at multiple local optima required which are above the statistical threshold.

Optimisation- and integration-based profile likelihood calculation



Integration-based profile likelihood calculation

Lagrange function of constraint optimisation problem

$$\ell(\theta) = J(\theta) + \lambda(g(\theta) - c),$$

with Lagrange multiplie $\lambda \in \mathbb{R}$, yielding the first order optimality conditions,

$$\begin{aligned} \nabla_\theta J(\theta) + \lambda \nabla_\theta g(\theta) &= 0 \\ g(\theta) &= c \end{aligned}$$

The optimal point depends on c: $\theta = \theta(c)$ and $\lambda = \lambda(c)$

Integration-based profile likelihood calculation

Differentiation of the optimality condition yields the differential algebraic equation (DAE)

$$\underbrace{\begin{pmatrix} \nabla_\theta^2 J(\theta_c) + \lambda_c \nabla_\theta^2 g(\theta_c) & \nabla_\theta g(\theta_c) \\ \nabla_\theta g(\theta_c)^T & 0 \end{pmatrix}}_{:=M(\theta_c)} \begin{pmatrix} \dot{\theta}_c \\ \dot{\lambda}_c \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \underbrace{- \begin{pmatrix} \gamma \nabla_\theta J(\theta_c) \\ 0 \end{pmatrix}}_{=\text{“stabilisation”}}$$

The solution of this DAE for a starting point which solves the constraint optimisation problem for $c = c_0$ yields the profile θ_c for $c \in [c_0, c_{\mathrm{end}}]$.

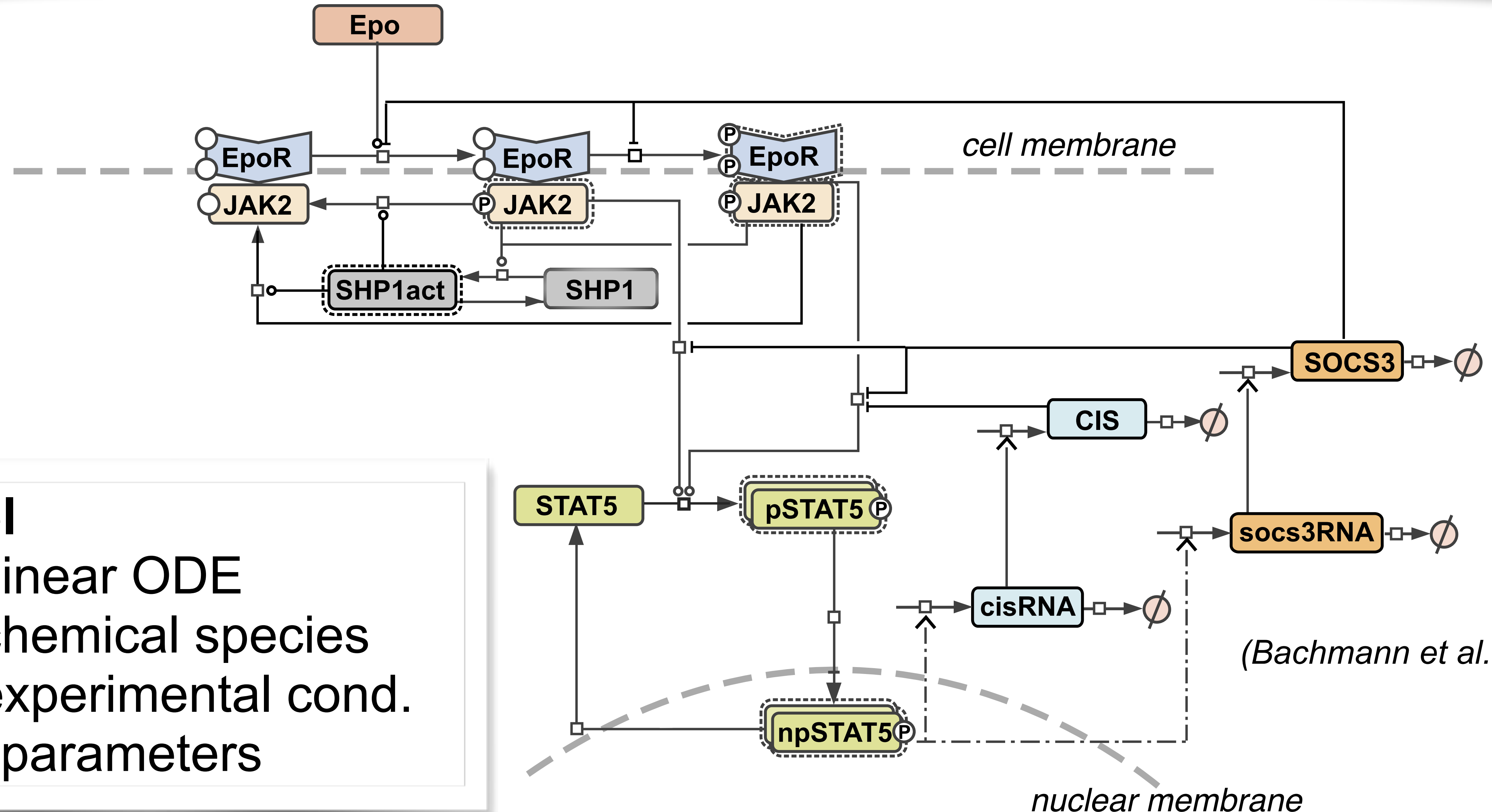
Some applications of profile likelihoods and profile posteriors

Model-based analysis of Epo-signaling

S. Hug, A. Raue, J. Hasenauer, J. Bachmann, U. Klingmüller, J. Timmer, and F. J. Theis. High-dimensional Bayesian parameter estimation: Case study for a model of JAK2/STAT5 signaling, *Mathematical Biosciences,* 246(2):293-304, 2013.

