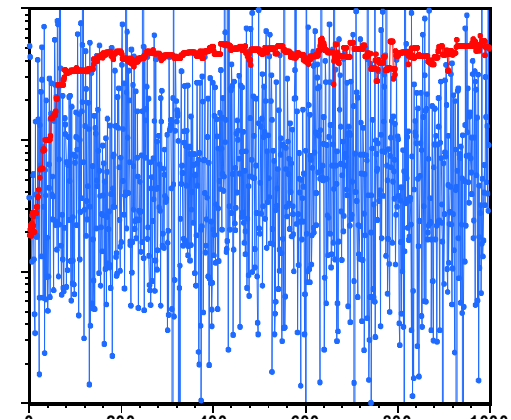
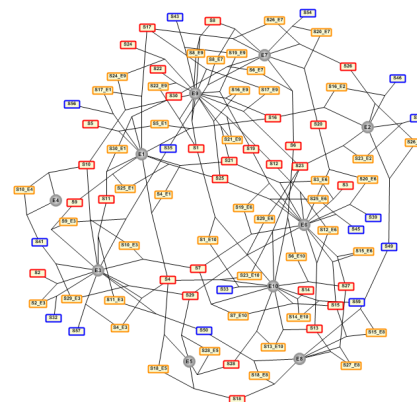


