

Importance of Assessing Model Convergence and Practical Identifiability During Model Development

Andreas Raue and Andrew Stein
October 2018, 2017 – ACoP

With contributions from:
Kostas Biliouris and Stas Tsitkov

Key Message:

- **Likelihood Waterfalls** are recommended for assessing optimizer convergence and the existence of local minima.
- **Likelihood Profiles** are recommended for assessing the identifiability of pharmacometrics and systems pharmacology models.
- Maximum off-diagonal correlation of covariance matrix > 0.98 seems to be sign of convergence or identifiability issues. More examples needed.

Outline

- **Log Likelihood Waterfall (LLW)**

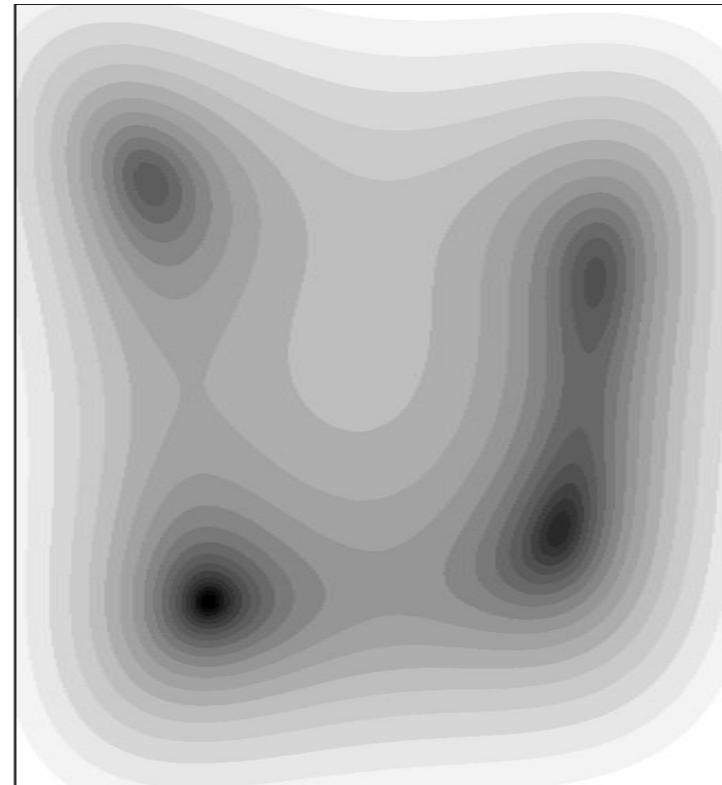
- Introduce Log Likelihood Waterfall
 - Demonstrate LLW with examples

- **Log Likelihood Profiling (LLP)**

- Introduce Log Likelihood Profile
 - Demonstrate LLP with examples

Background

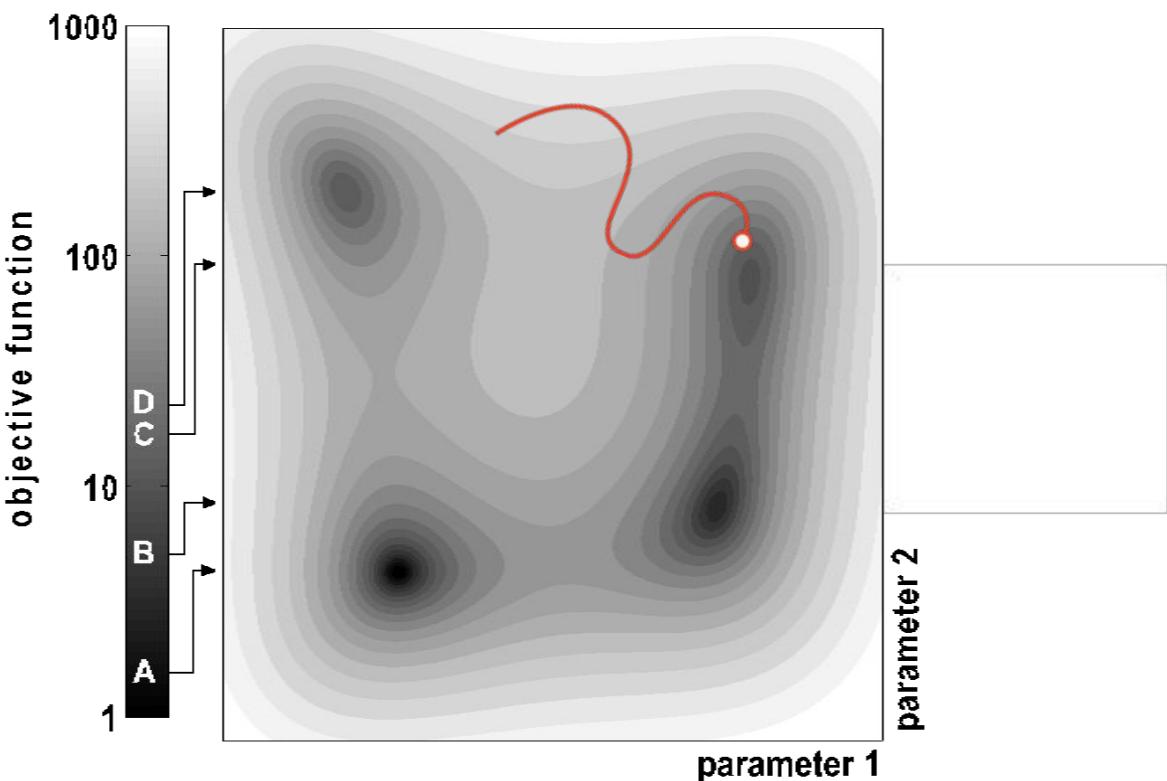
- Pharmacometrists and Systems Pharmacologists commonly follow the recipe that if the optimizer stops with no warnings and if diagnostics (visual predictive check, residuals) look reasonable, then the model is to be trusted.
- However, standard diagnostics do not test for convergence or over-parameterization of the model.
- Consequences of over-parameterization:
 - Underestimate the true uncertainty in model predictions
 - Errors and overconfidence in model extrapolation to other dosing regimens
 - Incorrect physiological interpretation of data.



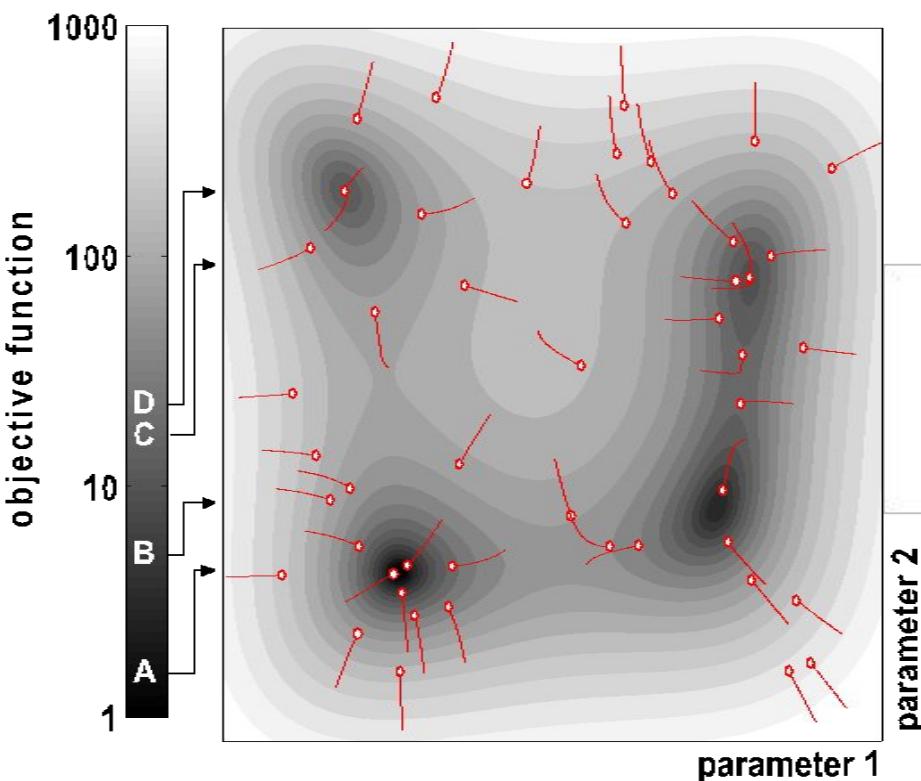
Himmelblau, D. et al. Applied nonlinear programming
(McGraw-Hill, New York, 1972)

How to test for convergence?