Biological system

Background: Used during cancer therapy to reduce side effects.
Problem: Increases also survival probability of cancer cells.



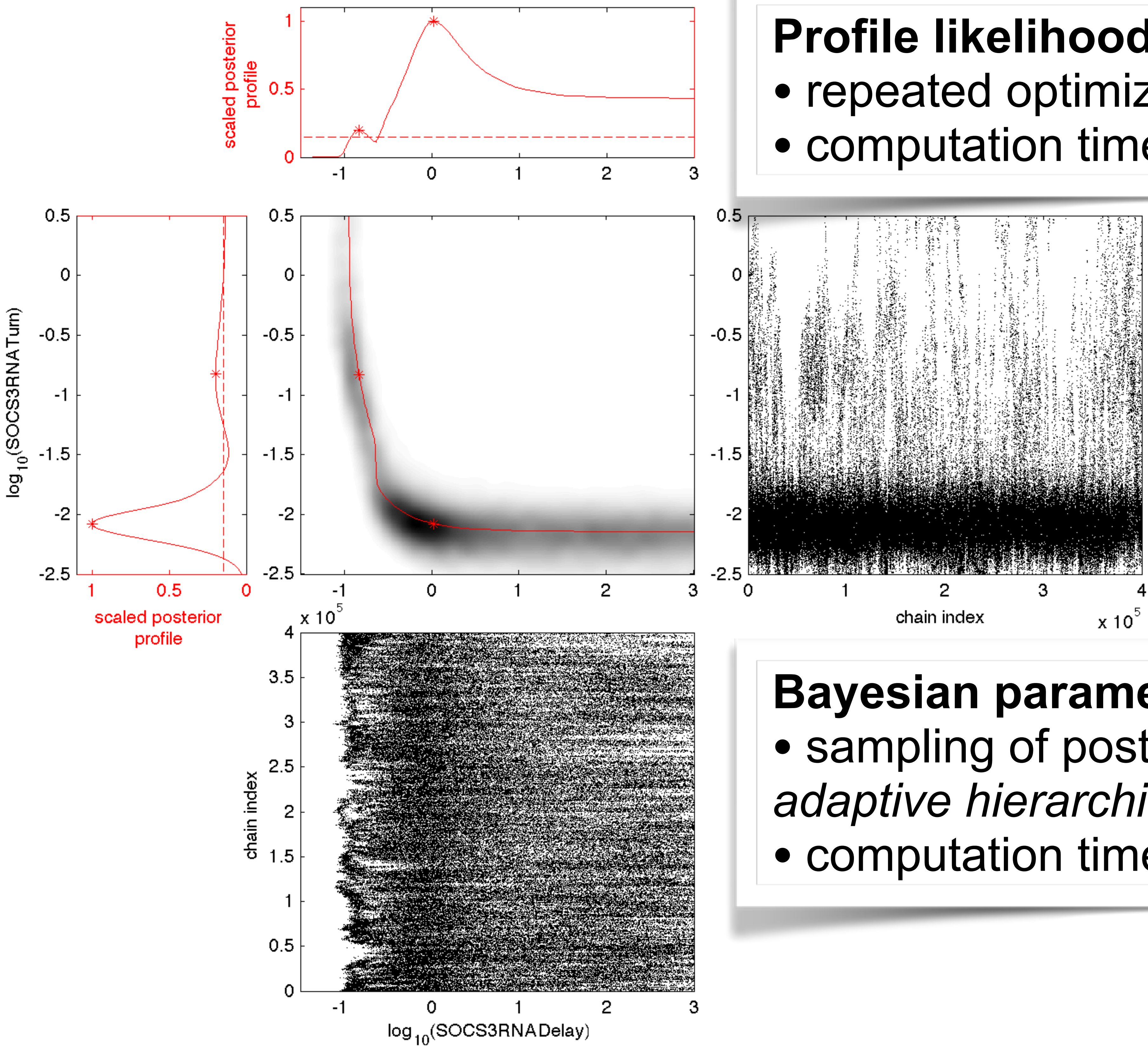
(Bachmann et al., 2007)

Model

- nonlinear ODE
- 25 chemical species
- 24 experimental cond.
- 113 parameters

Key question: Optimal Epo dosis during chemotherapy?
⇒ need for predictive models

