# Systems biology and beyond
# –
# From data to mathematical models

Susanna Röblitz
Computational Systems Biology Group
ZIB

March 1st, 2013

Zuse Institute Berlin

DFG Research Center
Matheon

# Overview

1. Who we are and what we are doing
2. A typical research example: Mathematical modelling of the human menstrual cycle
3. Further research topics
4. Software

# Who we are...

**Zuse Institute Berlin (ZIB)**, Takustr. 7, 14195 Berlin `http://www.zib.de`

- Founded by P. Deuflhard in 1986 as a research institute for Computational Mathematics
- Basic funding by the State of Berlin; strong support by third party funding
- About 200 employees (50% research, 50% service)
- R&D branches: Numerical Mathematics, Discrete Mathematics, Computer Science
- High-level services for high-performance computing

# Group Members

## Head

Susanna Röblitz

## Staff

Rainald Ehrig

Claudia Stötzel

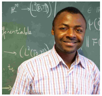
Thomas Dierkes

Mascha Berg

## Students

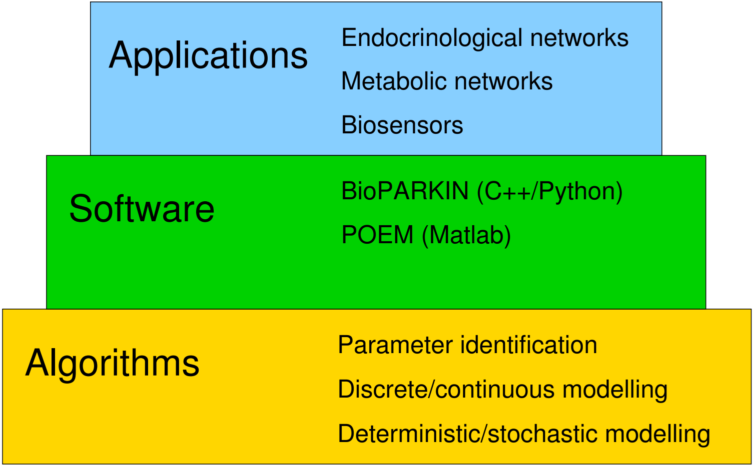
Julia Plöntzke

Gabriel Müller



Dany Pascal Moualeu

# Activities



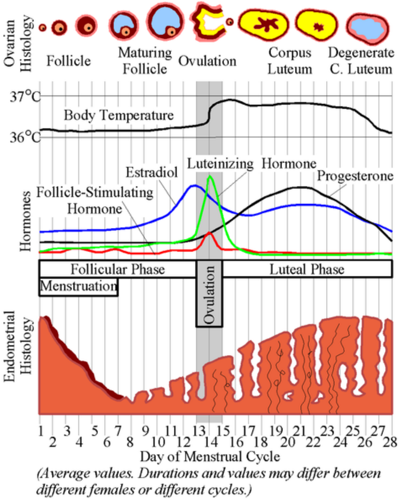
**Applications**
- Endocrinological networks
- Metabolic networks
- Biosensors

**Software**
- BioPARKIN (C++/Python)
- POEM (Matlab)

**Algorithms**
- Parameter identification
- Discrete/continuous modelling
- Deterministic/stochastic modelling

# Mathematical modelling of the human menstrual cycle

# The human menstrual cycle



*(Average values. Durations and values may differ between different females or different cycles.)*

(http://www.websters-online-dictionary.org/definitions/Menstrual Cycle)

# Aims

First step

- a model for the idealized cycle of an idealized woman
- calculation of hormone profiles and follicle development over time

What can we use such a model for?

- find parametrizations for individual women, study of infertility or genetic differences
- simulation of external effects (administration of drugs, design of hormone therapies)

# Overview

Models available before

- smaller compartment models: fewer (though essential) compartments coupled by coarse interactions

  [Schlosser/Selgrade 2000, Harris 2001]

- small biochemical reaction models (such as receptor binding): often not fully integrated into total model

  [Clément 2001]

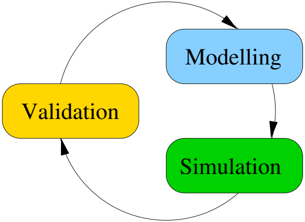
ZIB Model **GynCycle**

- fully integrated compartment model: containing most relevant couplings, both physiological and biochemical ones
- modifications for special purposes:
    - model reduction (to focus on specific aspects)
    - model expansion (to catch further effects)