A single parameter estimation can get “trapped” in local minima

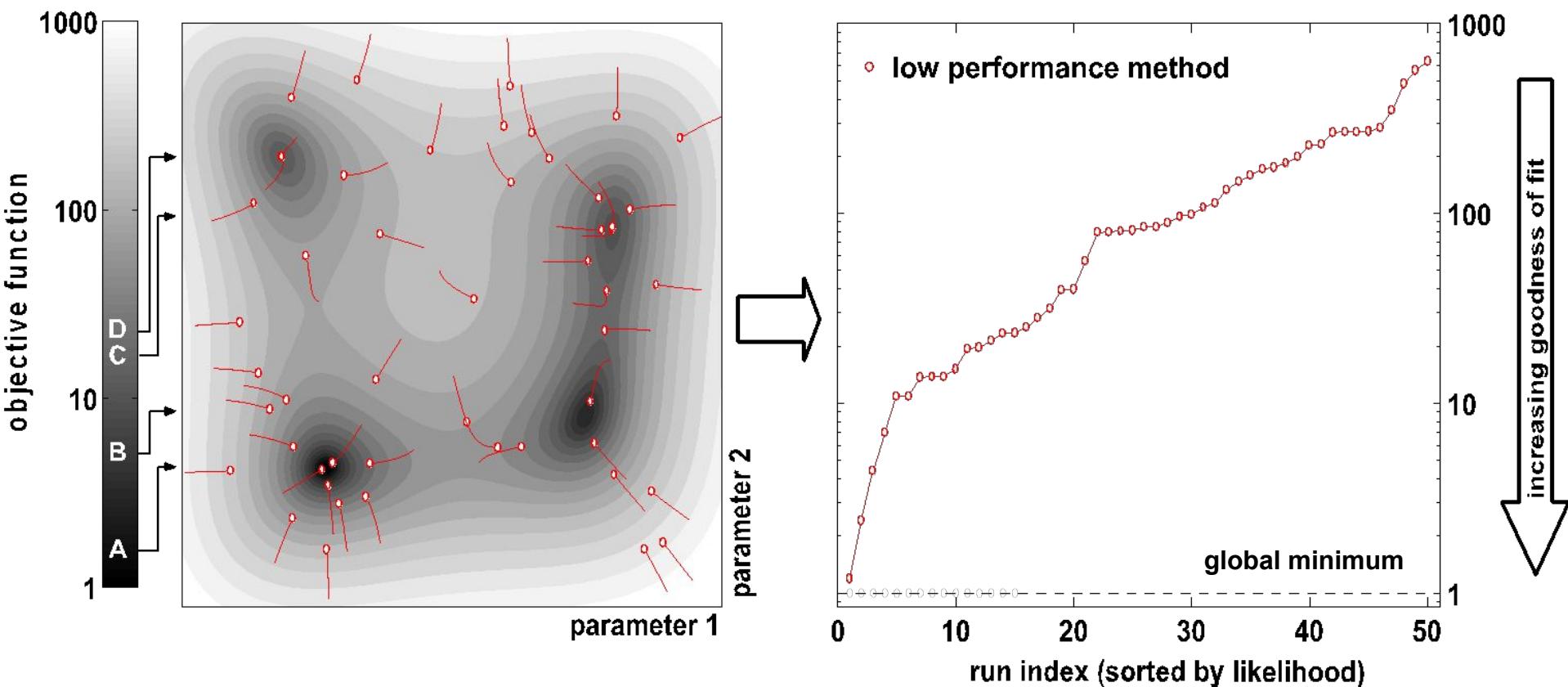


Multiple parameter estimations can help finding better solutions

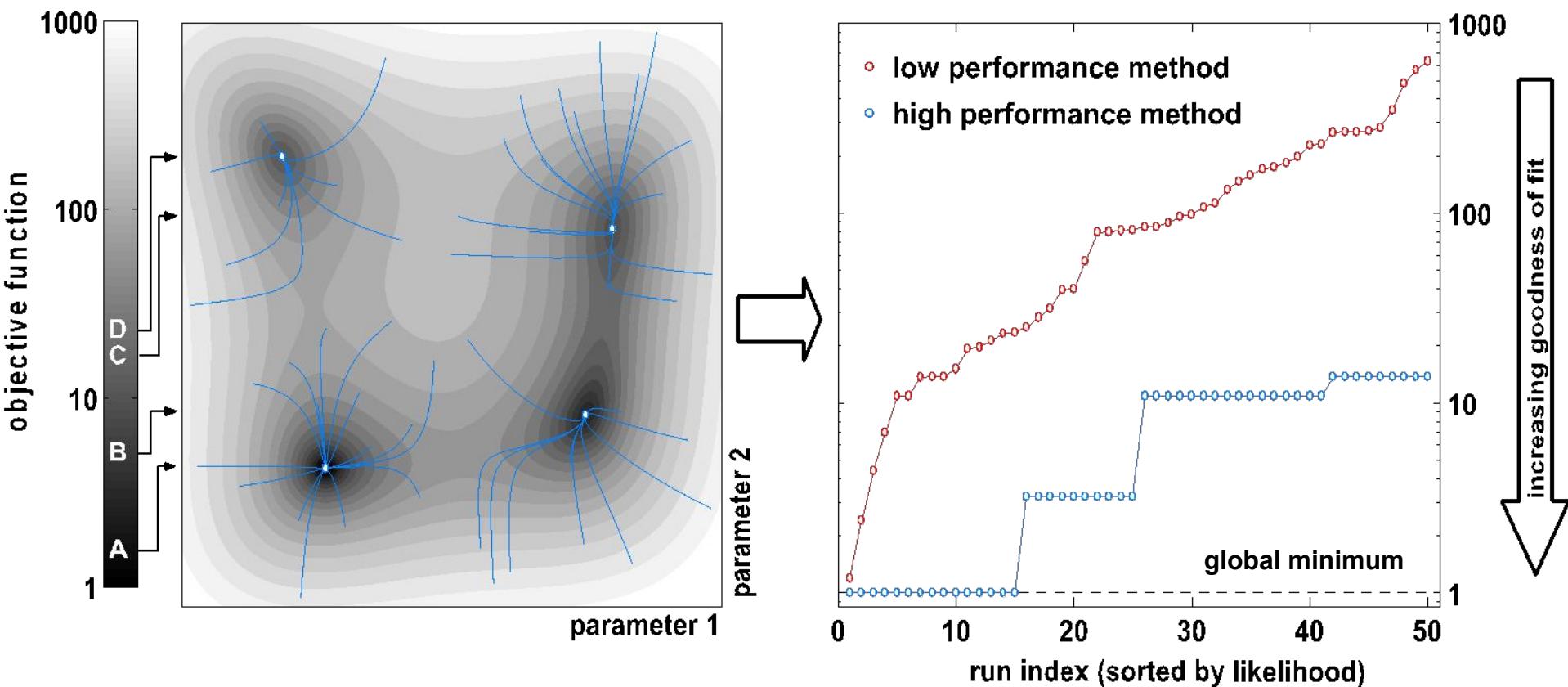


- Did we find the best solution (global minimum) ?
- How many runs are necessary ?
- How well does the method work ?

QC for Parameter Estimation: Convergence Waterfall Plot

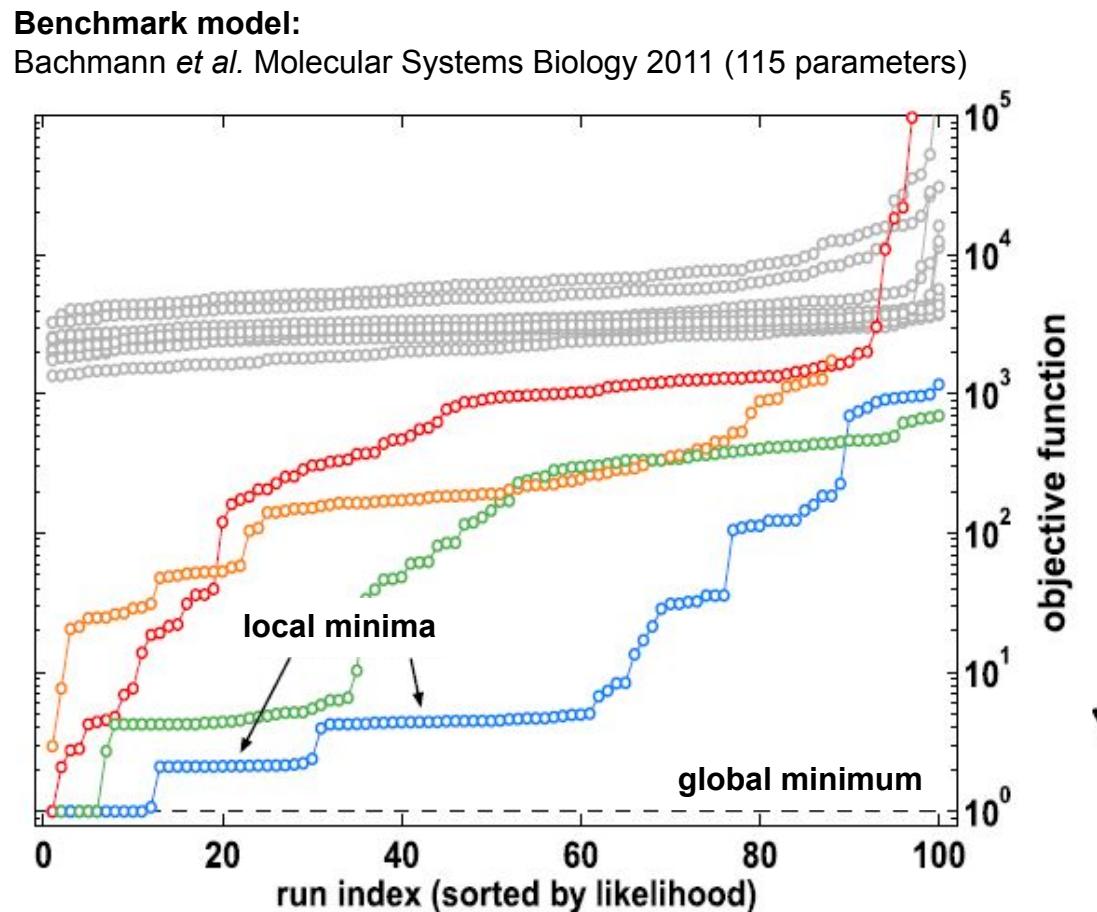


QC for Parameter Estimation: Convergence Waterfall Plot



Convergence Waterfall Plot for Larger Benchmark Model

- stochastic optimization (12 different algorithms)
- deterministic optimization (finite differences)
- hybrid optimization
- deterministic optimization (sensitivity equations)



Convergence was assessed for 6 of models in NONMEM using PsN

Name	Number of Fixed Effects Fit (theta)	Number of Random Effects Fit (eta)
Indirect Response v1	5	1
Indirect Response v2	5	2
Custom PKPD Model	5	1
TMDD (Binding)	14	5
Mechanistic Model #1	16	10
Mechanistic Model #2	13	6

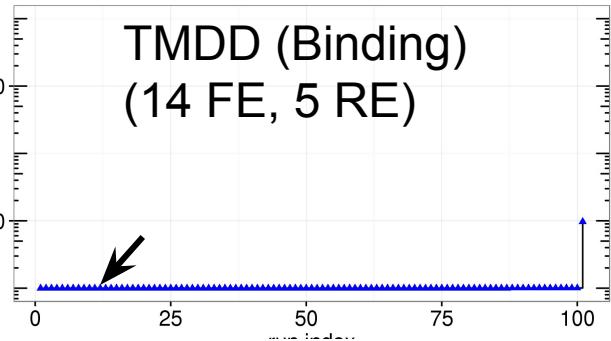
Log Likelihood Waterfall generated with PsN command:

```
parallel_retries-4.2.0 filename.mod -threads=101  
-degree=0.99 -min_retries=100 -picky -seed=12345
```

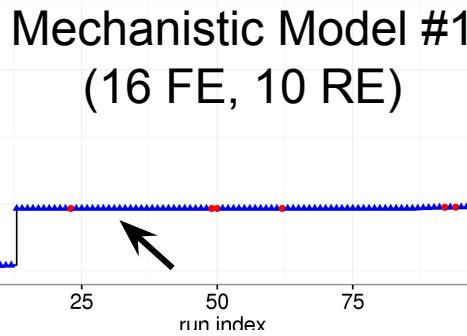
Log Likelihood Waterfalls

Run Time:
10 min – 3 days

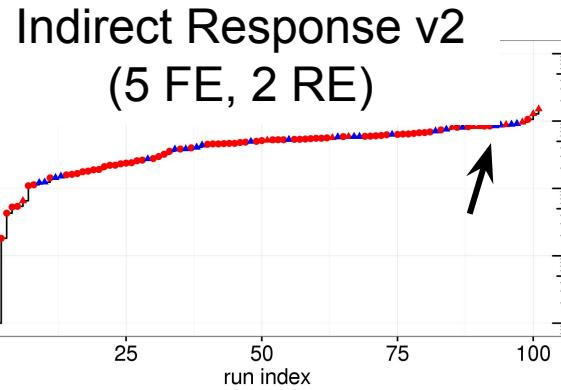
Objective Function Difference + 1



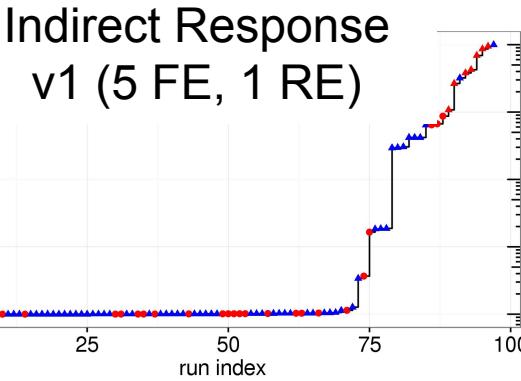
Objective Function Difference



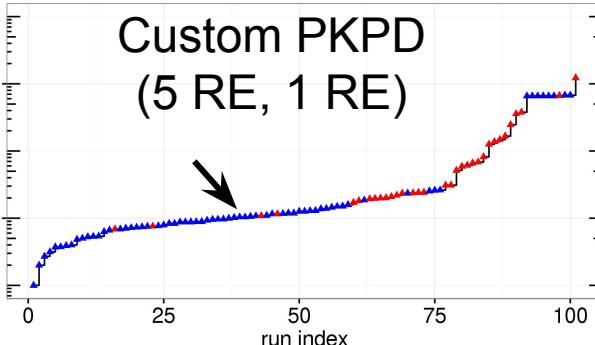
● Success
● Failure
→ Original Fit



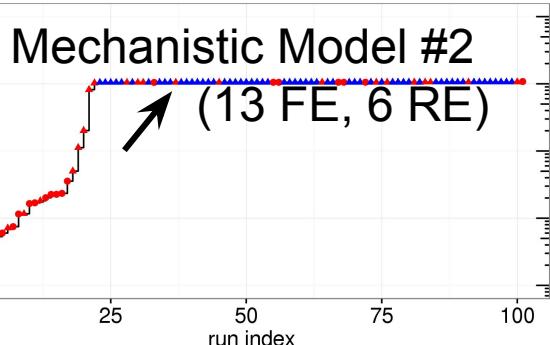
Objective Function Difference



Objective Function Difference



Objective Function Difference



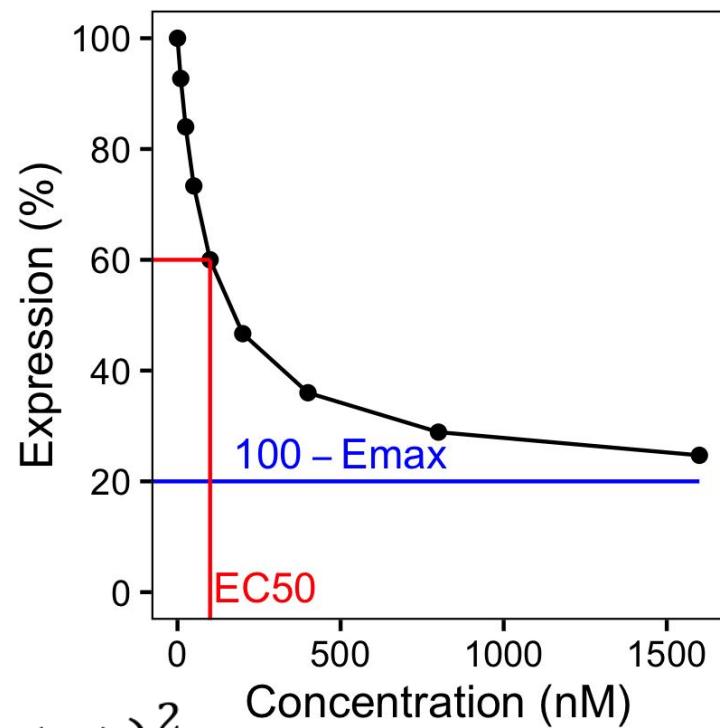
run index (1-100)

run index (1-100)