Profile likelihoods and Bayesian methods



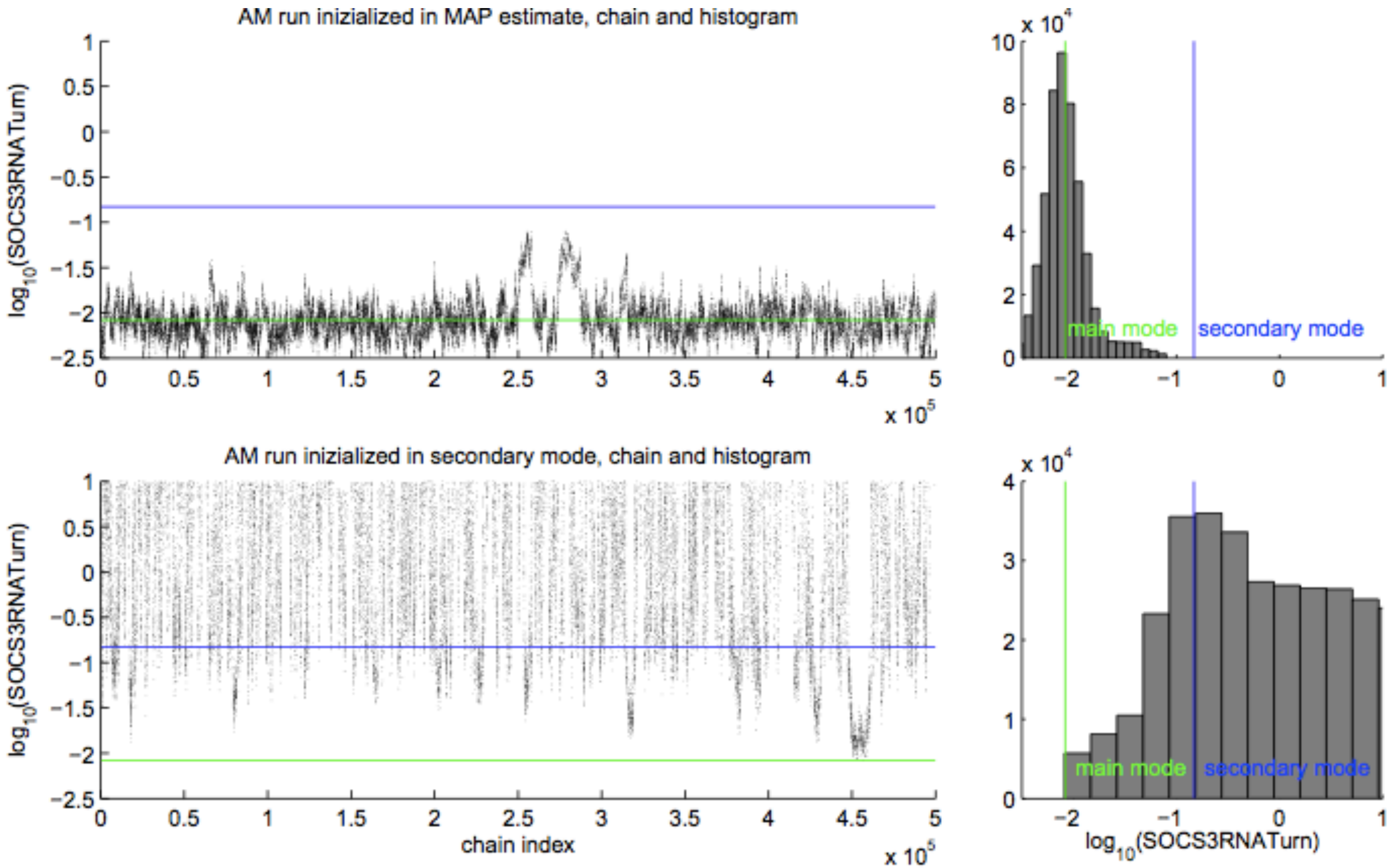
Profile likelihood calculation

- repeated optimization: ~60,000 times
- computation time: ~1 day

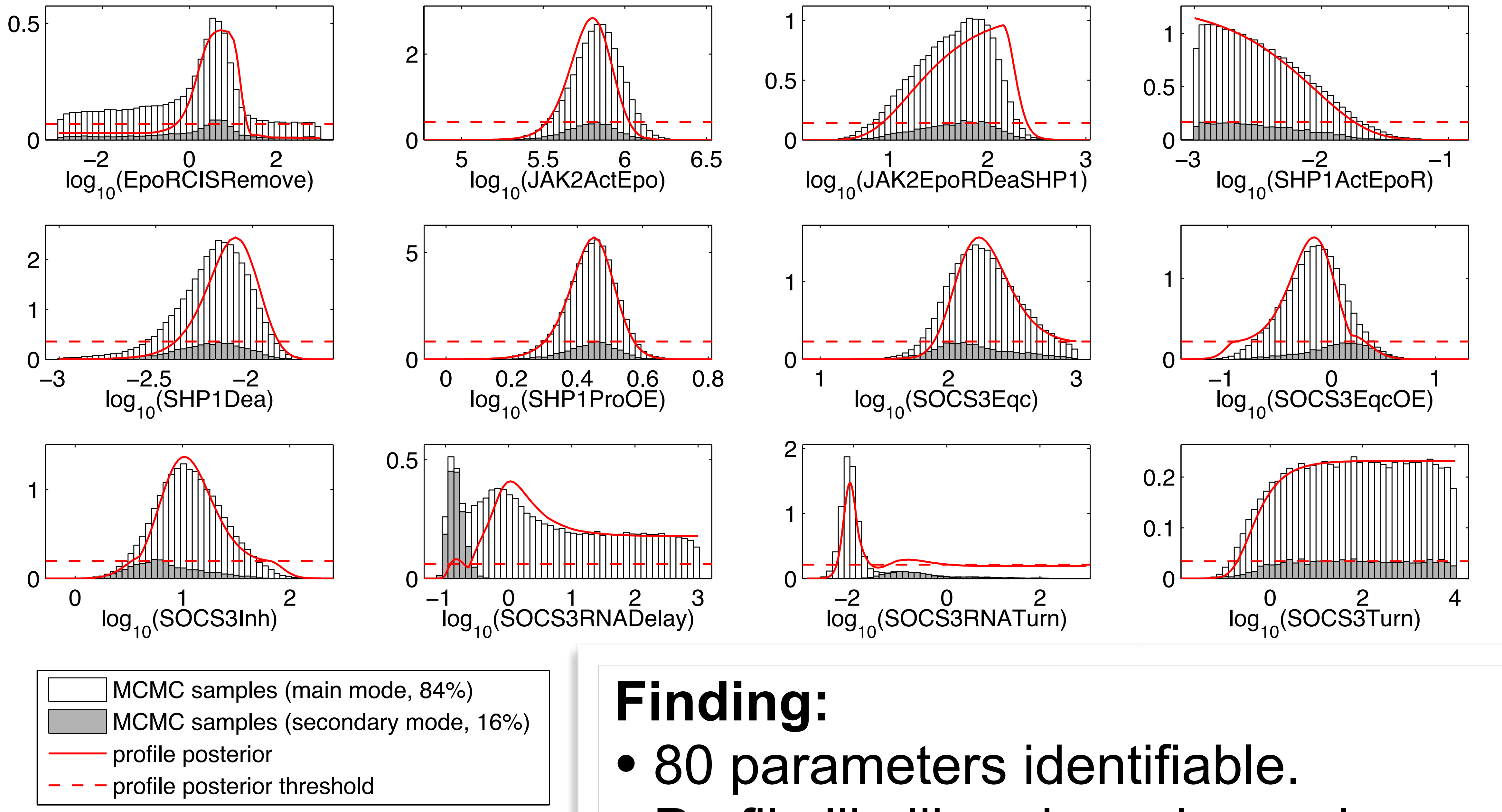
Bayesian parameter estimation

- sampling of posterior distribution using *adaptive hierarchical sampling*
- computation time: ~20 day