# Systems biology approach

- the cycle as a whole system rather than only focusing on individual parts
- investigate how the components function together
- find reliable abstraction levels that display the most important mechanisms

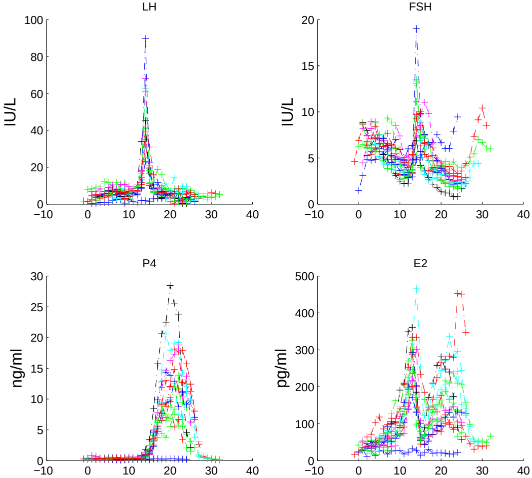
Mathematical task::



- Model development and parametrization such that simulation results match given measurement data

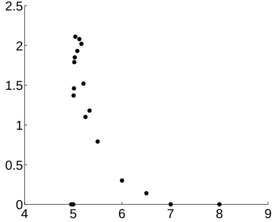
# Data

# Data for normal cycle

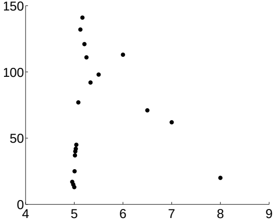


measurement values for 12 women over one cycle (Pfizer R&D)

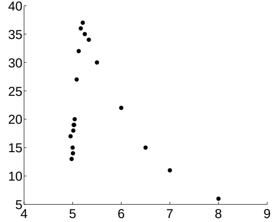
# Data for single dose nafarelin



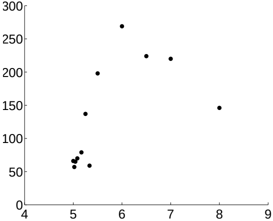
nafarelin



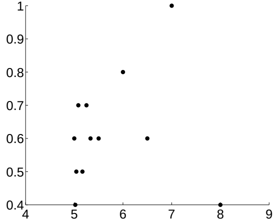
LH



FSH



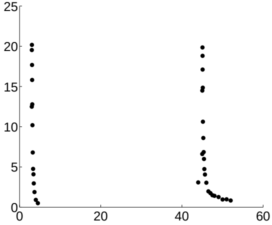
E2



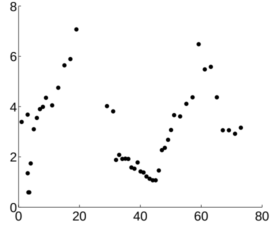
P4

GnRH-Agonist: activation of GnRH receptor resulting in (initially) increased secretion of FSH and LH, followed by a drop in gonadotropin secretion caused by receptor downregulation

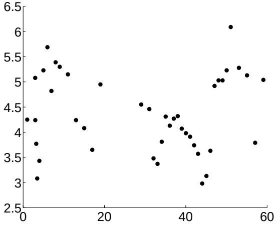
# Data for single and multiple dose cetrorelix



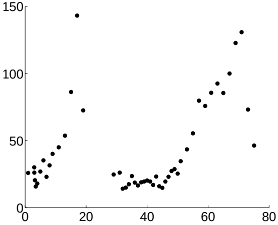
cetrorelix



LH



FSH



E2

GnRH-Antagonist: competitive and reversible binding to GnRH receptors, immediate drop in gonadotropin secretion

# Typical problems

- different physical units, sometimes not even convertible
- missing measurement errors
- missing information about the cycle day
- averaged data for women with different cycle length or in different stages of the cycle

ideal measurement: data for a single woman together with the length of her cycle, the stage of the cycle (days since last menses), and measurement errors