Likelihood profiles

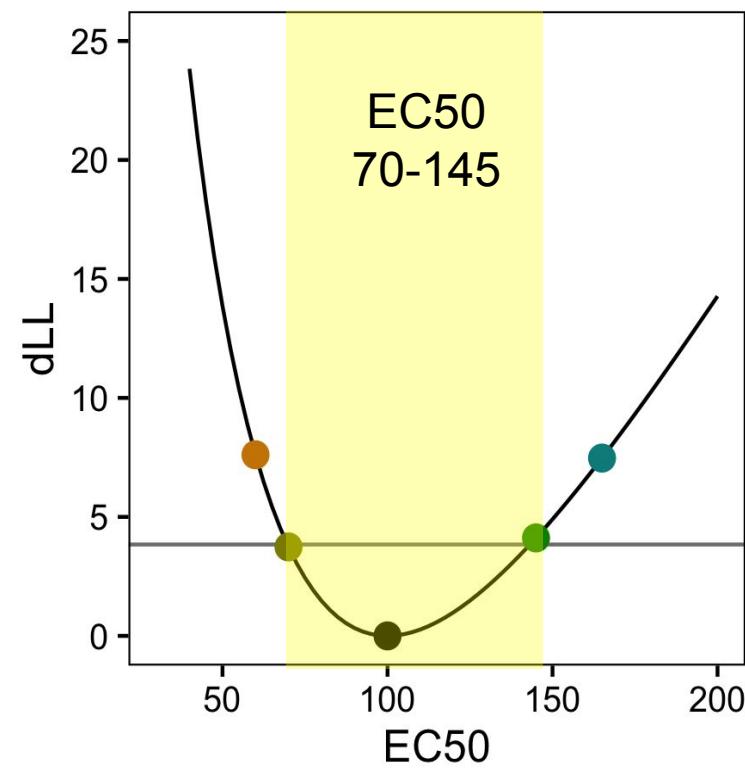
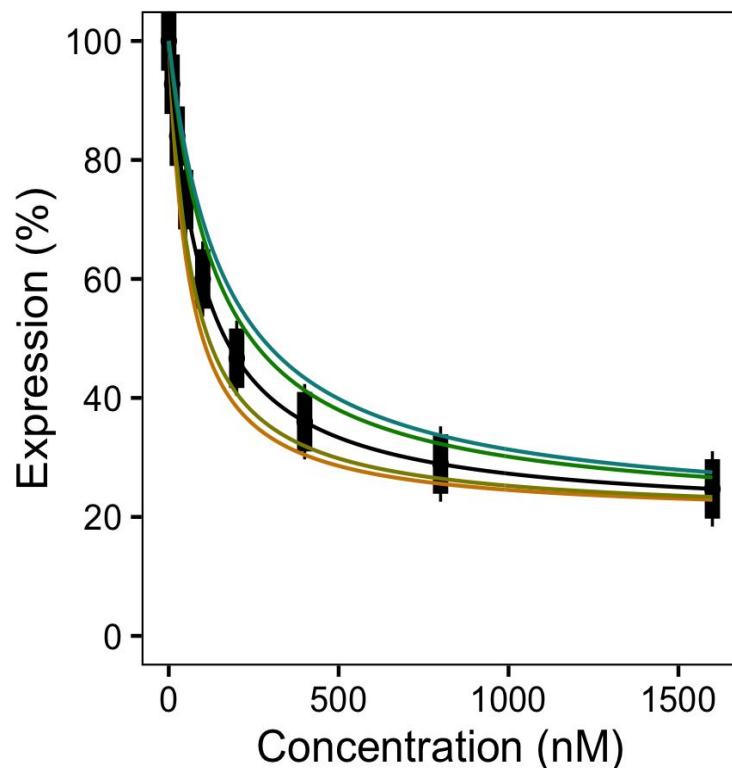
LLP Example: Emax model

- $E(C) = 100\% - \frac{E_{max} \cdot C}{EC50 + C}$
- There are two parameters:
 - Emax = maximum drug effect (80%)
 - EC50 = concentration for 50% effect (100 nM)
- Assume additive error
 - $\sigma = 5\%$
- Log likelihood ($-2LL$) $\sim \sum_i \frac{1}{\sqrt{2\sigma^2}} (y_i - E(C_i))^2$
- $dLL = -2LL - \min(-2LL)$



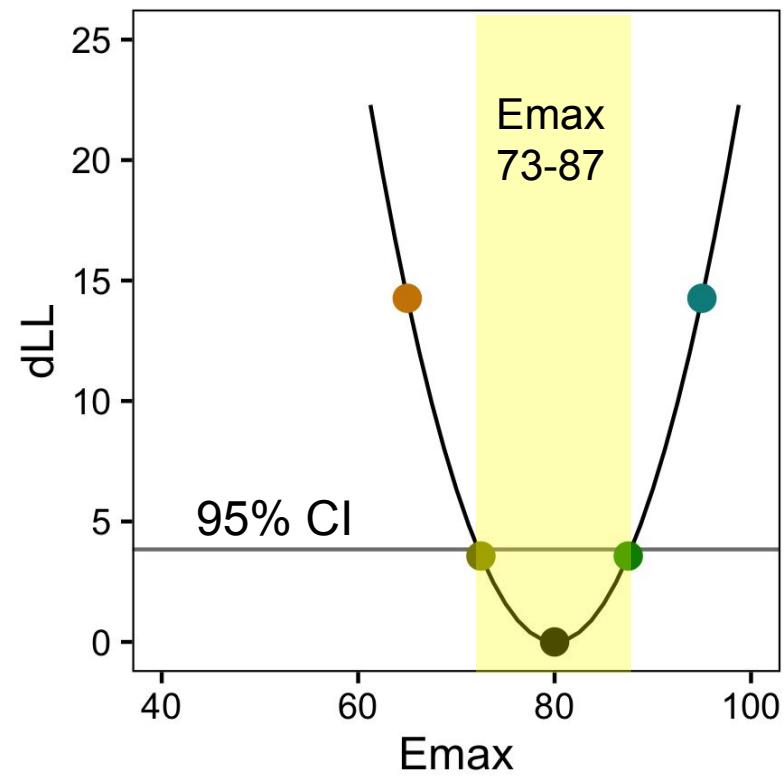
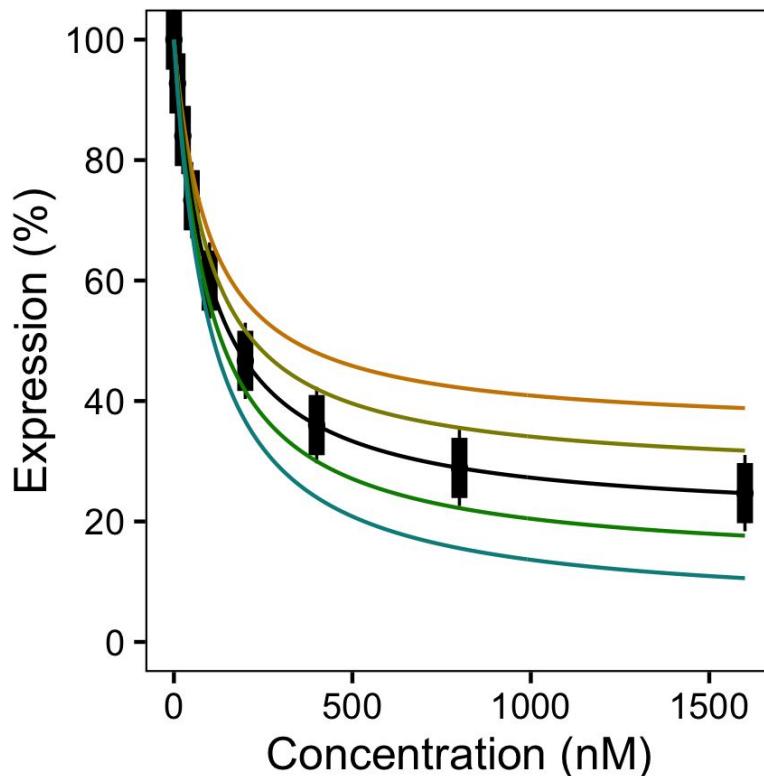
Calculation of standard confidence intervals for EC50

- Vary EC50 keeping Emax = 80%

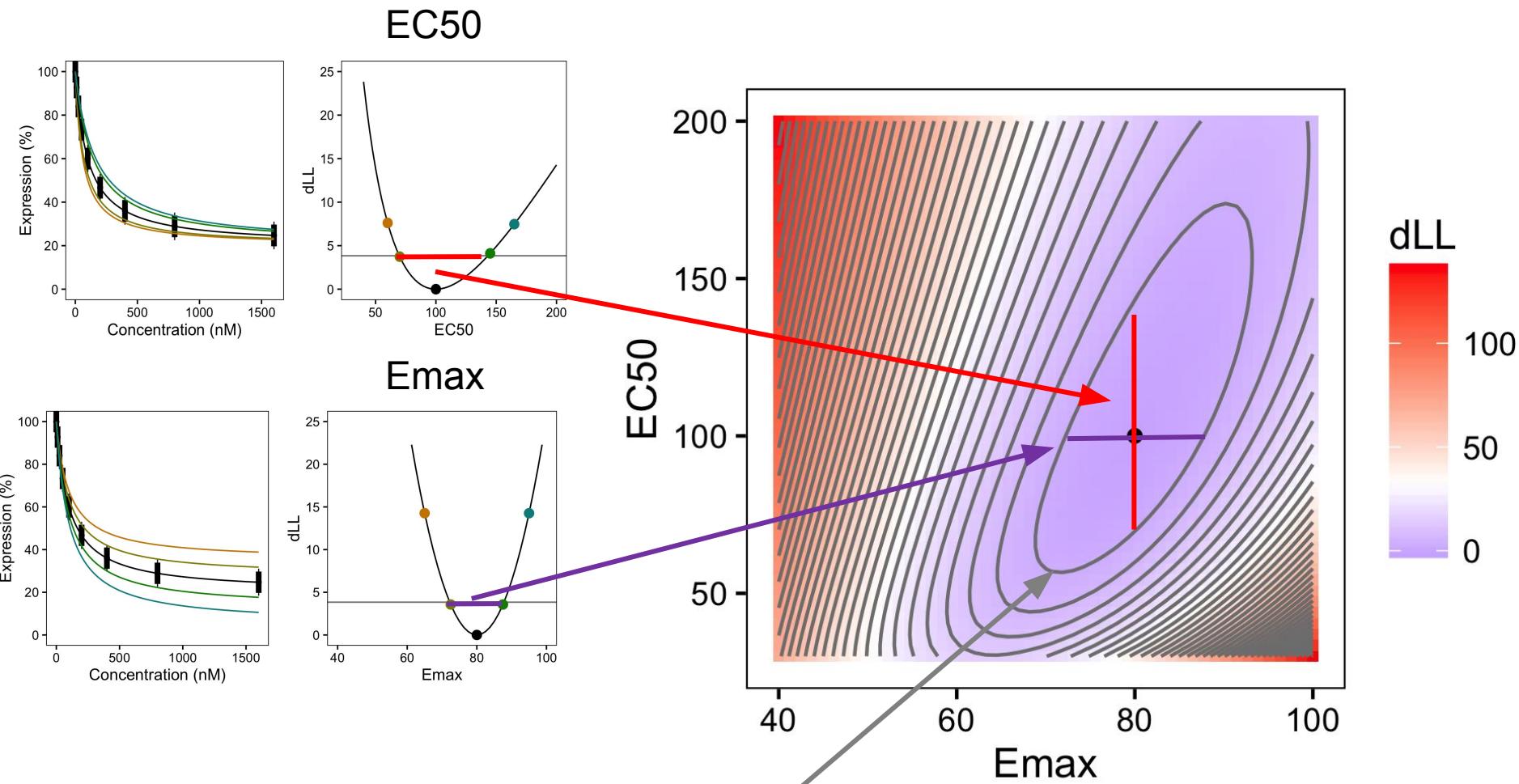


Calculation of standard confidence intervals for Emax

- Vary Emax keeping EC50 = 100 nM
- 95% confidence intervals can be estimated by region where $dLL < \chi^2(\text{deg. freedom } = 1, \alpha=.95) = 3.84$