Sampling properties of single-chain methods



Parameter uncertainties using profile likelihoods and Bayesian methods



Finding:

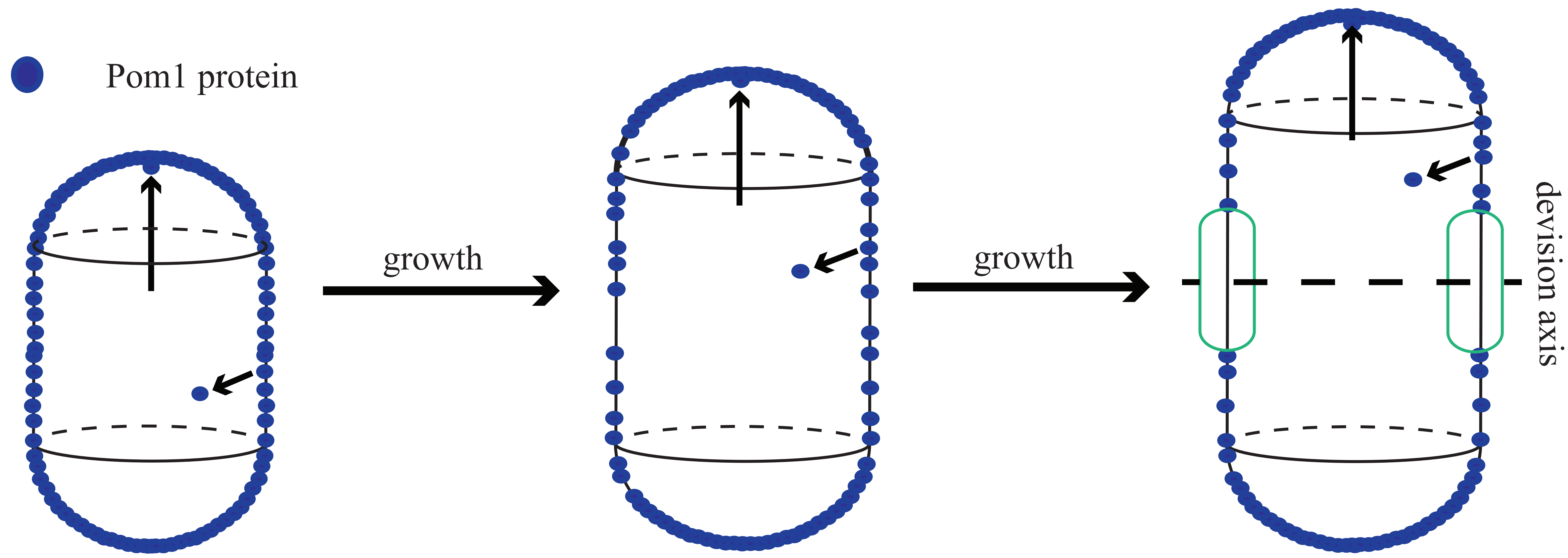
- 80 parameters identifiable.
- Profile likelihoods and sample histograms agree well.
- Mode weights are different.

Model-based analysis of Pom1p gradient formation

R. Boiger, J. Hasenauer, S. Hross and B. Kaltenbacher. Integration based profile likelihood calculation for PDE constrained parameter estimation problems. *Inverse Problem,* 32(12):125009, 2016.