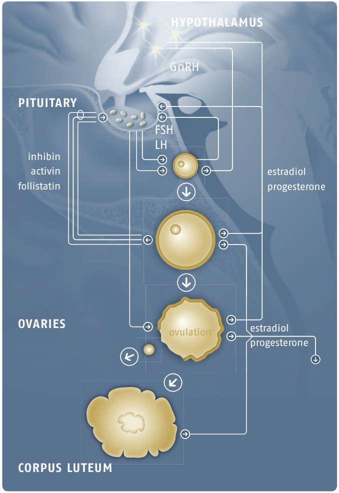
# Model Development

# Components and compartments considered

Compartments: blood, ovaries, uterus, pituitary, hypothalamus

Components:

- Estradiol
- Progesterone
- Inhibin A and B
- LH + receptor binding
- FSH + receptor binding
- GnRH + receptor binding
- 6 follicular stages
- 6 luteal stages (corpus luteum)



# Differential equations

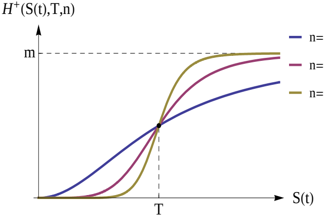
Abstract mathematical formulation:

$$y'(t) = f(t, y(t), p), \qquad y(t=0) = y_0$$

Production, clearance, synthesis and release usually depend on other components. To model stimulation or inhibition we use

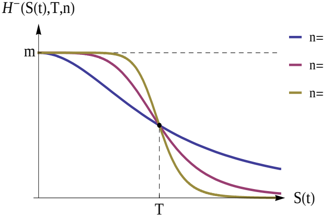
positive Hill functions

$$H^+(S(t), T, n) := \frac{S(t)^n}{S(t)^n + T^n}$$

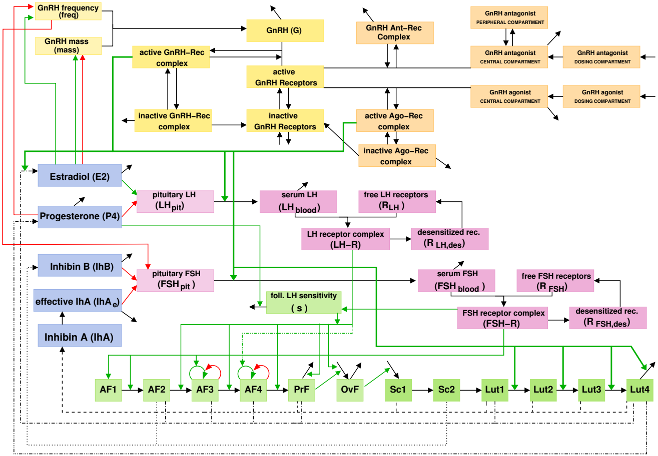


negative Hill functions

$$H^-(S(t), T, n) := \frac{T^n}{S(t)^n + T^n}$$



# Current model



GnRH frequency (freq)
GnRH mass (mass)
GnRH (G)
GnRH Ant–Rec Complex
GnRH antagonist PERIPHERAL COMPARTMENT
GnRH antagonist CENTRAL COMPARTMENT
GnRH antagonist DOSING COMPARTMENT
active GnRH–Rec complex
active GnRH Receptors
GnRH agonist CENTRAL COMPARTMENT
GnRH agonist DOSING COMPARTMENT
inactive GnRH–Rec complex
inactive GnRH Receptors
active Ago–Rec complex
inactive Ago–Rec complex
Estradiol (E2)
Progesterone (P4)
pituitary LH ($LH_{pit}$)
serum LH ($LH_{blood}$)
free LH receptors ($R_{LH}$)
LH receptor complex (LH–R)
desensitized rec. ($R_{LH,des}$)
Inhibin B (IhB)
pituitary FSH ($FSH_{pit}$)
serum FSH ($FSH_{blood}$)
free FSH receptors ($R_{FSH}$)
effective IhA ($IhA_e$)
foll. LH sensitivity (s)
FSH receptor complex (FSH–R)
desensitized rec. ($R_{FSH,des}$)
Inhibin A (IhA)
AF1
AF2
AF3
AF4
PrF
OvF
Sc1
Sc2
Lut1
Lut2
Lut3
Lut4

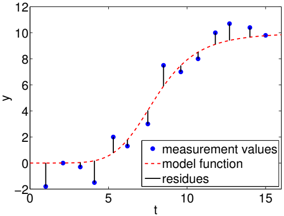
**2007**: 49 DDEs, 208 parameters, 9 identifiable [Reinecke,Deuflhard (2007)]