MCSim: un logiciel de simulation et inférence statistique

Frédéric Yves Bois

INERIS-UTC



MCSim est au départ un logiciel de simulation

SBML

Model
Definition
File

Simulation
Definition
File

Mod

C
File

*C
compiler*

Executable

Fichier de
sortie



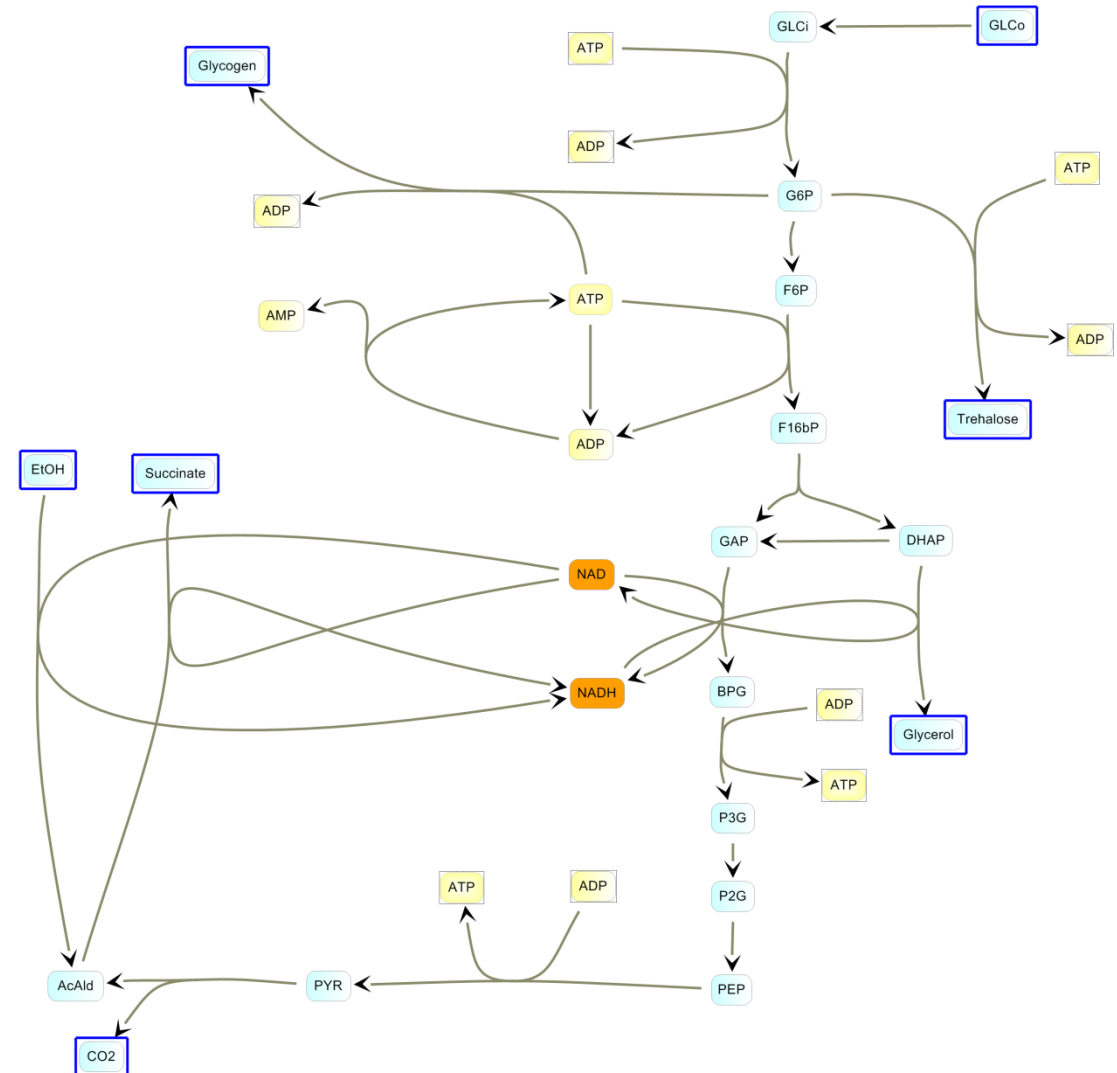
MCSim permet de faire

- Des simulations simples ($y=f(t, \)$) ou $y'=f(y,t, \)$ par intégration numérique.
- Des simulations Monte Carlo simples
- Des simulations en série pour une grille imposée
- Des simulations MCMC (Metropolis Hasting) y compris pour des modèles hiérarchiques.
- Identification de design expérimentaux "optimaux" sur la base de simulations prédictives.

Stats: Bayesian calibration of SBML models

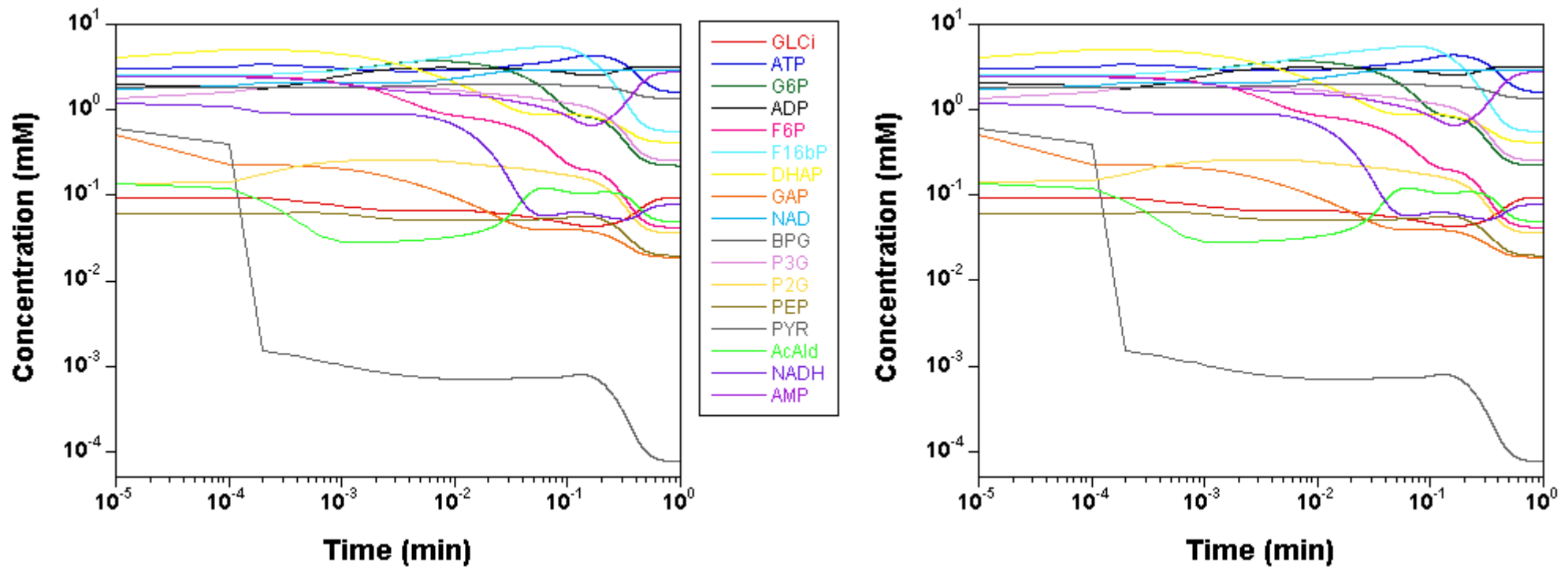
Example of the yeast glycolysis model (Pritchard and Kell, 2002)