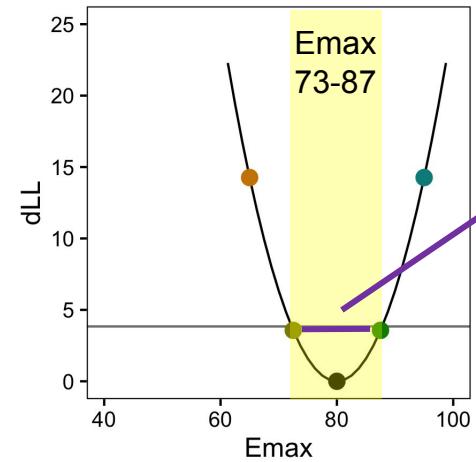
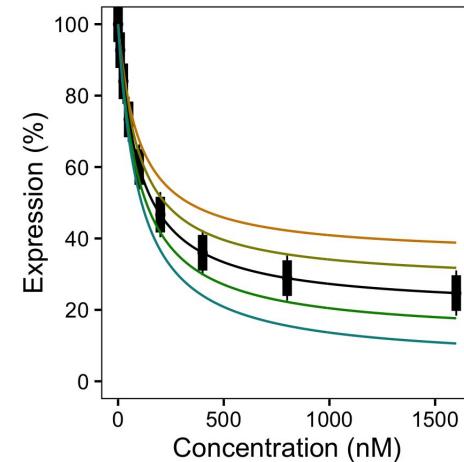
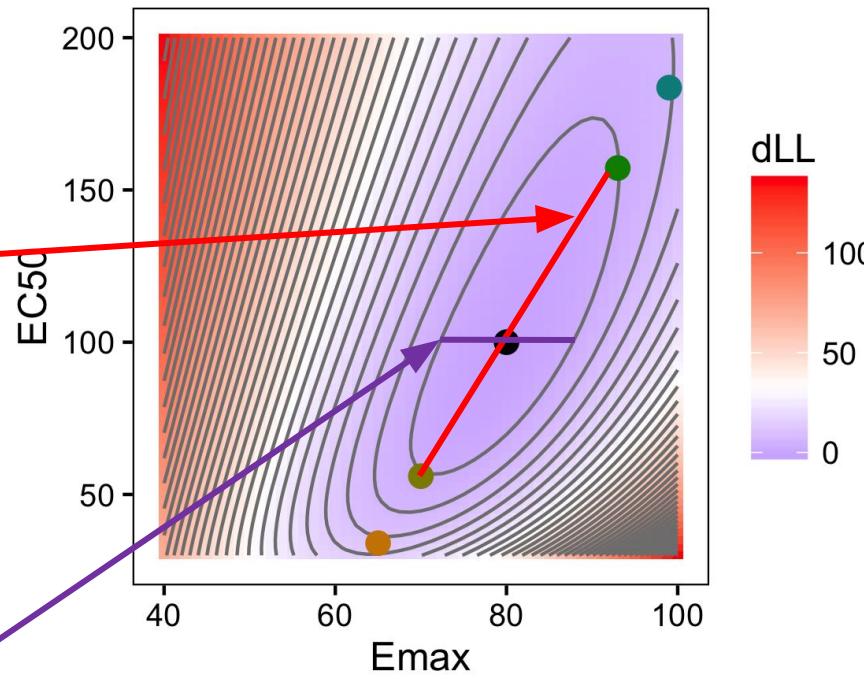
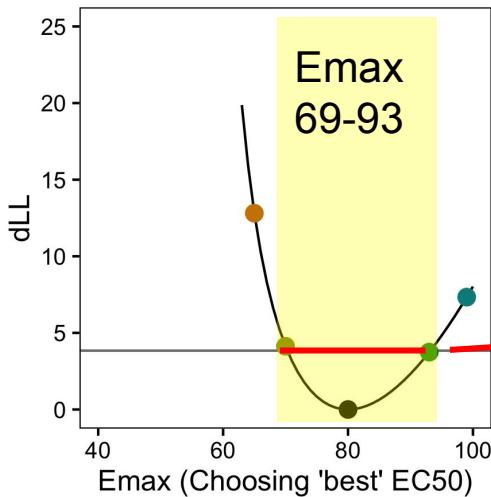
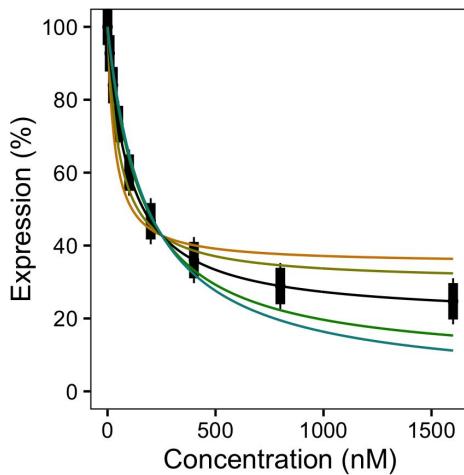


2d Likelihood space for Emax model



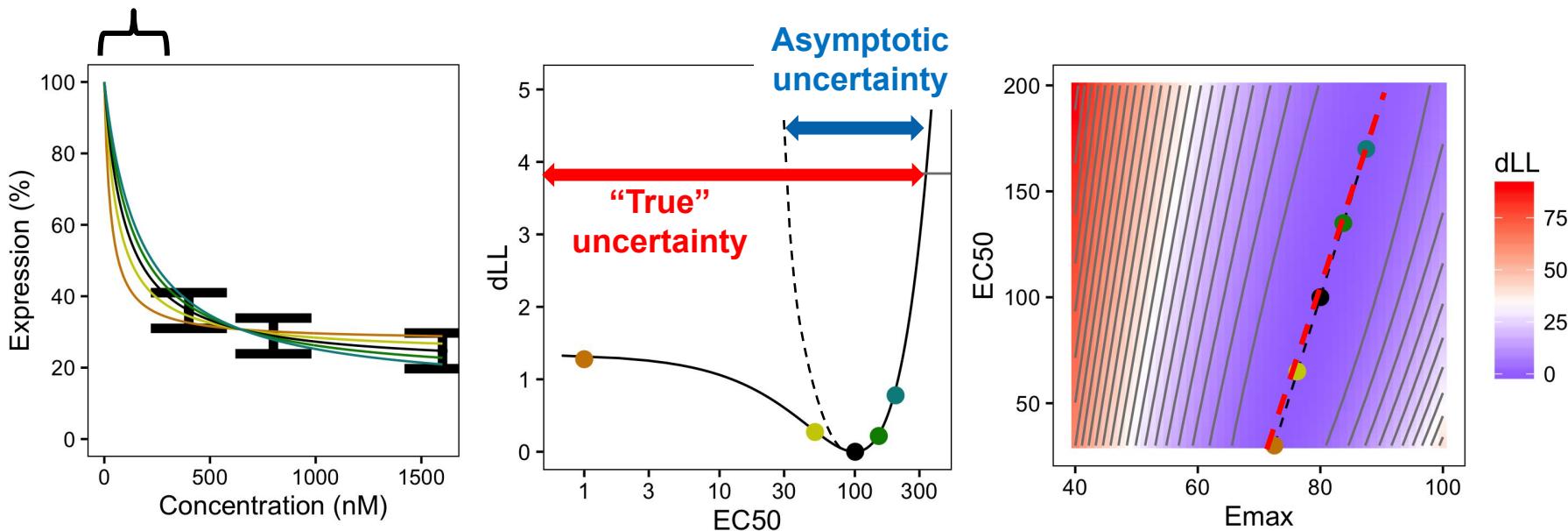
First contour is where $dLL = 3.84$
 $p = 0.05$ for χ^2 distribution with one degree of freedom

Likelihood profile in “most uncertain dimension” shows more uncertainty.



Given limited data, likelihood profiling shows EC50 is practically unidentifiable

Data below 400 nM
unavailable



A practically identifiable parameter is one where:

- Lower confidence limit > 0
- Upper confidence limit $< \infty$

Likelihood Profiles: A Method to Assess Parameter Precision

Goal: Evaluate precision of fit

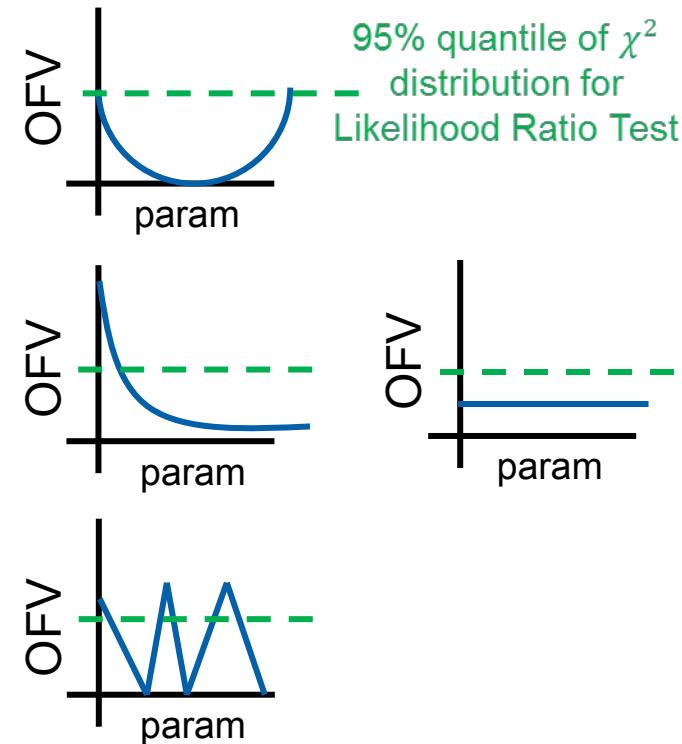
Procedure:

1. Perform a fit to find the n optimal parameters that describe data
2. Perturb and fix estimate of parameter i
 - Re-fit $n-1$ other parameters
 - Record OFV of new estimate