**2012**: 33(+8) ODEs, 114 parameters, 63 identifiable [Röblitz et al. (2013)]

# Parameter Identification

# A nonlinear least squares problem

Model $y(t,p) = (y_1(t,p), \ldots, y_n(t,p))$
Parameters $p = (p_1, ..., p_q)$
Data $x = (x_1, ..., x_m), m \geqslant q$
mostly $m \gg q$, data compression



Least squares formulation:

$$\|F(p)\|_2^2 = \sum_{i=1}^{m} \left( \frac{x_i - y(t_i,p)}{\delta x_i} \right)^2 \rightarrow \min$$

$\delta x_i$ : measurement accuracy for the $x_i$ (never forget!)
**relative** measurement accuracy:

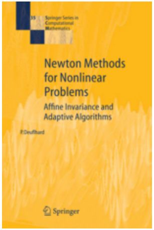
$$\delta x_i = \varepsilon_x x_i$$

$\varepsilon_x$ mostly $10^{-1}$ to $10^{-3}$ in experiments

# Gauss-Newton algorithm

- solution of the nonlinear least squares problem by error-oriented global Gauss-Newton method

$$\|F'(p^{(k)})\Delta p^{(k)} + F(p^{(k)})\|^2 \rightarrow min$$
$$p^{(k+1)} = p^{(k)} + \lambda_k \Delta p^{(k)}, k = 0, 1, 2, \ldots$$



Newton Methods for Nonlinear Problems

Affine Invariance and Adaptive Algorithms

P. Deuflhard

Springer

[Deuflhard: Newton Methods for Nonlinear Problems, 2004]

- sequence of linear least squares problems with $(m \times q)$ Jacobian matrix $F'(p)$
- good **initial guess** required (model decomposition)
- $F'(p)$ gives us some hints whether the current combination of model and data will permit an actual identification of the parameters

# Identifiability

- the rows of $F'(p)$ contain the sensitivities of the measured components w. r. t. the parameters $p = (p_1, \ldots, p_q)$

$$F'_{ij}(p) = \frac{\partial}{\partial p_j} y_{k_i}(t_i, p), \quad i = 1, \ldots, m, \quad k_i \in \{1, \ldots, n\}$$

- solution of linear LSQ problems by QR factorization with column pivoting

$$F'(p)\Pi = QR, \quad r_{11} \geqslant r_{22} \geqslant \ldots \geqslant r_{qq}$$

- detection of linear dependencies by monitoring the subcondition numbers

$$\mathrm{sc}_j = r_{11}/r_{jj}$$

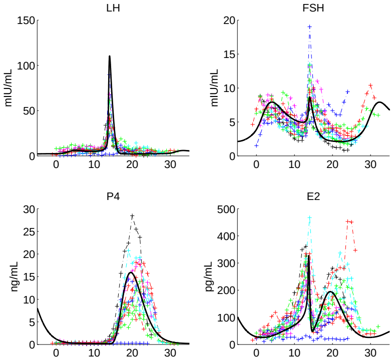
**identifiable parameters**: $\varepsilon_x \mathrm{sc}_j < 1$ [Deuflhard/Sautter 1980]

- estimation of incompatibility factor $\kappa < 1$ as asymptotic convergence rate

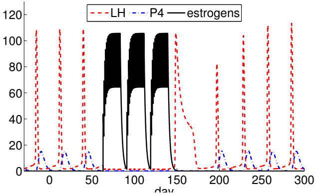
# Simulation Results

# Results

- normal cycle simulation



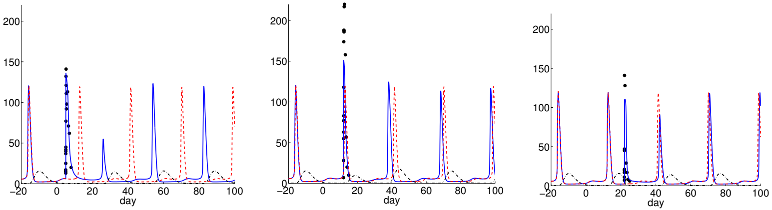
- simulating the effect of birth control pills