- 17 state variables
- 95 parameters
- 19 reactions

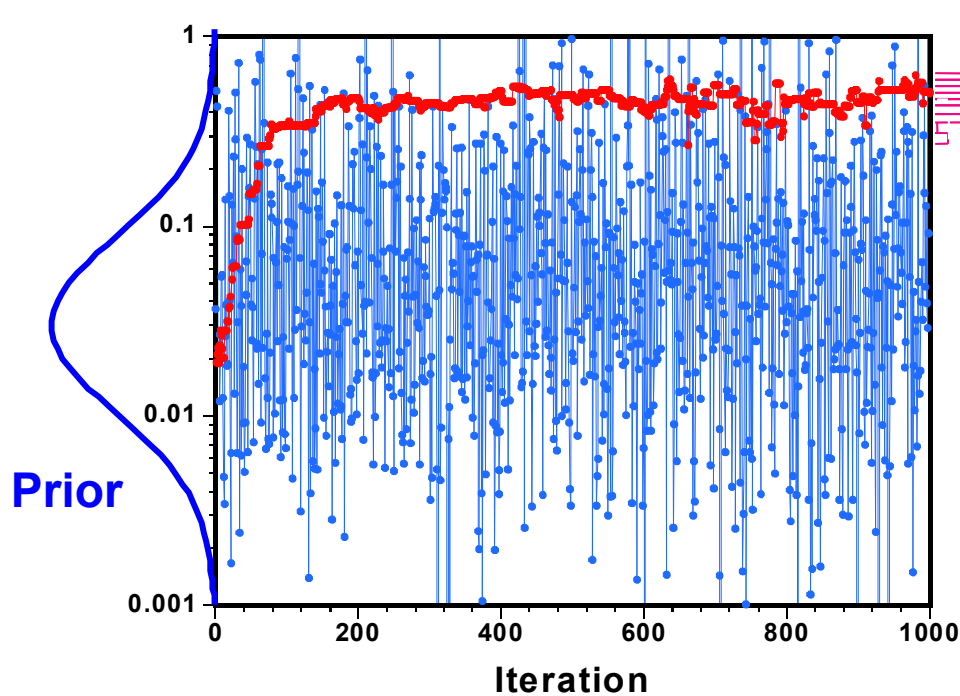


Time-course simulations

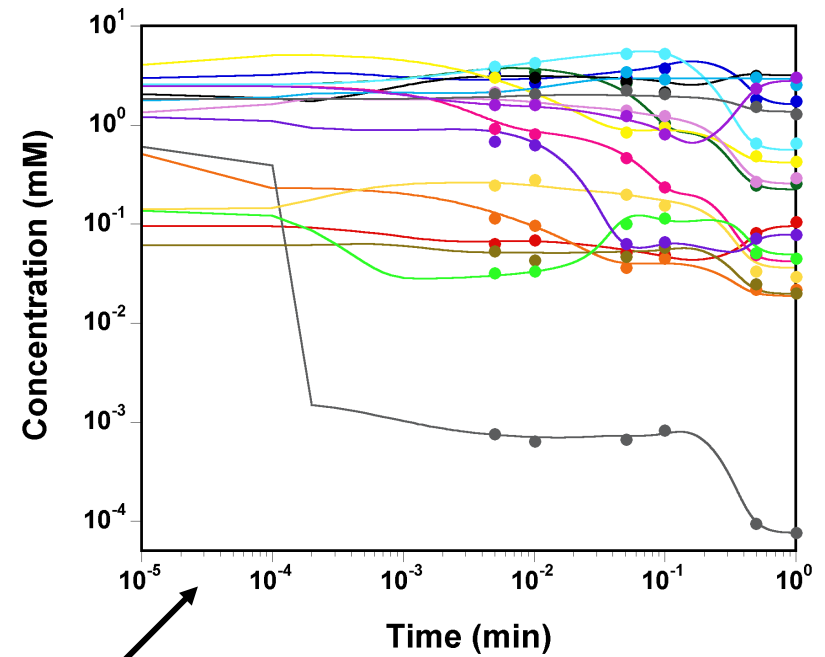
Exhibits complex and nonlinear behaviour