Identifiable

Non-
Identifiable

Poor
Convergence

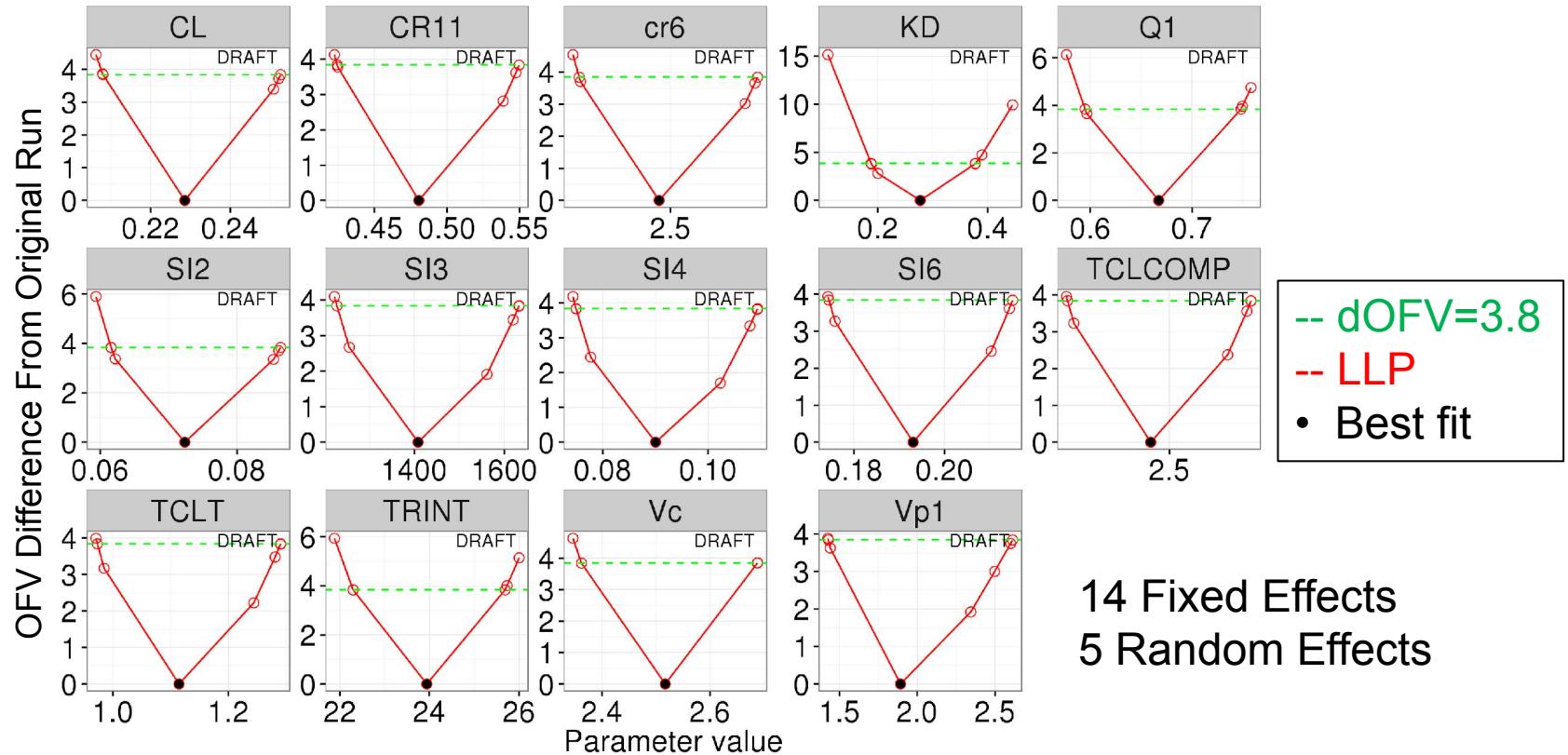


OFV: Objective Function Value

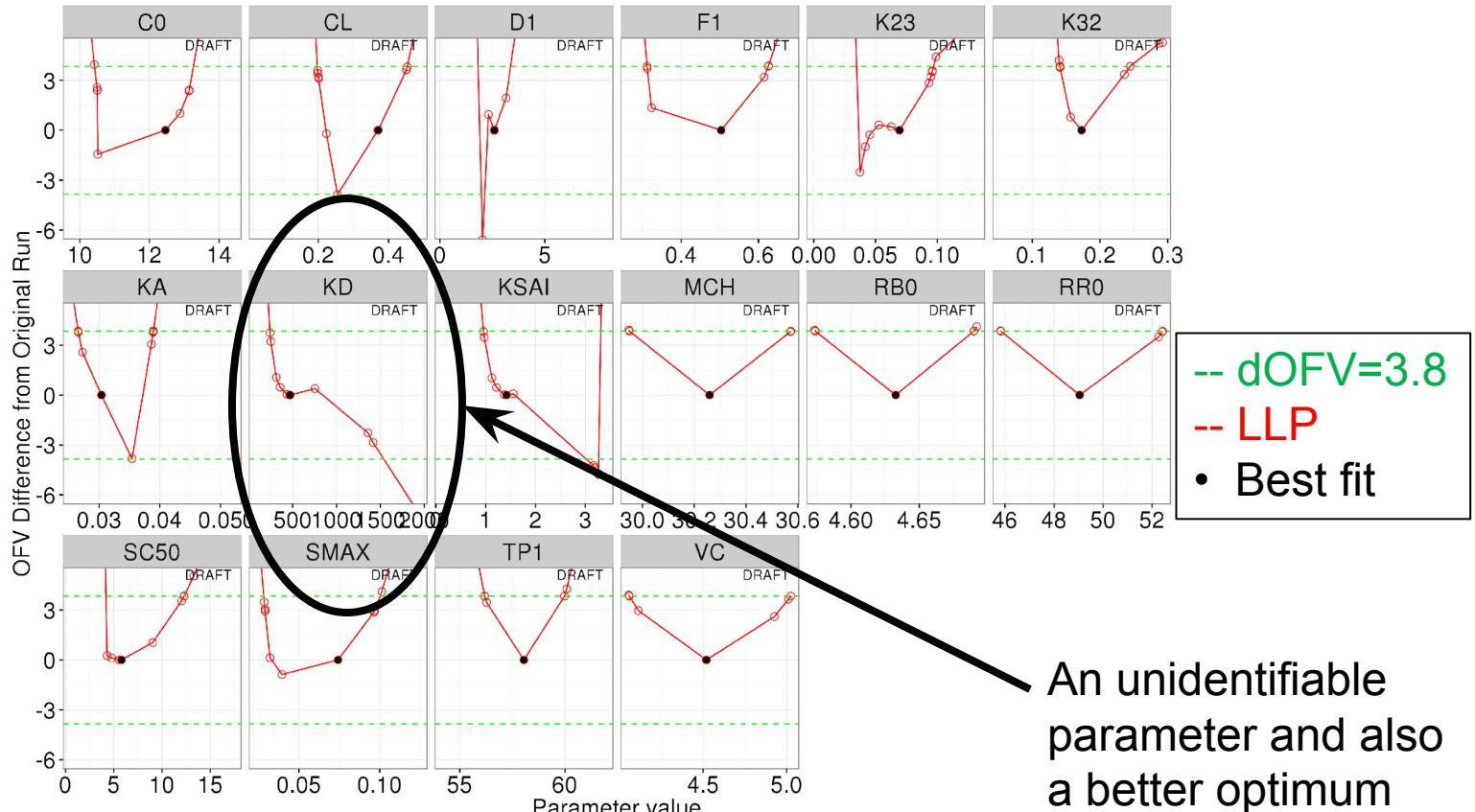
Method for generating the likelihood profile

- In Monolix 2017R1, there will be an “llp” function
- In Matlab, one can use the package Data2Dynamics.
- In Nonmem and PsN, you can use “llp” function
- PSN command: llp-4.2.0 file_name.mod
-clean=3 -thetas=1,2,3,4,5 -omegas=1

Example: Identifiable Model



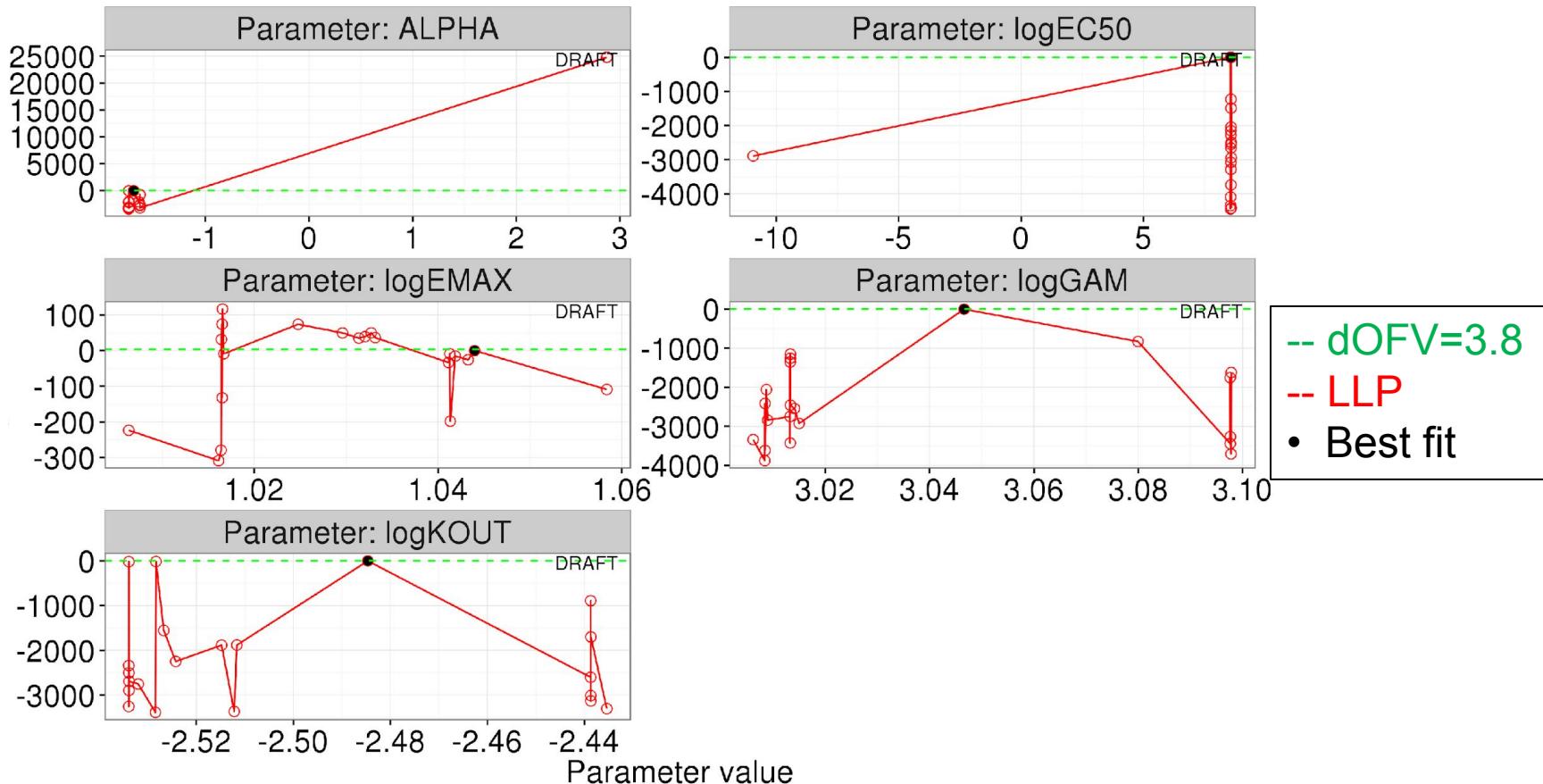
Example of model with unidentifiable parameters



An unidentifiable parameter and also a better optimum

16 Fixed Effects,
10 Random Effects

Example of model with poor convergence



NONMEM output had no errors or warnings

Minimization successful for original NONMEM estimate

Likelihood profiles can take a long time to compute in NONMEM

- We looked at 8 models
- Time to compute: 10 hours – 10 days (with 500 cores)
- There are other diagnostics for identifiability based on a single NONMEM fit.
 - Condition number of Covariance Matrix (Inverse of Fisher Info Matrix)
 - Maximum off-diagonal term of correlation of Covariance Matrix
 - Question: which Fisher Information Matrix to use? And do we trust it?