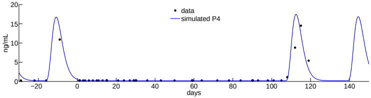
# Applications

Analyzing the role of dose and timing of certain drugs

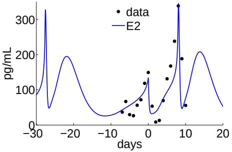
- single dose agonist (nafarelin)



- multiple dose agonist (nafarelin)



- single dose antagonist (cetrorelix)

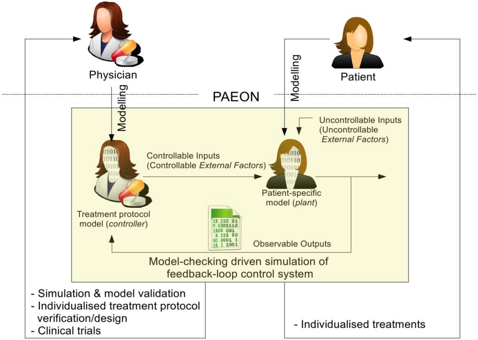


# Perspective

## PAEON: Model-Driven Computation of Treatments for Infertility Related Endocrinological Diseases

- 3 years EU project
- development of patient-specific models
- cooperation partners: Università di Roma "La Sapienza", Hochschule Luzern, Universitätsspital Zürich,

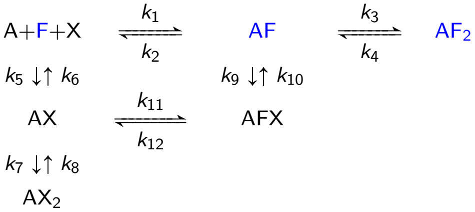
  Med. Hochschule Hannover



# Further topics

# Model-based spectrometer calibration for toxin determination in food

A: antibody
X: analyte (mycotoxins)
F: fluorescence-marked analyte

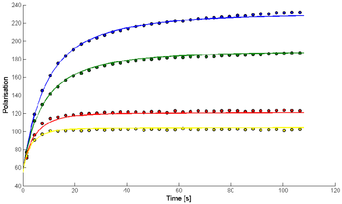


$$\begin{tikzcd}
A+F+X \arrow[r, shift left, "k_1"] \arrow[d, shift right, "k_5"'] & AF \arrow[l, shift left, "k_2"] \arrow[r, shift left, "k_3"] \arrow[d, shift right, "k_9"'] & AF_2 \arrow[l, shift left, "k_4"] \\
AX \arrow[u, shift right, "k_6"'] \arrow[r, shift left, "k_{11}"] \arrow[d, shift right, "k_7"'] & AFX \arrow[l, shift left, "k_{12}"] \arrow[u, shift right, "k_{10}"'] & \\
AX_2 \arrow[u, shift right, "k_8"'] & &
\end{tikzcd}$$

aokin spectrometer FP470

Polarisation of fluorescent light:

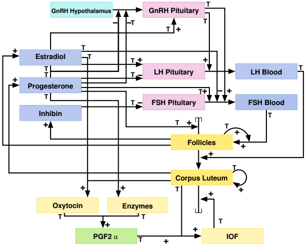
$$P = \frac{p_F[F]Q_F + p_{AF}[AF]Q_{AF} + 2p_{AF_2}[AF_2]Q_{AF_2}}{[F]Q_F + [AF]Q_{AF} + 2[AF_2]Q_{AF_2}}$$



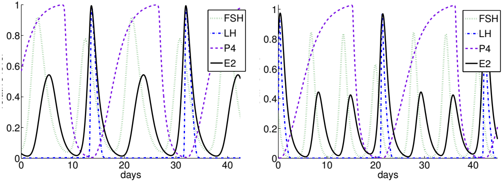
# BovCycle

## A mathematical model of the bovine estrous cycle

15 ODEs, 60 parameters

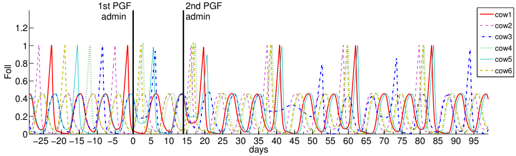


GnRH Hypothalamus
GnRH Pituitary
Estradiol
LH Pituitary
LH Blood
Progesterone
FSH Pituitary
FSH Blood
Inhibin
Follicles
Corpus Luteum
Oxytocin
Enzymes
PGF2 α
IOF



