

Approximate Bayesian Computing for Parameter Inference in Hybrid Discrete-Continuum Models

Nick Jagiella¹, Dennis Rickert¹, Fabian J. Theis^{1,2,*}, Jan Hasenauer^{1,2,**}

1 Institute of Computational Biology, Helmholtz Zentrum München, München, Germany

2 Zentrum Mathematik, Technische Universität München, Garching, Germany

* fabian.theis@helmholtz-muenchen.de

** jan.hasenauer@helmholtz-muenchen.de

Contents

Abstract

To do.

Author Summary

To do.

Introduction

To do.

Results

ABC implementation for computationally demanding models

We used a parallelized version of the well established ABC SMC method for parameter inference (see Material and Methods section). Fig. 1 illustrates the pipeline used for parallelization. There are many possibilities how to parallelize (multi-core, GPU, cluster, etc.). As we aim for resource-wise and computationally expensive models, we make use of a queue-mediated cluster architecture. Here a master is running the actual ABC SMC routine and is outsourcing the time and resource consuming

viele Parallelisierungsansätze, point out that m consecutive ones are only accepted if all intermingled simulations finished

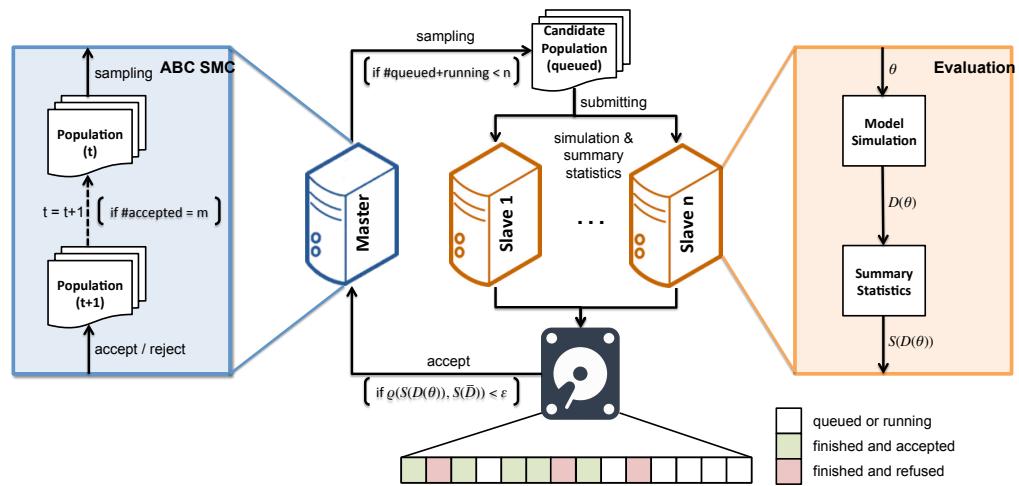


Figure 1. Pipeline using cluster. The ABC method runs on a master machine. Each iteration t new candidate parameters are drawn from a prior distribution and submitted to a queue. Slaves pick up candidates, simulate the model and calculate the summary statistics / distance to data. Then the master accepts those candidate with distances below a threshold ϵ and replaces all finished model evaluations with new samples on the queue.

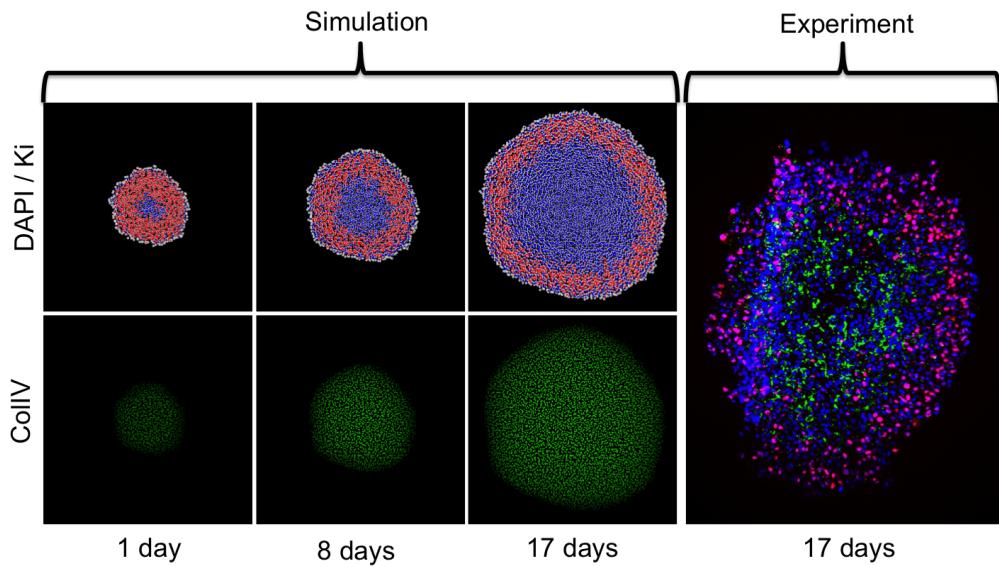


Figure 2. Histological information in in-silicio and in-vitro data.