Fisher Information Matrix for assessing parameter uncertainty

- Fisher Information Matrix is the expected second moment of the log likelihood (f), given data (x) and parameters (θ)

$$- I(\theta) = E \left[\left(\frac{\partial}{\partial \theta} \log f(x; \theta) \right)^2 \right] = \int \left(\frac{\partial}{\partial \theta} \log f(x; \theta) \right)^2 f(x; \theta) d\theta$$

- Covariance matrix is the inverse of the Fisher Information Matrix



Nonmem estimates covariance matrix in three ways

1. S Matrix: outer product of gradients

- $S^{-1} = COV^{-1} = I(\theta) = \int \left(\frac{\partial}{\partial \theta} LL(x, \theta) \right)^2 f(x; \theta) d\theta$
- $\approx \left(\frac{\partial}{\partial \theta} LL(x, \hat{\theta}) \right)^2$
- matrix notation $\approx \sum_i (\nabla LL)(\nabla LL)^T$, i = patient id

If there are fewer patients than parameters, S and Sandwich will be singular. Thus R should be used.

2. R Matrix: Hessian

- $R^{-1} = COV^{-1} = I(\theta)^{-1} = \int \left(\frac{\partial}{\partial \theta} LL(x, \theta) \right)^2 f(x; \theta) d\theta$
- Under certain regularity conditions, we can integrate by parts
- $= \int \frac{\partial^2}{\partial \theta^2} LL(x, \theta) f(x; \theta) d\theta$
- $\approx \frac{\partial^2}{\partial \theta^2} LL(x, \hat{\theta})$

3. Sandwich Matrix: $R^{-1}SR^{-1}$ = Nonmem Default

Choose matrix by: \$COVARIENCE MATRIX=R (or MATRIX=S).
Without this line, Sandwich matrix is the default.

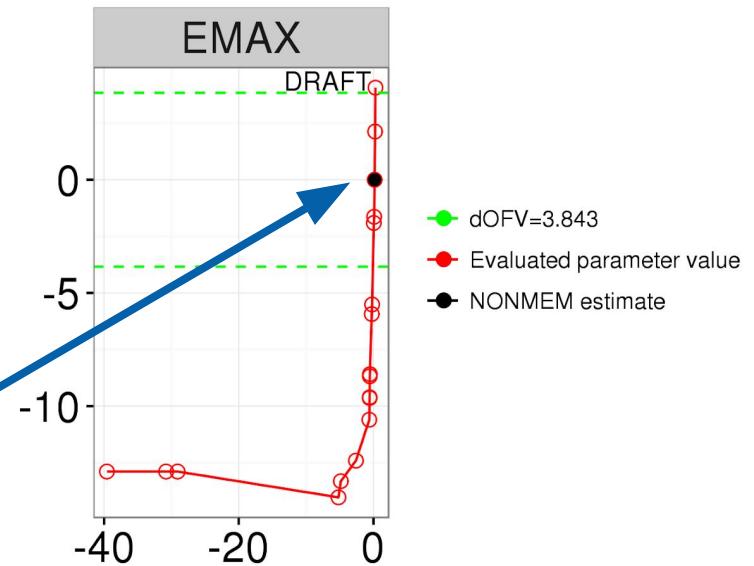
NONMEM can hide problems in the covariance step

OPTIMIZATION WAS COMPLETED

```
Number of Negative Eigenvalues in Matrix= 1  
Most negative value= -182.568507686780  
Most positive value= 7738.92745498663  
Forcing positive definiteness  
Root mean square deviation of matrix from original= 4.308439995024584E-002
```

“Forcing Positive Definiteness”
In Hessian matrix

Original run not a minimum

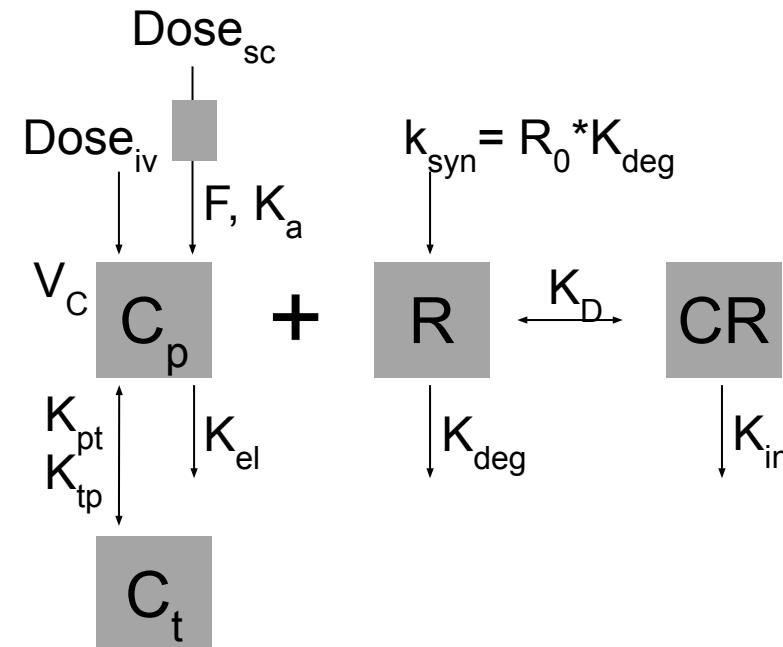


Correlation > 0.98 suggests unidentifiability or poor convergence

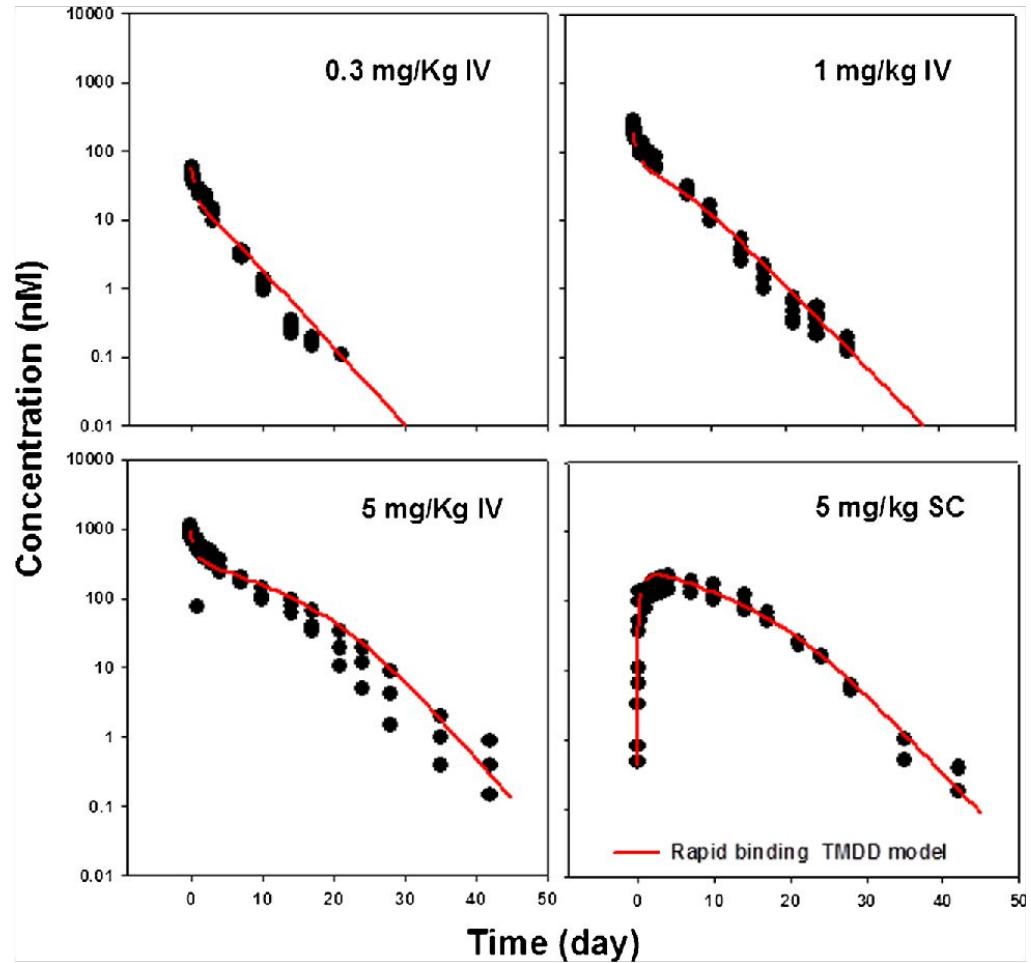
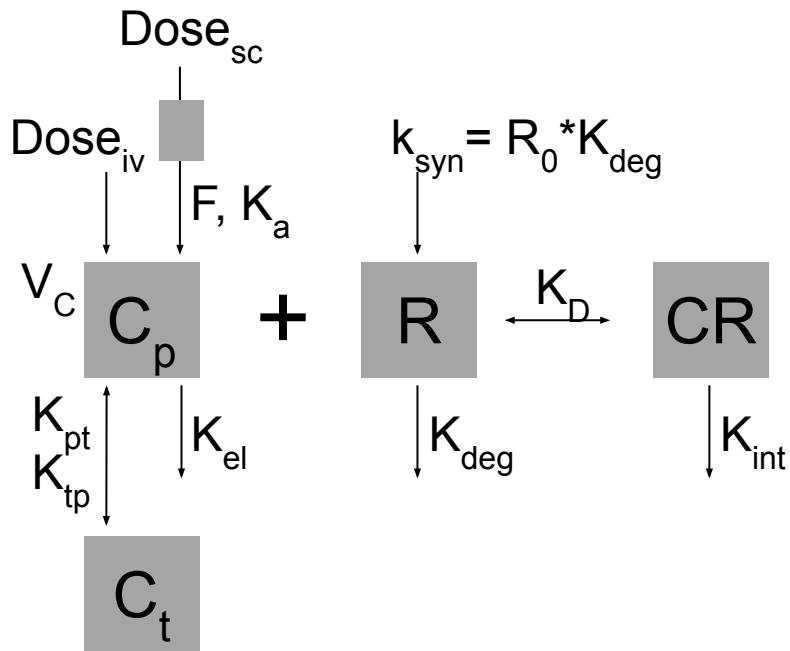
Model	Condition Number	Max Correlation	Model Type
TMDD #2	6320	-0.953	Identifiable
PK: 2 compartment	843.7	0.935	Identifiable
TMDD	172	0.946	Identifiable
Indirect Response v1	191	0.946	Identifiable
Indirect Response v2	712	0.93	Poor Convergence
Mechanistic Model #2	1000000	0.998	Poor Convergence
Custom PKPD	872	0.982	Unidentifiable
Mechanistic Model #1	20,000	0.987	Unidentifiable

Only 8 examples: more investigation is needed

Target Mediated Drug Disposition (TMDD) Model



TMDD model for antibody drug¹



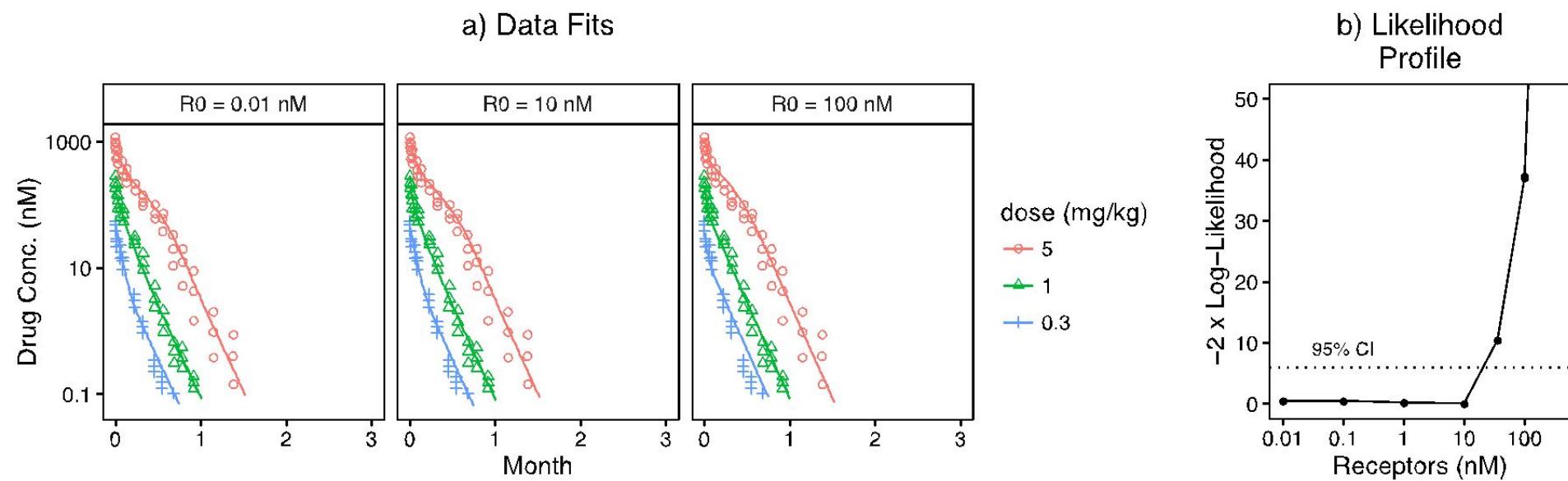
1. Singh et al., *AAPS Journal*, 17, 389 (2015), Monolix