analysis of follicular wave patterns

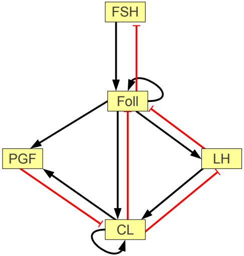
Cooperations:
FU Veterinary Medicine
U Wageningen



numerical validation of synchronization protocols

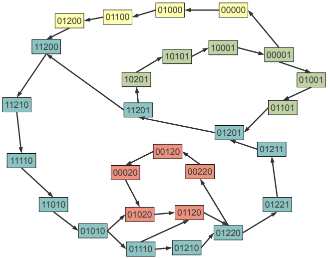
[Boer, Stötzel, Röblitz, Deuflhard, Veerkamp, Woelders, *J. Theoret. Biol.* 2011]
[Stötzel, Plöntzke, Heuwieser, Röblitz, *Theriogenology* 2012]

# Discrete/continuous modeling



Logic-based models:

- discrete variables and parameters
- efficient analysis methods
  - modularization – state space analysis
  - model checking – parameter/state space analysis



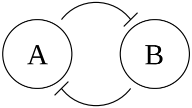
Integrated approach:

- parameter identification
- comprehensive state space analysis

with A. Bockmayr, H. Siebert (FU Berlin)

# Deterministic/stochastic modeling

## Rare events in chemical reaction systems



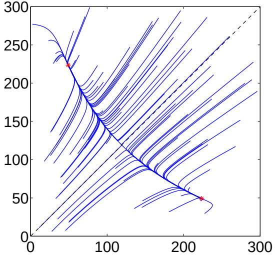
mutually repressing gene pair (e.g. bacteriophage-$\lambda$) with two competing proteins

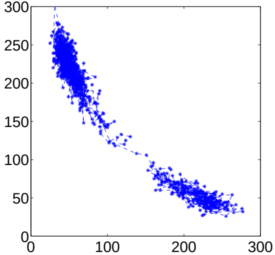
[Gardner et al., Nature 403 (2000)]

ordinary differential equations

$$A' = c_1/(c_2 + B^\beta) - c_3 A$$
$$B' = c_4/(c_5 + A^\gamma) - c_6 B$$

continuous-time discrete-space Markov chain (SSA [Gillespie, 1977])

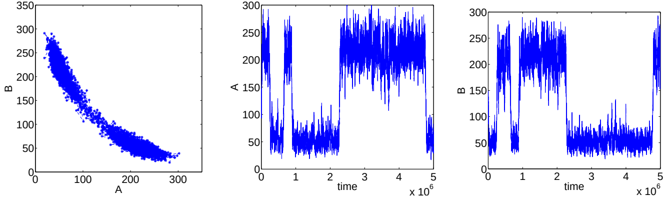




# Deterministic/stochastic modeling

only 5 transitions between $\{(A,B) : A > B\}$ and $\{(A,B) : A < B\}$ within $5 \cdot 10^4$ steps ⇒ poor statistics

29971 : 20016   (theoretically 1:1)



reduced description of the dynamical system in terms of nearly invariant (metastable) sets/functions

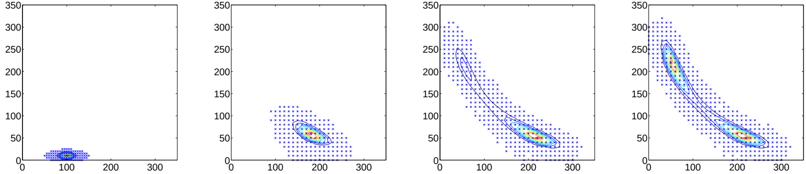
$$Q_c = 10^{-5} \begin{pmatrix} -0.3288 & 0.3288 \\ 0.3094 & -0.3094 \end{pmatrix}$$

# Deterministic/stochastic modeling: CME

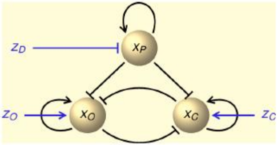
## Chemical Master Equation (CME)

$$\partial_t p(x,t) \;=\; \sum_{\substack{r=1\\ x-\nu_r\geqslant 0}}^{R} \alpha_r(x-\nu_r)\,p(x-\nu_r,t) - \sum_{\substack{r=1\\ x+\nu_r\geqslant 0}}^{R} \alpha_r(x)\,p(x,t)$$