Bayesian calibration is done via MCMC

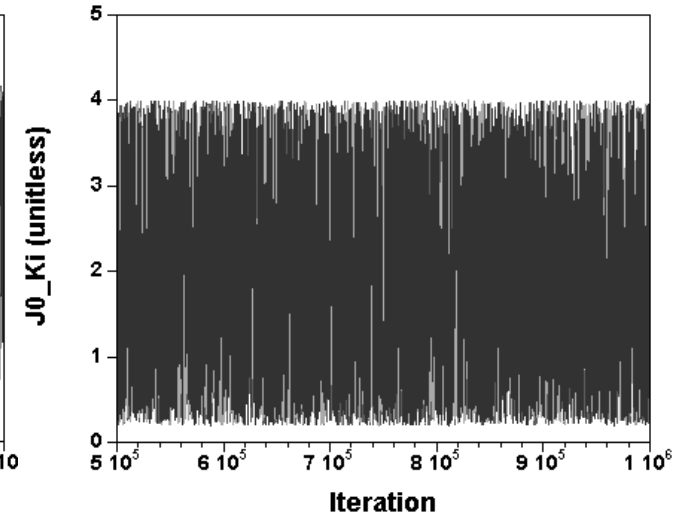
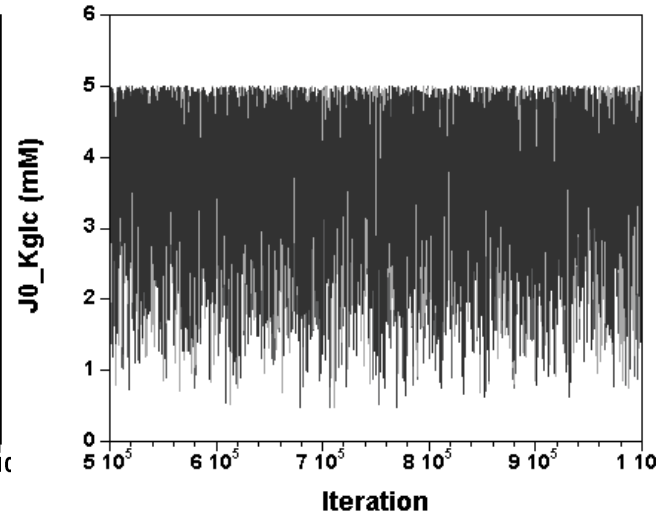
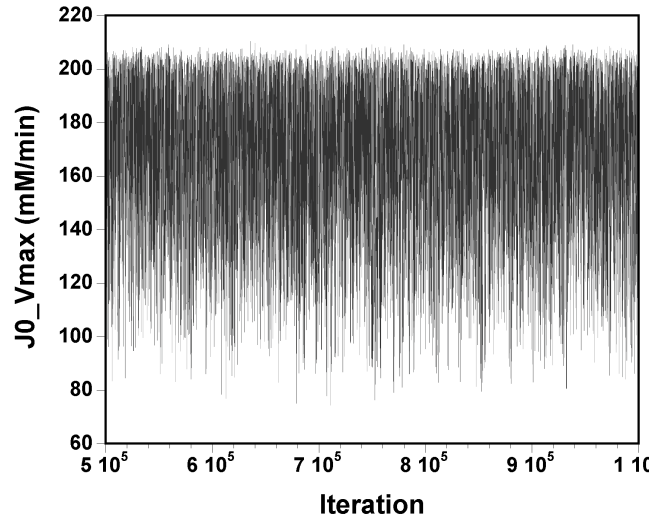