TMDD model fit to mAb data shows tight confidence intervals¹

Param.	Unit	Value (Rel. Standard Error %)
R_0	nM	1.17 (1)
K_{deg}	day ⁻¹	15.5 (8)
K_D	nM	9.37 (4)
K_{int}	day ⁻¹	5.07 (3)
K_{el}	day ⁻¹	0.14 (11)
K_{pt}	day ⁻¹	1.25 (10)
K_{tp}	day ⁻¹	0.941 (1)
V_c	L	0.132 (4)
K_a	day ⁻¹	0.657 (7)
F	—	0.802 (5)

1. Singh et al., *AAPS Journal*, 17, 389 (2015), Fits in Monolix

Likelihood profile shows that receptor density (R_0) can change by 1000x and be consistent with data



Param.	Unit	Value (Rel. Standard Error %)
R_0	nM	1.17 (1)

Large underestimate
of true uncertainty

Consequence of over-confident estimate of baseline receptor density (R0)

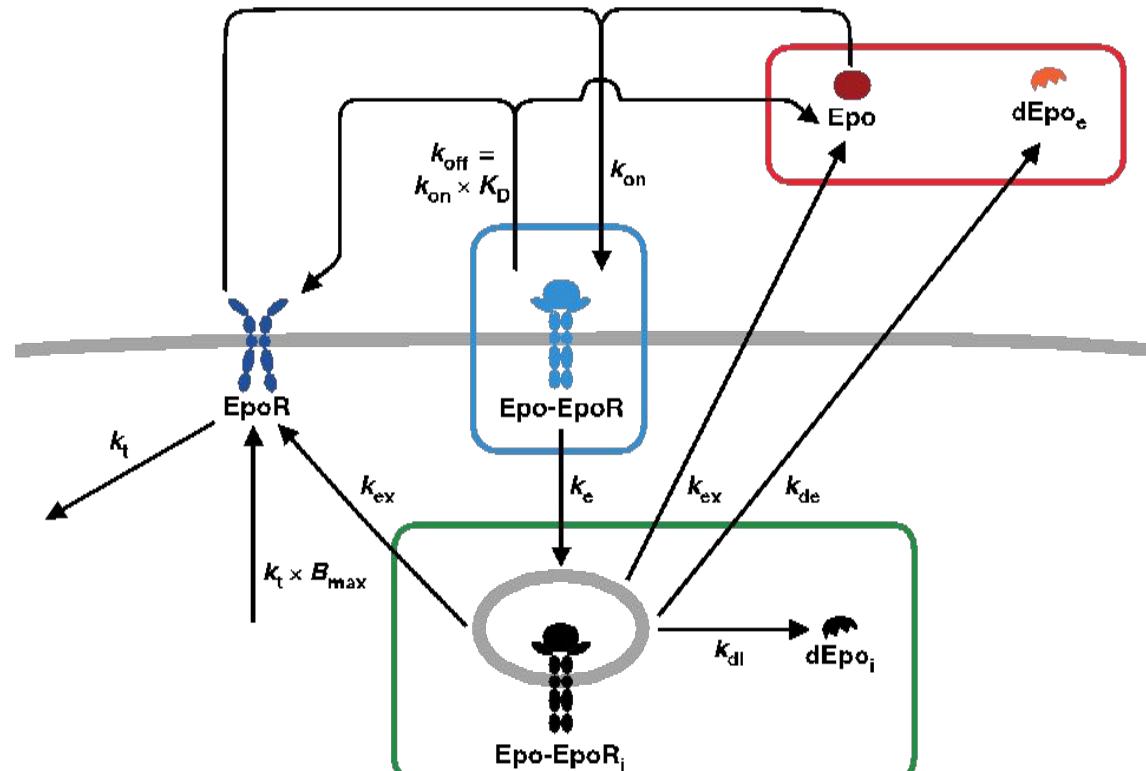
- Incorrect physiological interpretation of model (e.g. location or number of target receptors)
- Error when scaling nonlinear PK from cynomolgus monkeys to man.



**Likelihood profiling can
detect non-identifiability**

**and guide experiments to
resolve it**

Modeling Epo and EpoR interactions



10 parameters

	Synthesis	Binding	Endocyt	Exocyt	Degradation
EpoR: dR/dt	$=k_t \cdot B_{max}$	$-k_{on}L \cdot R + k_{off}(RL)$		$+k_{ex}(RL_i) - k_t R$	
Epo: dL/dt	$=$	$-k_{on}L \cdot R + k_{off}(RL)$		$+k_{ex}(RL_i)$	
Epo-EpoR: $d(RL)/dt =$		$k_{on}L \cdot R - k_{off}(RL) - k_e(RL)$			
Epo-EpoR-int: $d(RL_i)/dt =$				$+k_e(RL) - k_{ex}(RL_i) - (k_{de} + k_{di})(RL_i)$	

Reasons for non-identifiability

- Practical Identifiability
 - Measurement uncertainty
 - Time scales & limited sampling
 - Saturation
- Structural Identifiability
 - Non-observed states
 - Not directly observed states
 - Scaling factors

A simple example...

$$\begin{cases} \dot{x}_1(t) = -\theta_1 x_1 \\ \dot{x}_2(t) = +\theta_1 x_1 \end{cases}$$

$$x_1(0) = \theta_2 \text{ and } x_2(0) = 0$$

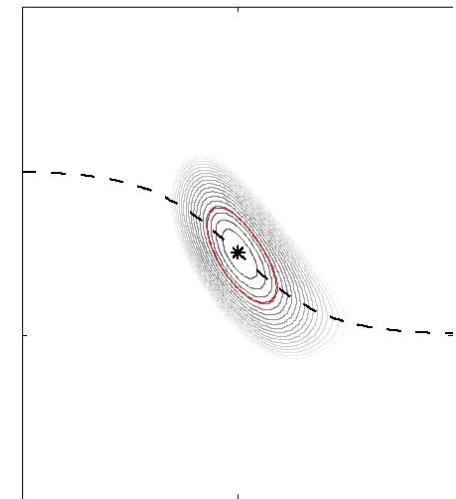
$$y_1(t) = \theta_3 x_2(t)$$

$$y_1(t) = \boxed{\theta_3 \theta_2} \exp(-\theta_1 t) \cdot (\exp(\theta_1 t) - 1)$$

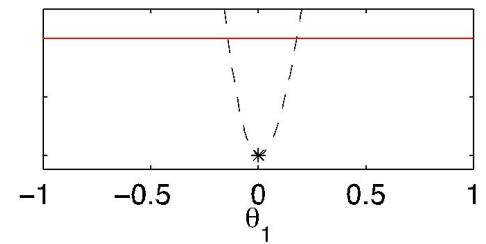
Structurally non-identifiable

Likelihood profiling can be used to detect both practical and structural non-identifiability

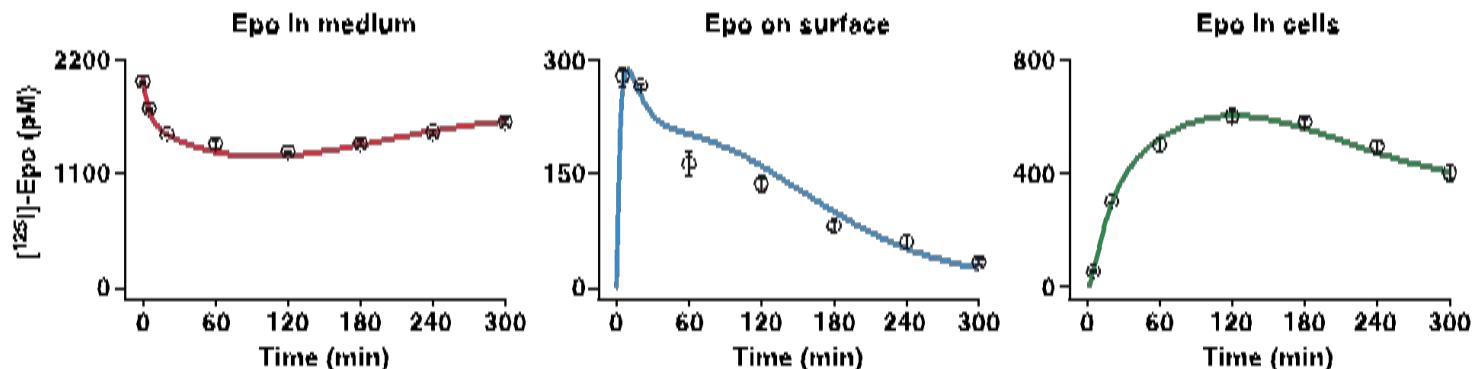
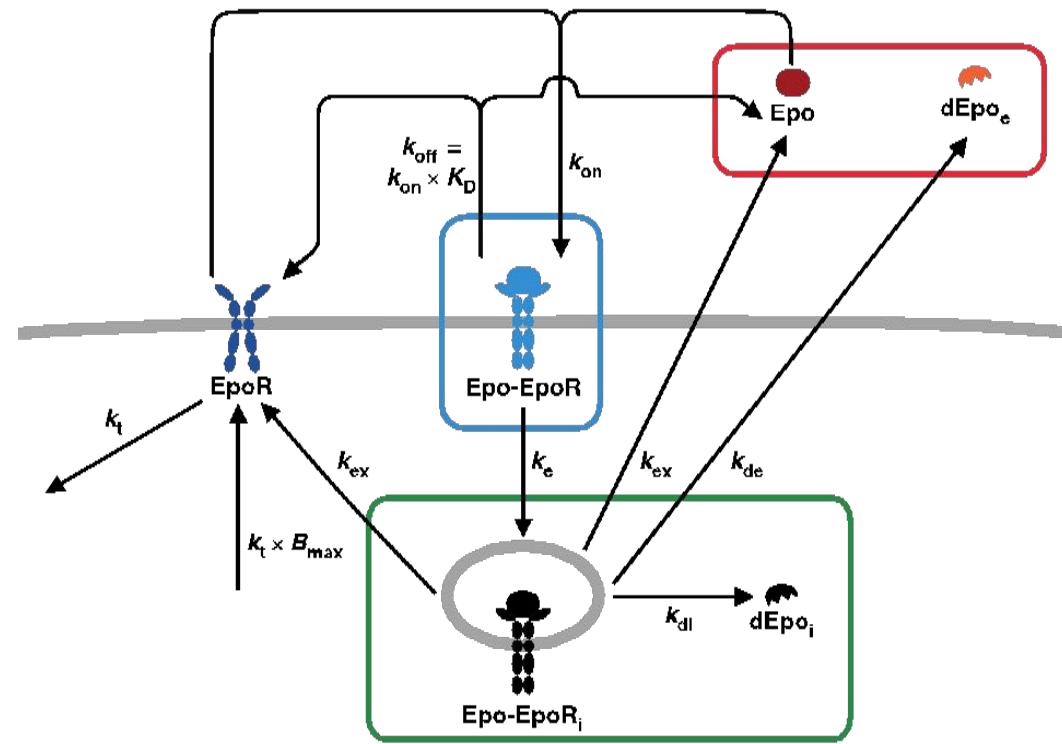
(e): parameters identifiable