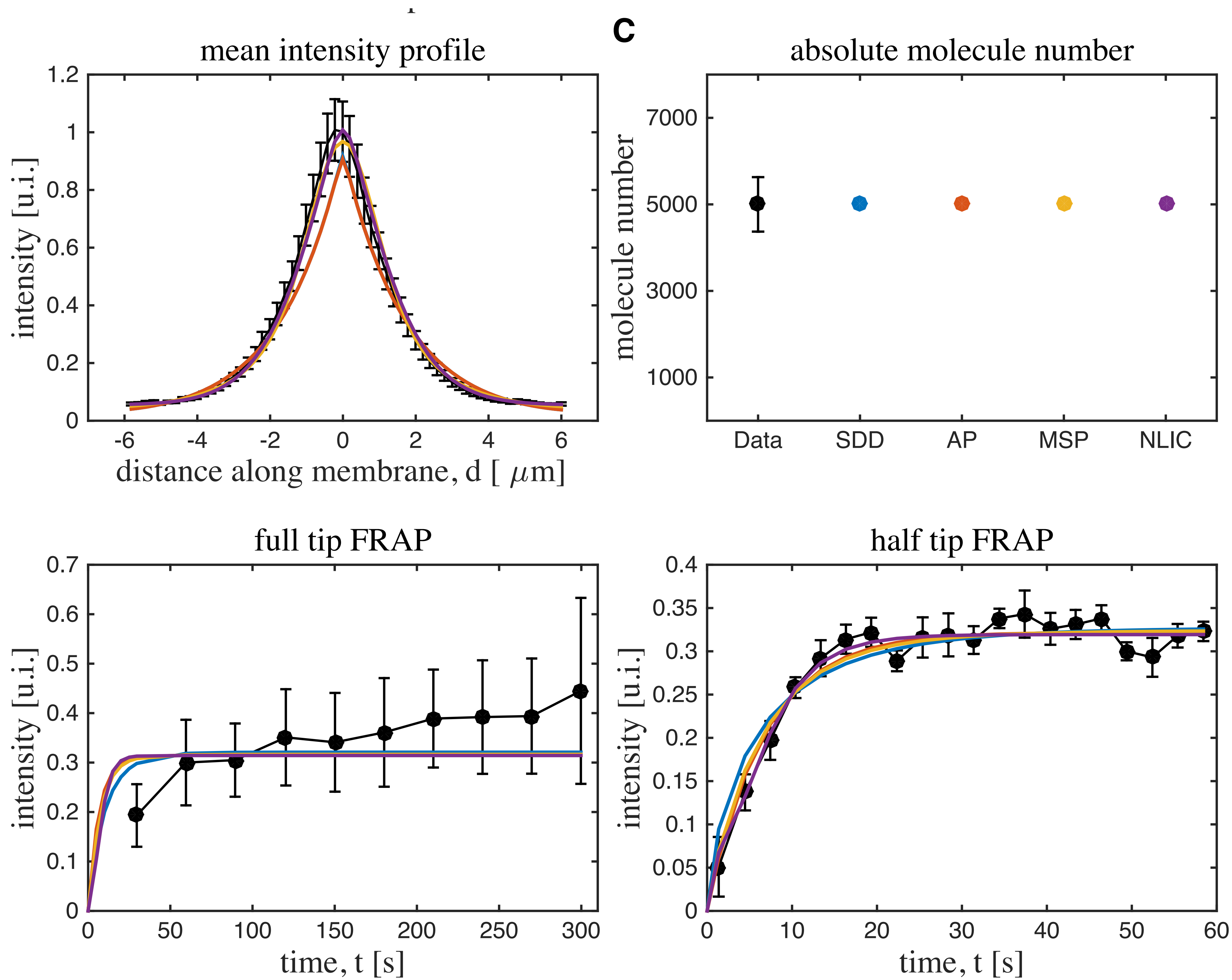
Biological system

Background: Pom1 controls cell division.
Problem: Competing hypotheses how the gradient is formed.

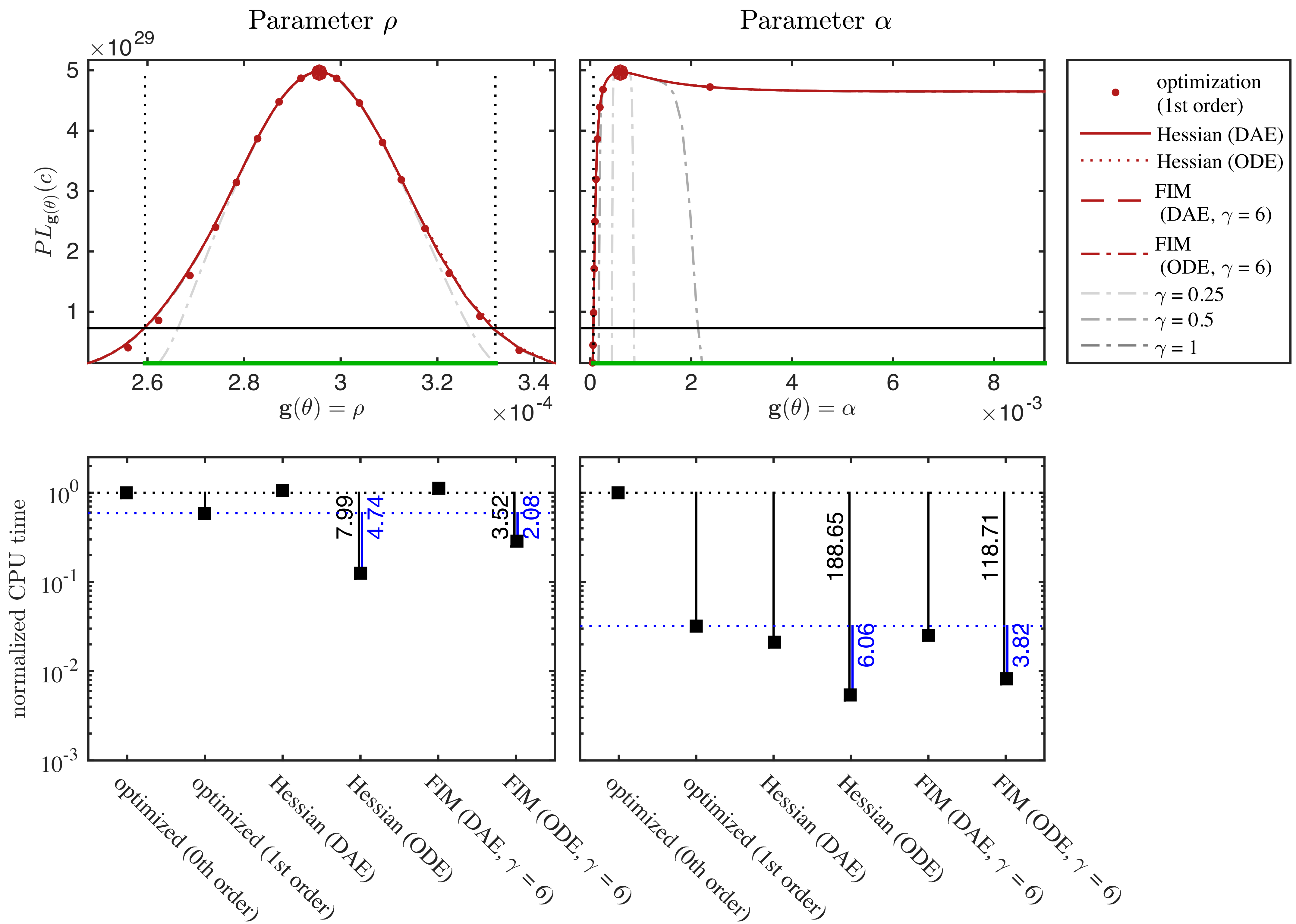


Key question: Which model topology provides a better description of the experimental and and yields testable hypotheses?