$$\text{Posterior} = \text{Prior} \times \text{Likelihood}$$



The data used

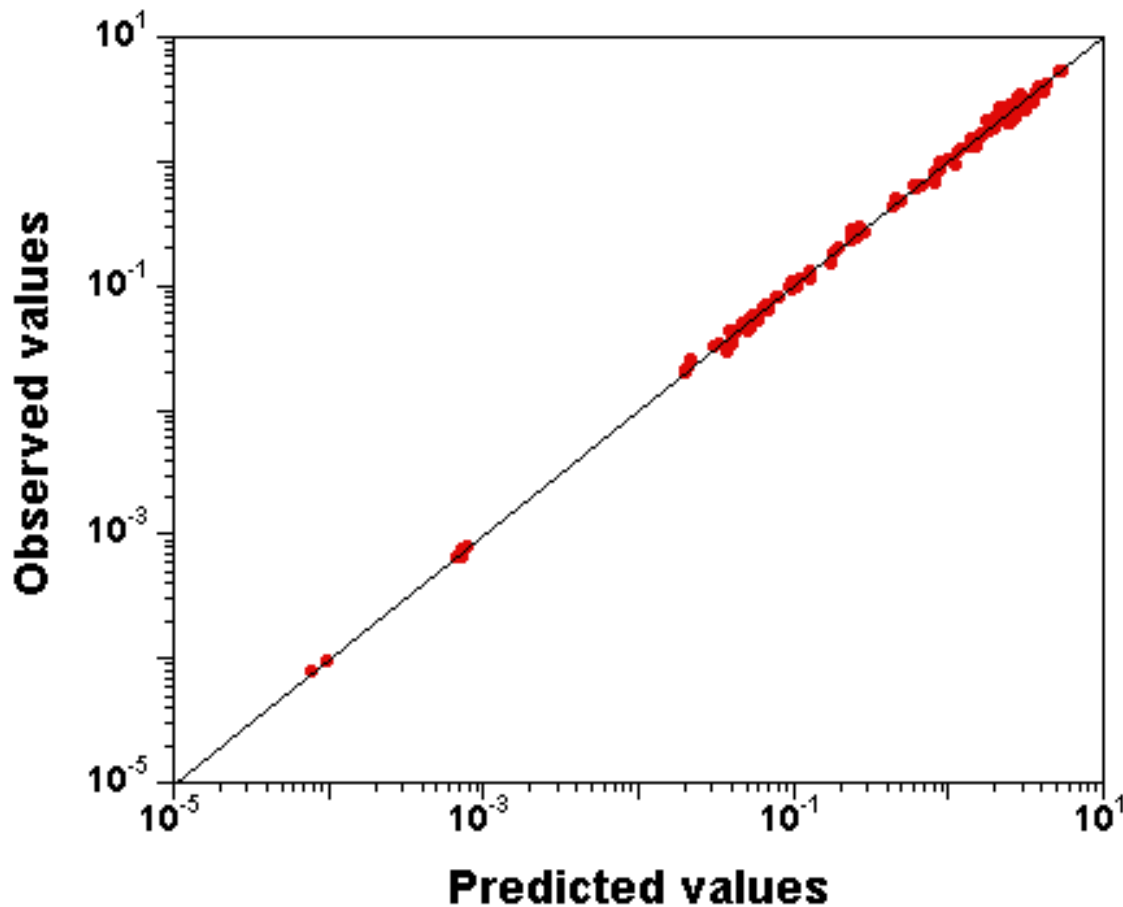
Trajectories at convergence



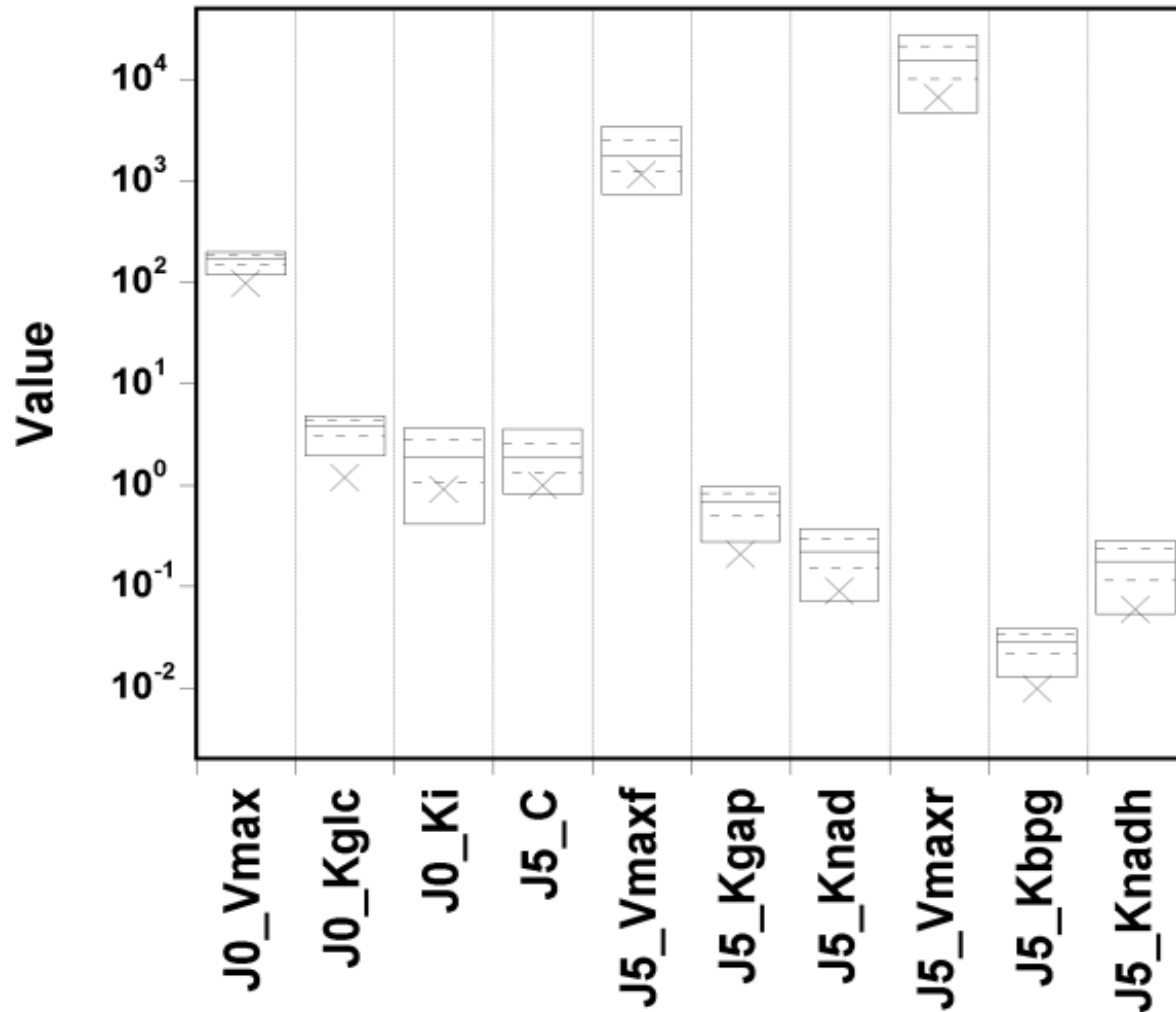
Parameters nicely converge to a stable joint distribution (here: the last 500,000 iterations of one million, 80 minutes on a i686 computer)