(f): profile likelihood of (e)

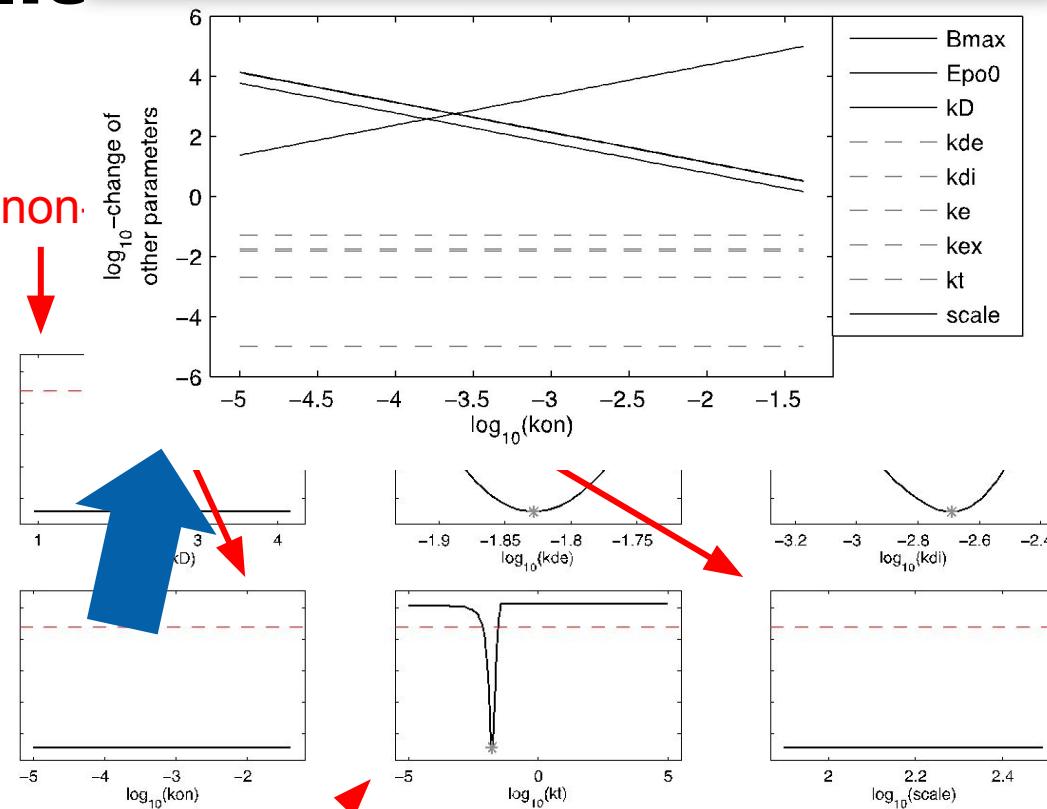
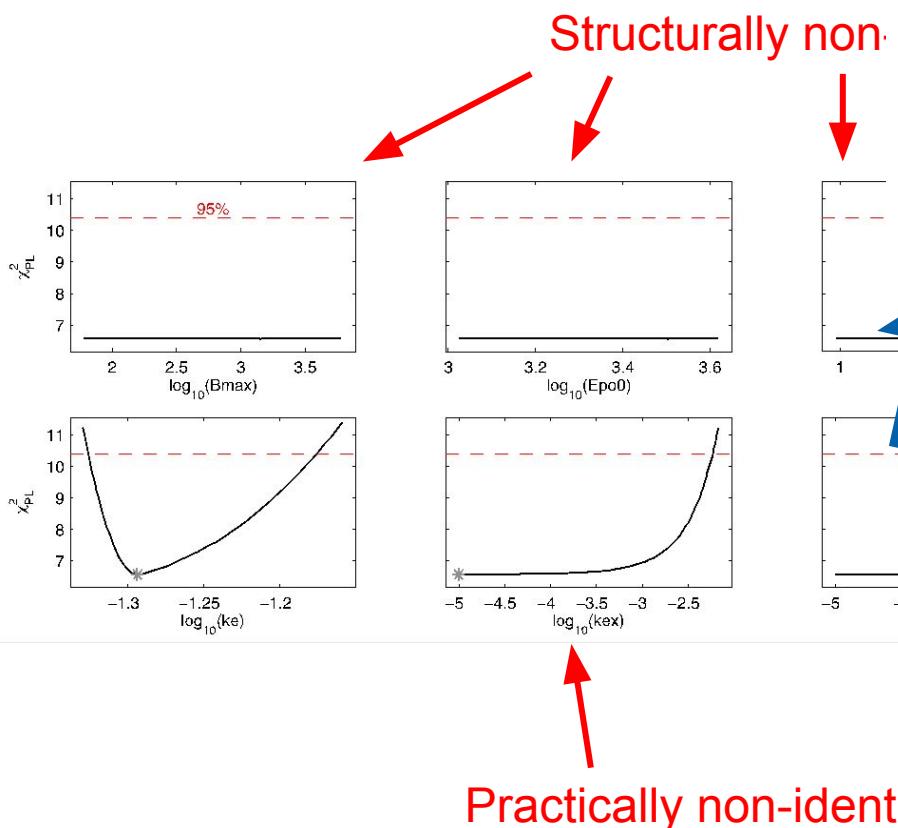


Data for Epo-EpoR model



Likelihood profiling is useful when applying models

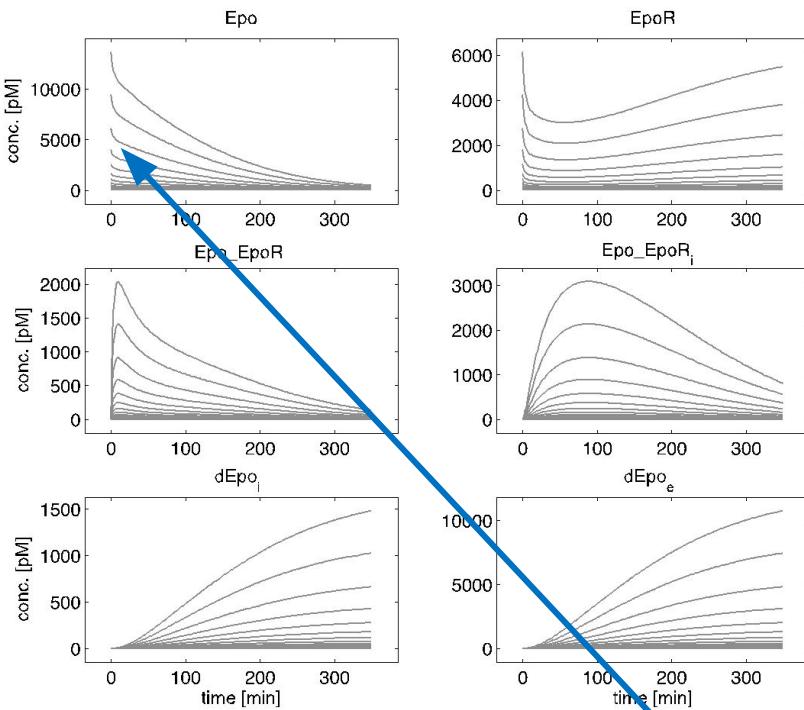
Functional relation between structurally non-identifiable parameters



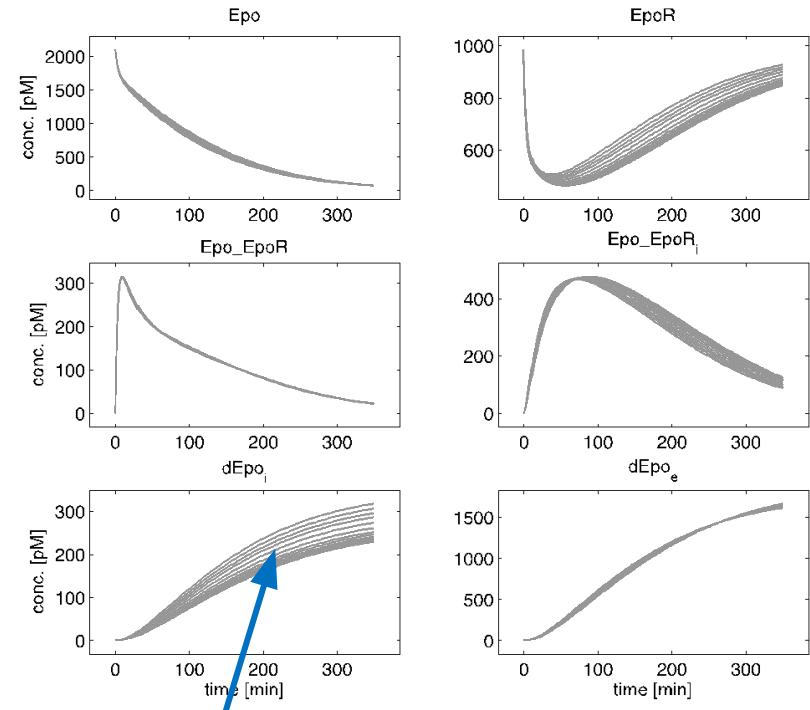
Practically non-identifiable

Parameter uncertainty translates to model dynamics - and can help to propose new experiments

kon: structurally non-identifiable



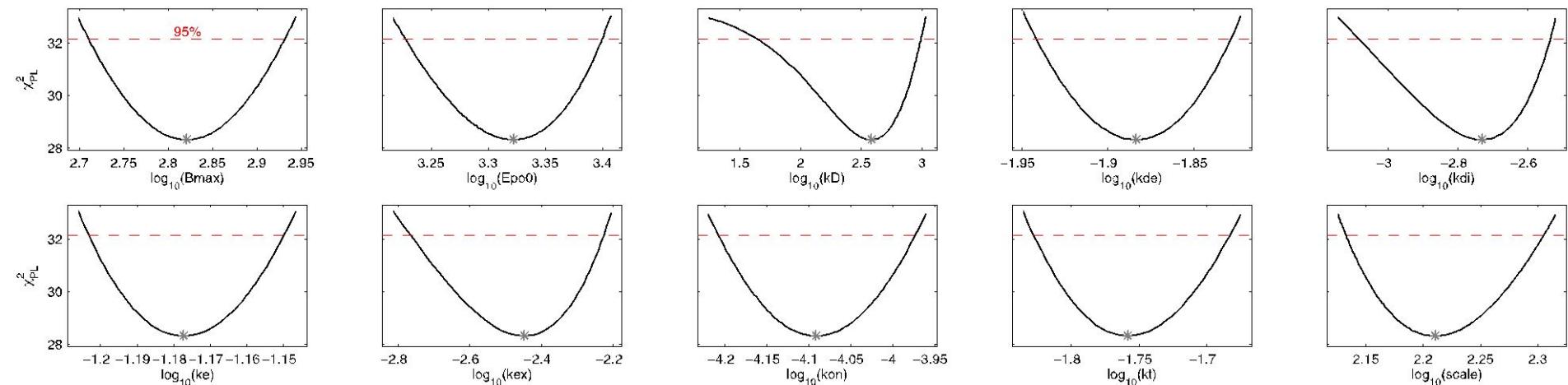
kex: practically non-identifiable



- Nondimensionalization

New measurements here will resolve uncertainty !

Given appropriate data, uncertainty in model parameters / predictions can be removed



All parameter identifiable ✓

Summary

Summary

- **Likelihood Waterfalls** are helpful for
 - Understanding if there are multiple local minima
 - Assessing whether the optimizer is converging to the local minima
- **Likelihood Profiles** are helpful in assessing parameter identifiability
 - Even if Nonmem (or any other optimizer) converges with no warnings, uncertainty may still be considerably underestimated.
- Maximum off-diagonal correlation of covariance matrix > 0.98 seems to be sign of convergence or identifiability issues. More examples needed.

Acknowledgements

- Geraldine Ayral (Lixoft)
- Alison Margolskee (Novartis)
- Frances Mentre (INSERM + U Paris)
- Bert Peletier (U. Leiden)
- Bernhard Steiert (Freiburg)
- Eduardo Sontag (Rutgers)