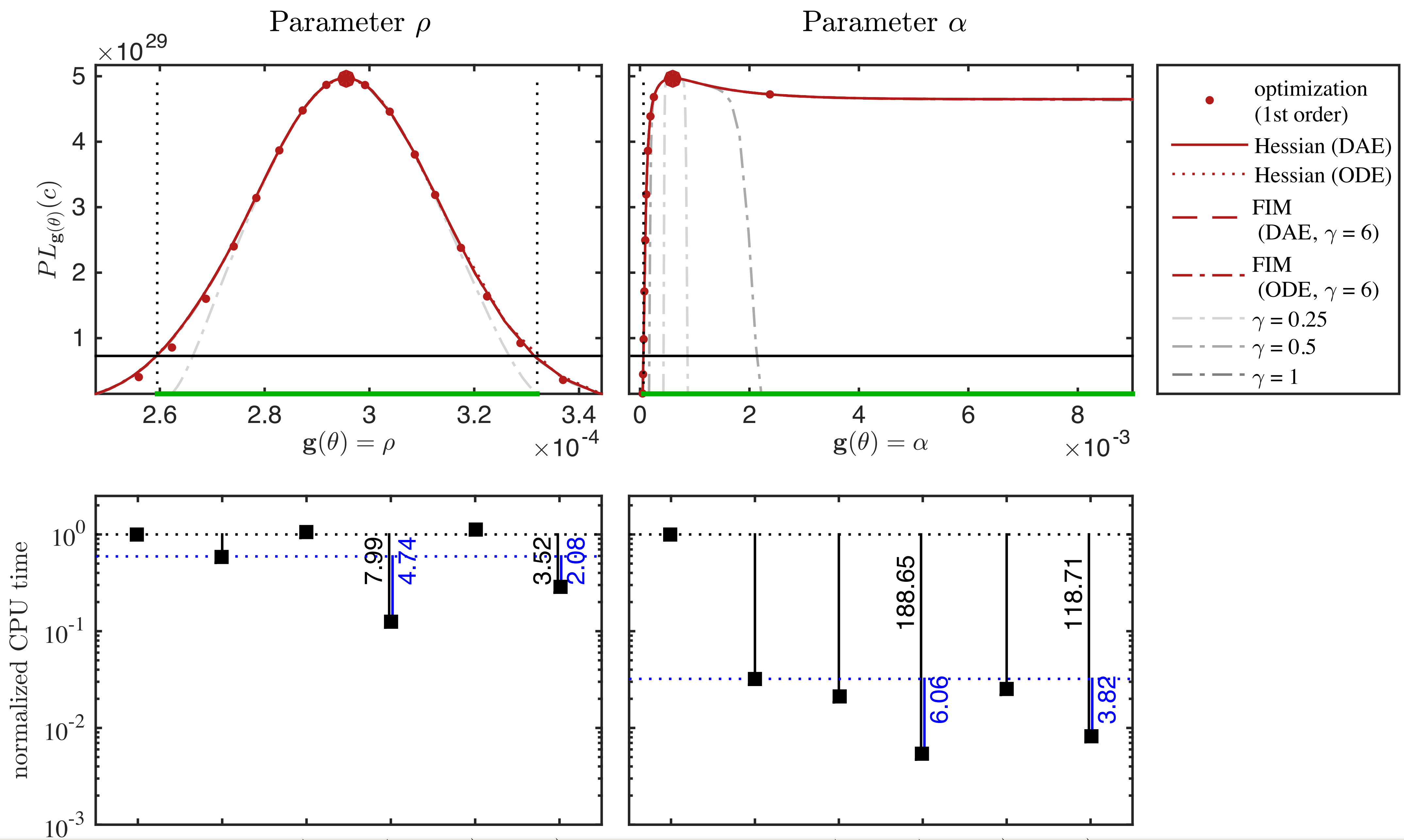
Comparison of different hypotheses



Computation time for different profile calculation methods



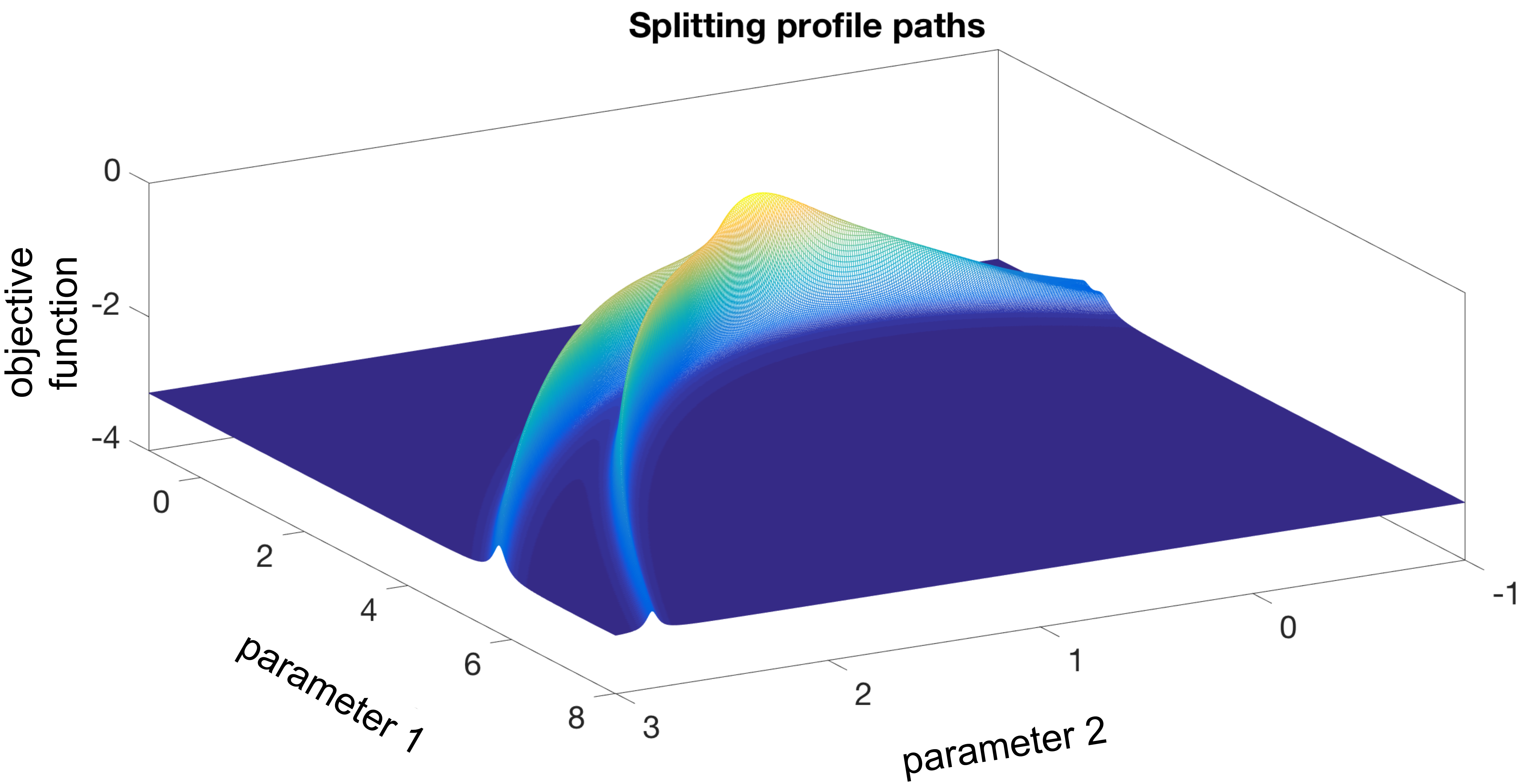
Computation time for different profile calculation methods



Integration-based methods outperform here optimisation-based approaches.