Fit to the data

Excellent: Predictions are very close to the observations
(10 parameters were sampled simultaneously)



Posterior distribution summaries



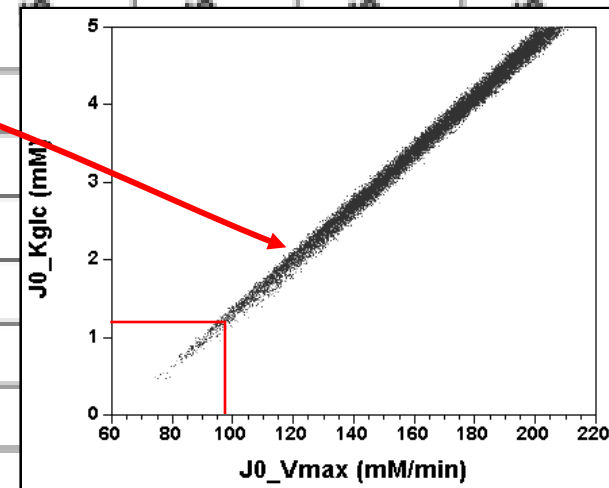
... **Bad!**

Remember that we have the right model and quite "clean" data.

What went wrong?

Posterior distributions' correlation matrix

	J0_Vmax_1	J0_Kglc_1	J0_Ki_1	J5_C_7	J5_Vmaxf_7	J5_Kgap_7	J5_Knad_7	J5_Vmaxr_7	J5_Kbpg_7	J5_Knadh_7
J0_Vmax_1	1									
J0_Kglc_1	0.997	1								
J0_Ki_1	0.32	0.28	1							
J5_C_7				1						
J5_Vmaxf_7				-0.72	1					
J5_Kgap_7				0.28	0.37	1				
J5_Knad_7				-0.24	0.24		1			
J5_Vmaxr_7				-0.24	0.20		-0.19	1		
J5_Kbpg_7					-0.10		0.17	0.20	1	
J5_Knadh_7				0.20	-0.12	0.10	0.45	0.35		1