Solution by meshfree discrete Galerkin methods



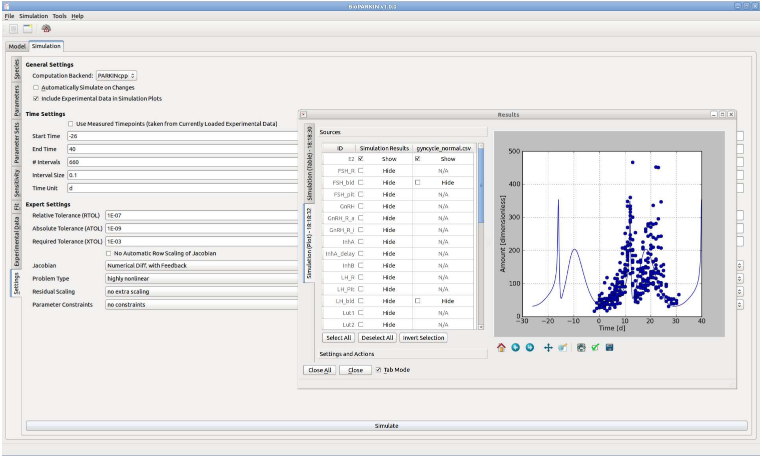
application to a model for differentiation of progenitor cells into bone or cartilage cells (with S. Waldherr, FU)



# Software

# BioPARKIN

integrated software environment for simulation and parameter identification www.bioparkin.zib.de



**General Settings**

Computation Backend: PARKINcpp

☐ Automatically Simulate on Changes

☑ Include Experimental Data in Simulation Plots

**Time Settings**

☐ Use Measured Timepoints (taken from Currently Loaded Experimental Data)

Start Time: -26

End Time: 40

# Intervals: 660

Interval Size: 0.1

Time Unit: d

**Expert Settings**

Relative Tolerance (RTOL): 1E-07

Absolute Tolerance (ATOL): 1E-09

Required Tolerance (XTOL): 1E-03

☐ No Automatic Row Scaling of Jacobian

Jacobian: Numerical Diff. with Feedback

Problem Type: highly nonlinear

Residual Scaling: no extra scaling

Parameter Constraints: no constraints

Sources

| ID | Simulation Results | gyncycle_normal.csv |
|---|---|---|
| E2 | ☑ Show | ☑ Show |
| FSH_R | ☐ Hide | N/A |
| FSH_bld | ☐ Hide | ☐ Hide |
| FSH_pit | ☐ Hide | N/A |
| GnRH | ☐ Hide | N/A |
| GnRH_R_a | ☐ Hide | N/A |
| GnRH_R_i | ☐ Hide | N/A |
| InhA | ☐ Hide | N/A |
| InhA_delay | ☐ Hide | N/A |
| InhB | ☐ Hide | N/A |
| LH_R | ☐ Hide | N/A |
| LH_Pit | ☐ Hide | N/A |
| LH_bld | ☐ Hide | ☐ Hide |
| Lut1 | ☐ Hide | N/A |
| Lut2 | ☐ Hide | N/A |

Settings and Actions

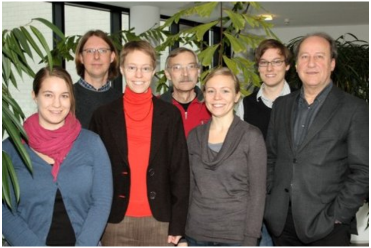
☑ Tab Mode

# BioPARKIN

- C++ stand alone library for numerical routines
- Graphical user interface for intuitive model handling
- cross-platform development for Windows, Mac OS X, Linux
- open source
- SBML import and export
- efficient integrator for differential-algebraic equations (LIMEX Linearly Implicit Euler method with EXtrapolation, highly cited in Google-Scholar)

  http://www.zib.de/en/numerik/software/codelib/ivpode.html
- sensitivity analysis based on variational equations
- parameter estimation via Gauss-Newton methods (tunable, in particular scaling of species and parameters possible)
- output of information on identifiability of parameters
- time-shifting of data and concatination of IVPs for multi-experiment simulations

# Thank you for your attention!



Contact:
Zuse Institute Berlin
Computational Systems Biology Group
http://www.zib.de/en/numerik/csb.html