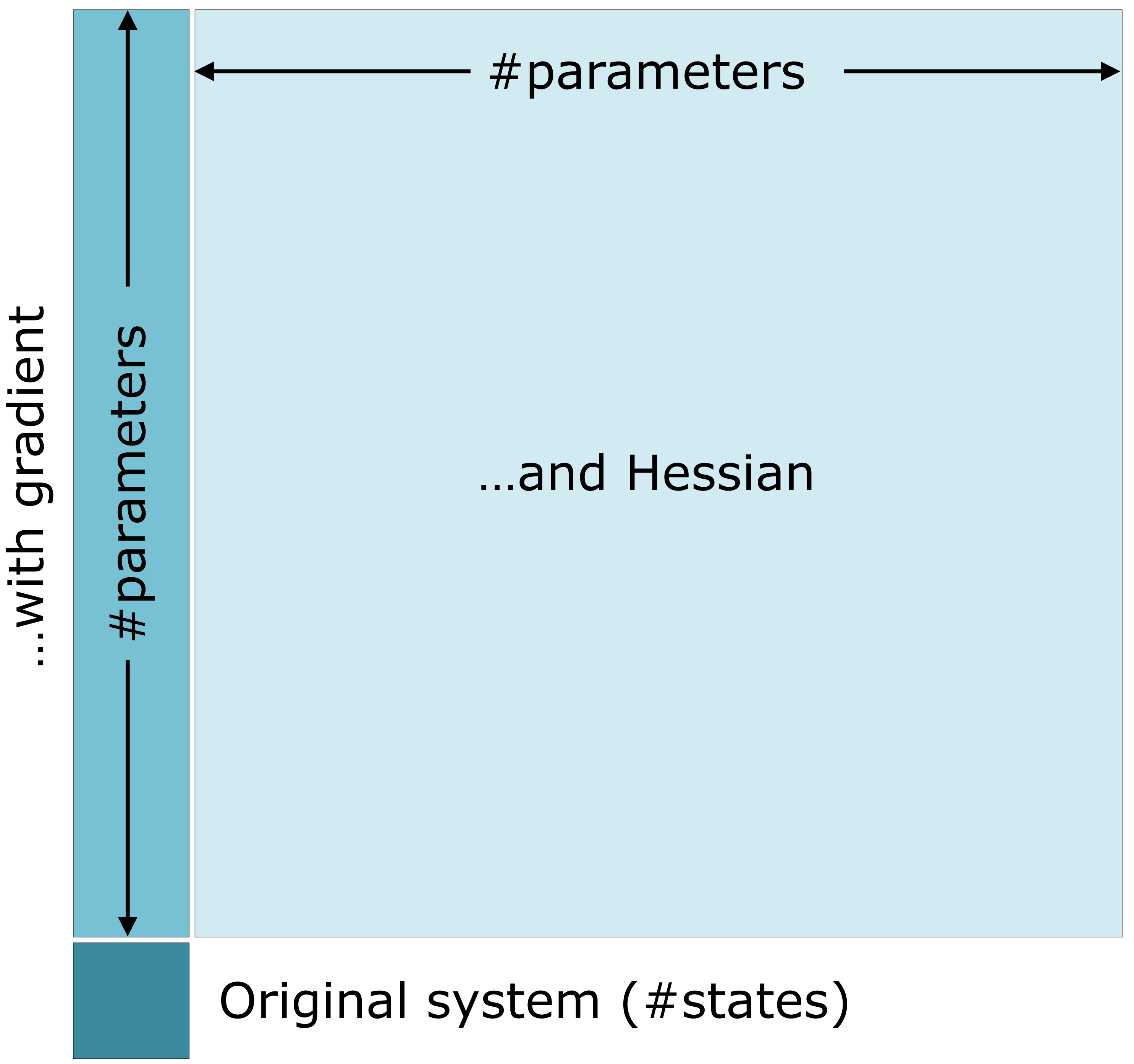
Some challenges and ideas

“Stiffness” of DAE for integration-based profile calculation

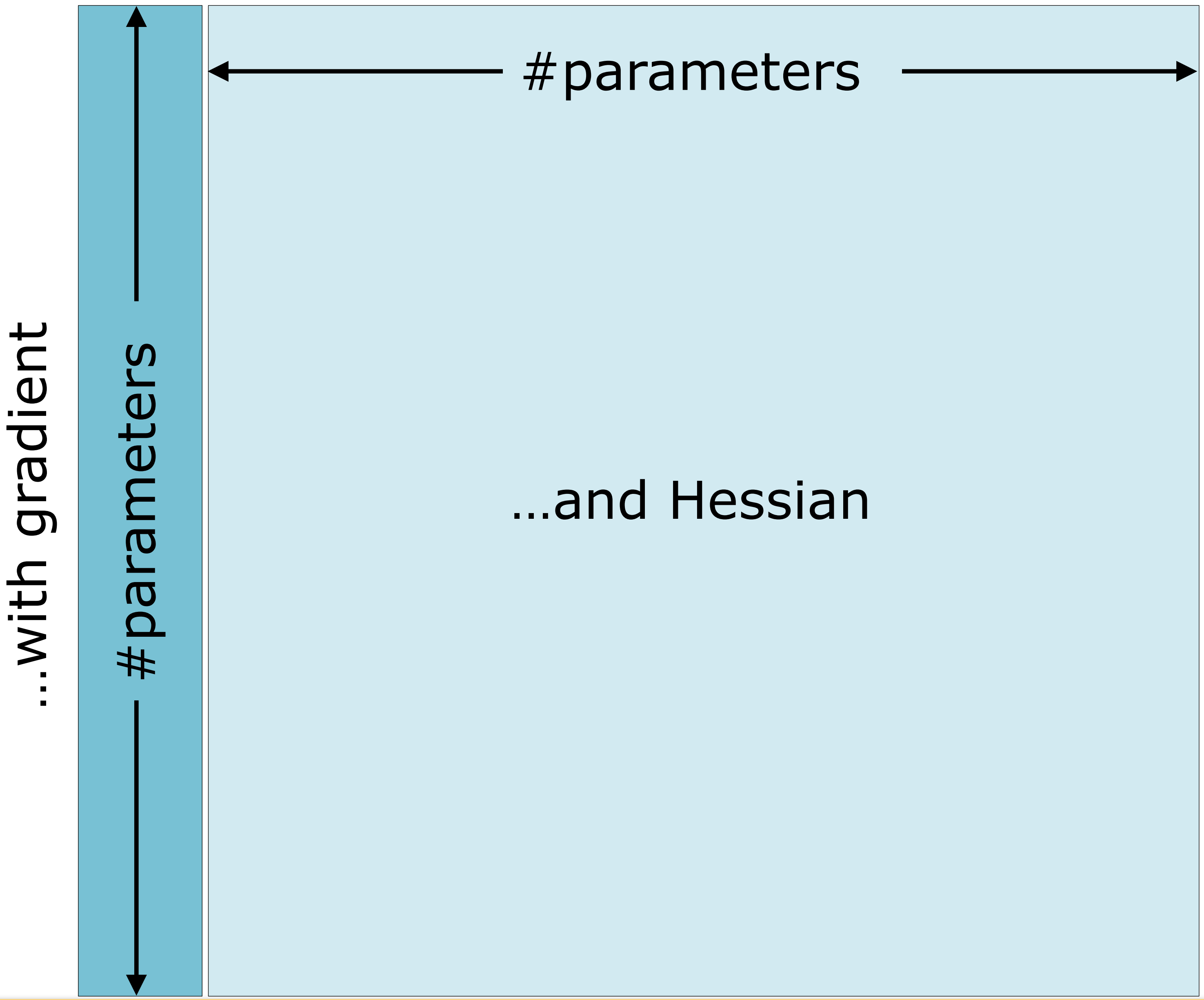


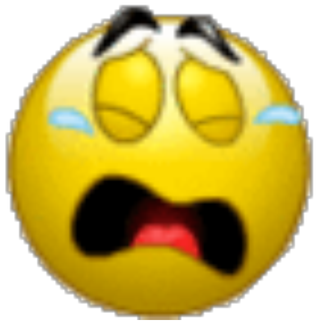
Hybrid profile calculation schemes seem to be promising.

Efficient calculation or reliable approximation of Hessian



Efficient calculation or reliable approximation of Hessian





2nd order adjoints and conjugated gradient methods might be an interesting approach.

Summary and conclusion

Summary and conclusion

- **Interpretation of profiles**
- **Calculation of profiles**
  - Optimisation-based method
  - Integration-based method
  - Hybrid method

  ⇒ Implemented in the MATLAB Toolbox PESTO
- **Comparison of profiles and marginals**
- **Comparison of computation time**

Personal conclusion / experience:

- Profile calculation nicely complements sampling-based approaches
- For problems for which efficient (local) optimisers are available, profiles calculation can be more efficient

Acknowledgment

Horizon 2020
European Union funding
for Research & Innovation

Helmholtz Zentrum München
Anna Fiedler
Fabian Fröhlich
Sabrina Hross
Sabine Hug
Paul Stapor
Fabian J Theis

Bundesministerium
für Bildung
und Forschung

e:Med
SYSTEMS MEDICINE

University of Freiburg
Andreas Raue
Jens Timmer