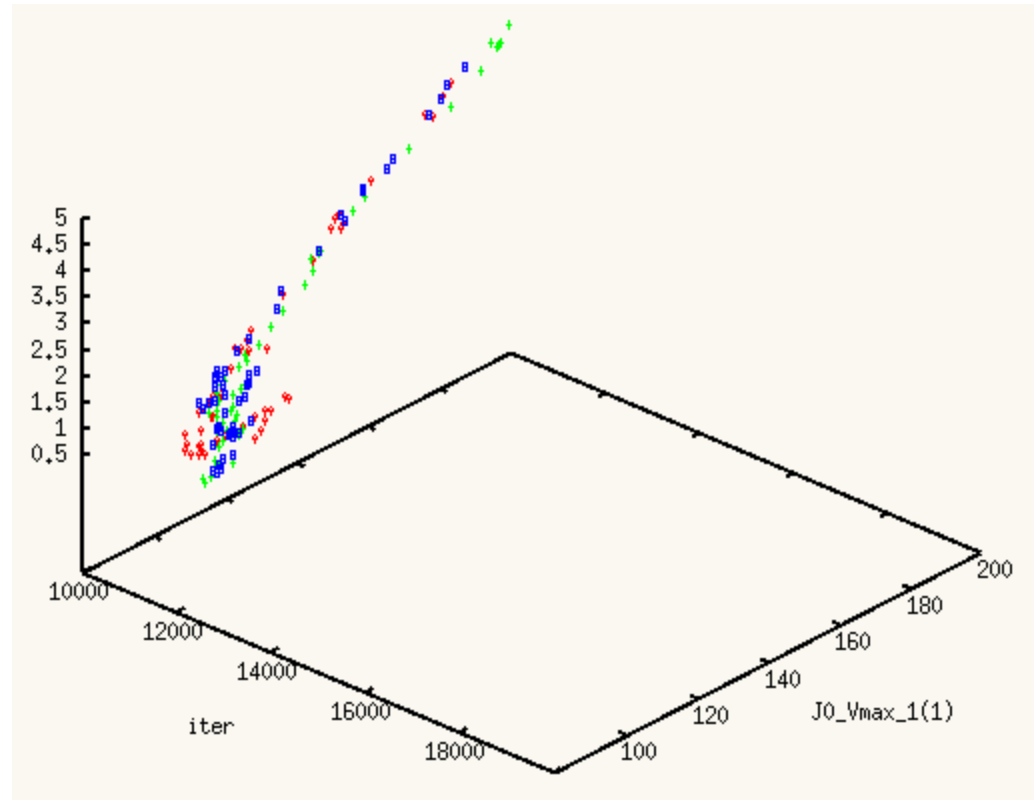
Design optimization

Duplicate of original design

iter	GLCo = 0.5						GLCo = 1						GLCo = 2						GLCo = 4						GLCo = 8						Chosen	Variance	SD
	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
0																															0	0,0598	0,245
1																															4	0,0419	0,205
2																														5	0,0348	0,186	
3																														2	0,0306	0,175	
4																														3	0,0276	0,166	
5																														10	0,0252	0,159	
6																														11	0,0235	0,153	
7																														9	0,0220	0,148	
8																														8	0,0207	0,144	
9																														17	0,0203	0,142	
10																														7	0,0200	0,141	
11																														16	0,0196	0,140	
12																														18	0,0191	0,138	
13																														25	0,0188	0,137	
14																														13	0,0186	0,137	
15																														1	0,0184	0,136	
16																														15	0,0183	0,135	
17																														14	0,0183	0,135	
18																														12	0,0186	0,136	
19																														20	0,0188	0,137	
20																														21	0,0194	0,139	
21																														19	0,0199	0,141	
22																														23	0,0214	0,146	
23																														26	0,0233	0,152	
24																														27	0,0254	0,159	
25																														6	0,0276	0,166	
26																														24	0,0302	0,174	
27																														22	0,0341	0,185	
28																														30	0,0413	0,203	
29																														28	0,0592	0,243	
30																														29	0,0941	0,307	

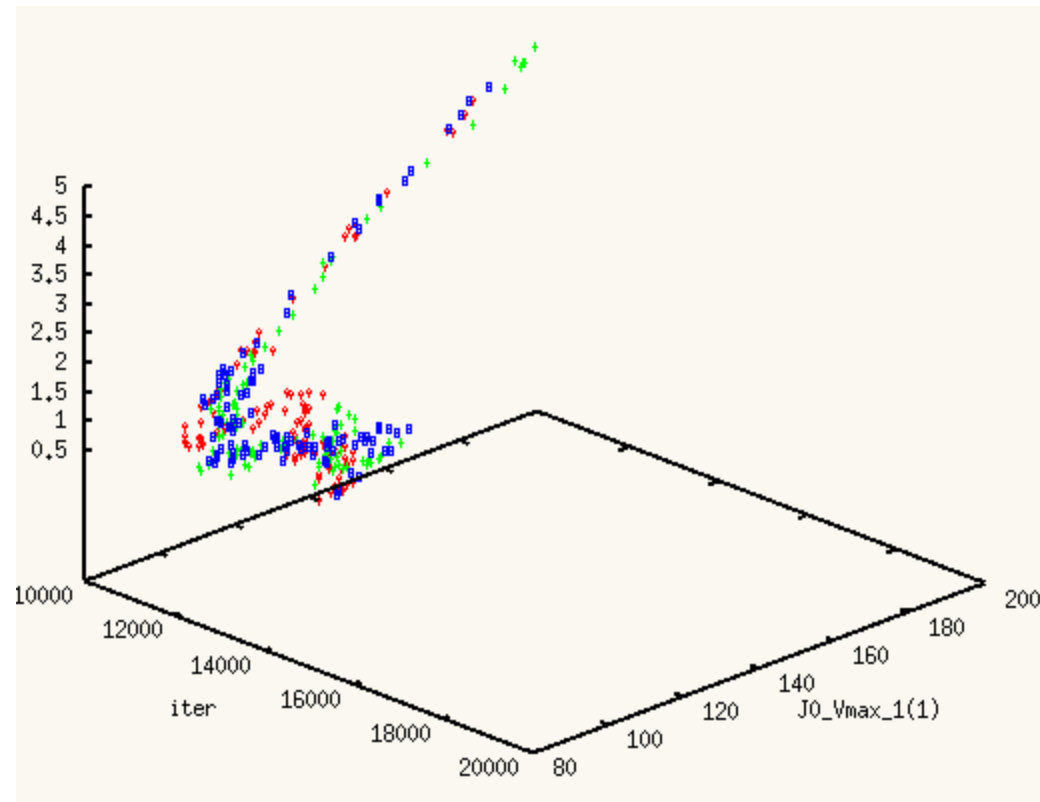
Firstly proposed design points

Moving to a new posterior distribution



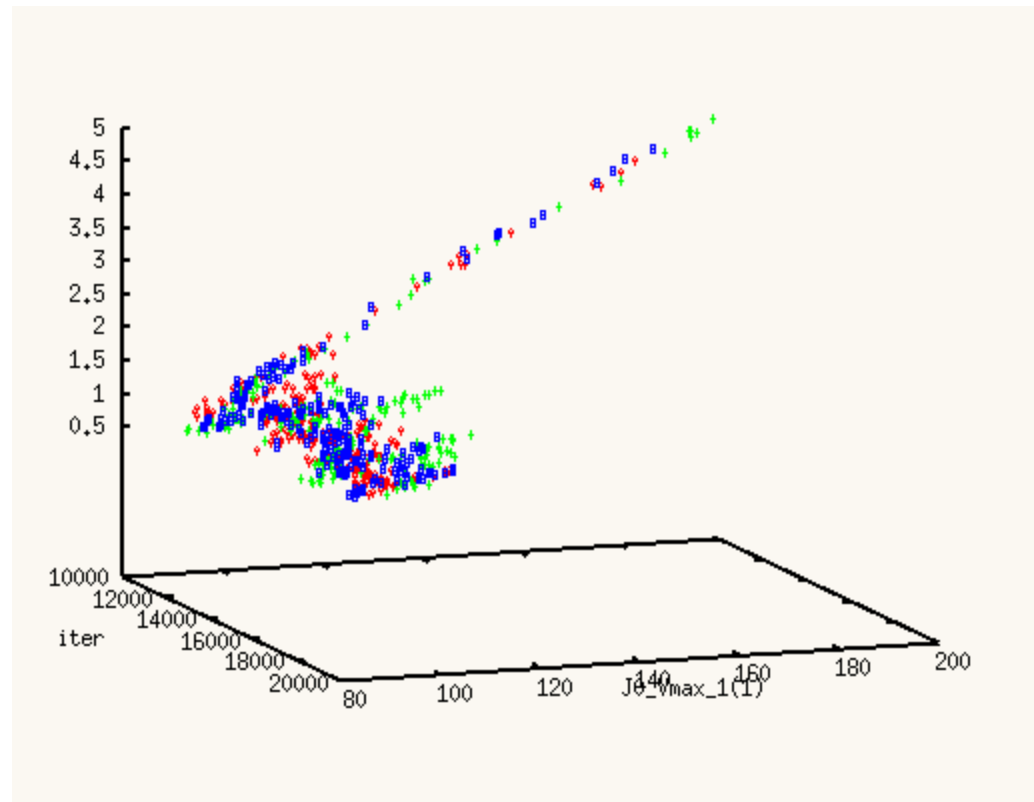
We start from a random point in the previous posterior but add the new data: we move to a new posterior.

Moving to a new posterior distribution



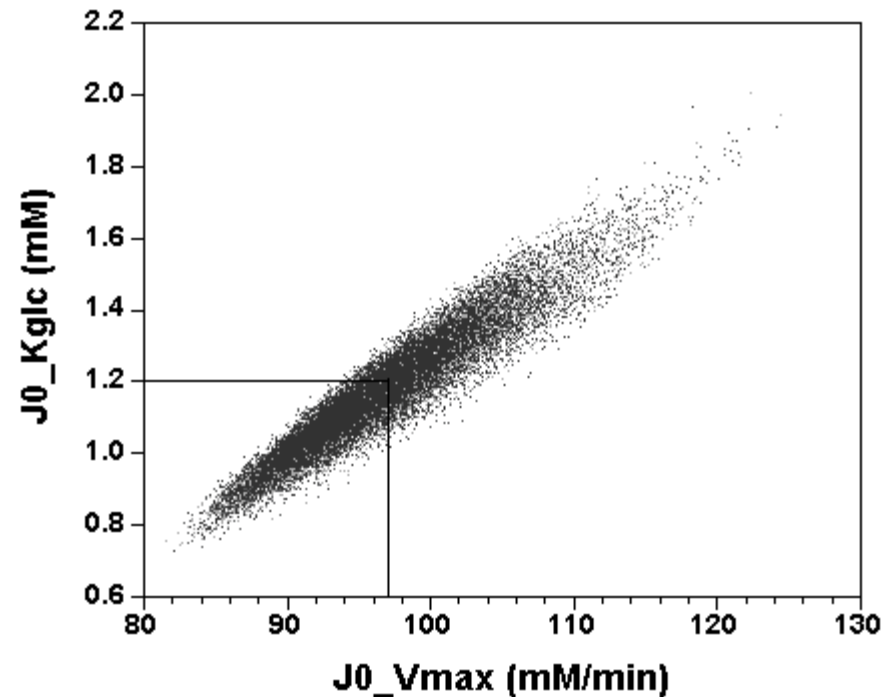
4000 iterations and a few seconds later...

Moving to a new posterior distribution



After about 10000 new iterations we converge to...

New (updated) posterior distribution



The correct answer!

Full story is at:

http://www.gnu.org/software/mcsim/supplement_bioinformatics_2009.html