Artificial data (2D)

Parameter Inference Fig. 2 shows the comparison of data and model prediction and its evolution over the iterations.

Fig. 3 shows the comparison of data and model prediction of the parameter populations evolving over iterations.

For the inference of the model parameters used to create the artificial data all parameters can be identified. Nevertheless, the coupling parameter between cellular model and extracellular matrix, e^{crit} , needs a lot of evaluations and remains with a very large uncertainty.

We also can observe that the acceptance rate of candidate samples becomes dramatically low for epsilon < distance at the optimum itself.

22

23

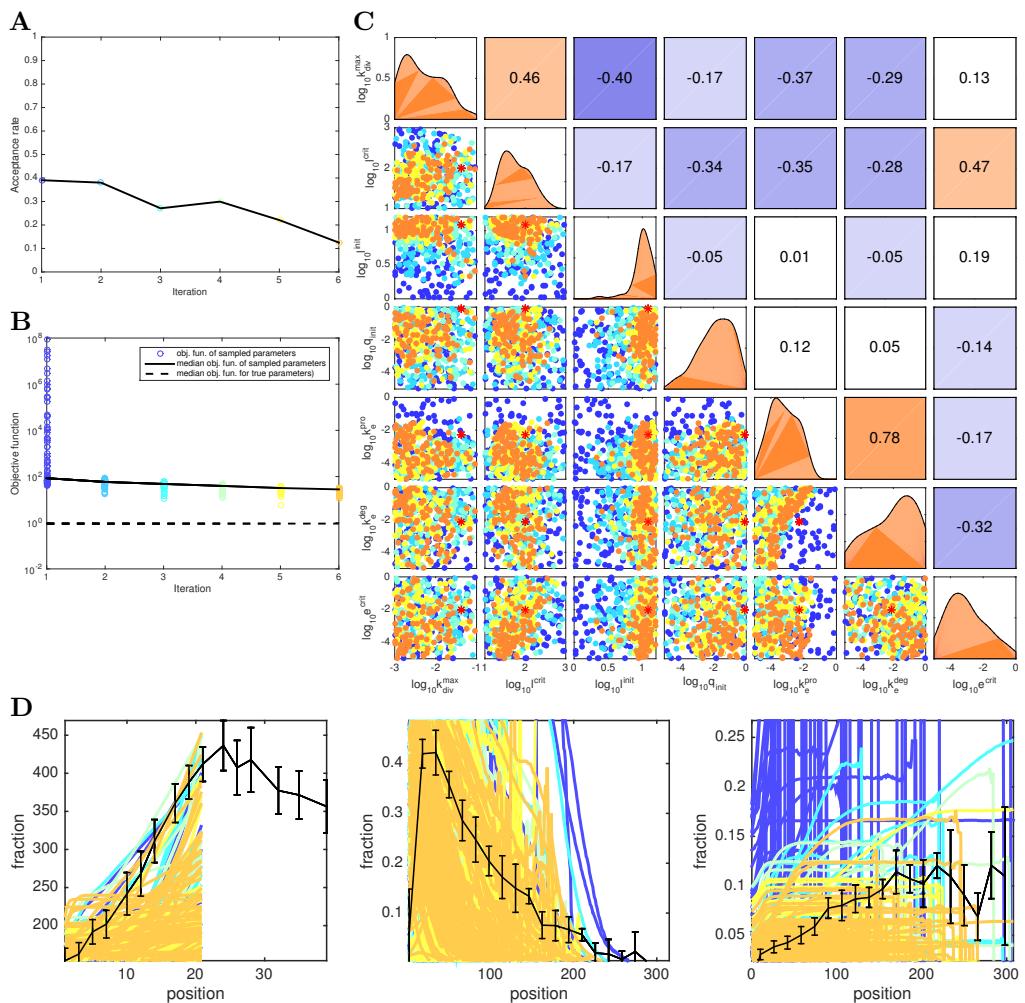


Figure 3. Artificial data and fits for 5 (color-coded) generation (complete dataset). **A** acceptance rate over iteration; **B** ϵ threshold (cyan) and distance of the accepted population (blue) over iteration; for comparison the median distance at the optimum is depicted (red); **C** scatter matrix for 5 (color-coded) generation. todo: color spectrum. **D** Artificial data and fits for 5 (color-coded) generations (complete dataset).

24

25

26

27

28

29

Sensitivity to Population/Sample Size Fig. 4 illustrates that the population size has an critical impact on the convergence of the algorithm. If the population size is chosen too small, in our case 20, then convergence can not be assured. On the other hand we observe no significant improvement for an increase from 100 to 200. So for the means of limiting the cluster load per iteration all following parameter inference runs will be done with a population size of 100.

30

31

32

33

Experimental data (2D)

Fig. 5

no histological info on proliferation (Ki67) leads to wrong predictions cellular kinetic param. kinetic ECM param. not identifiable without hist. info. on ECM (ColV)

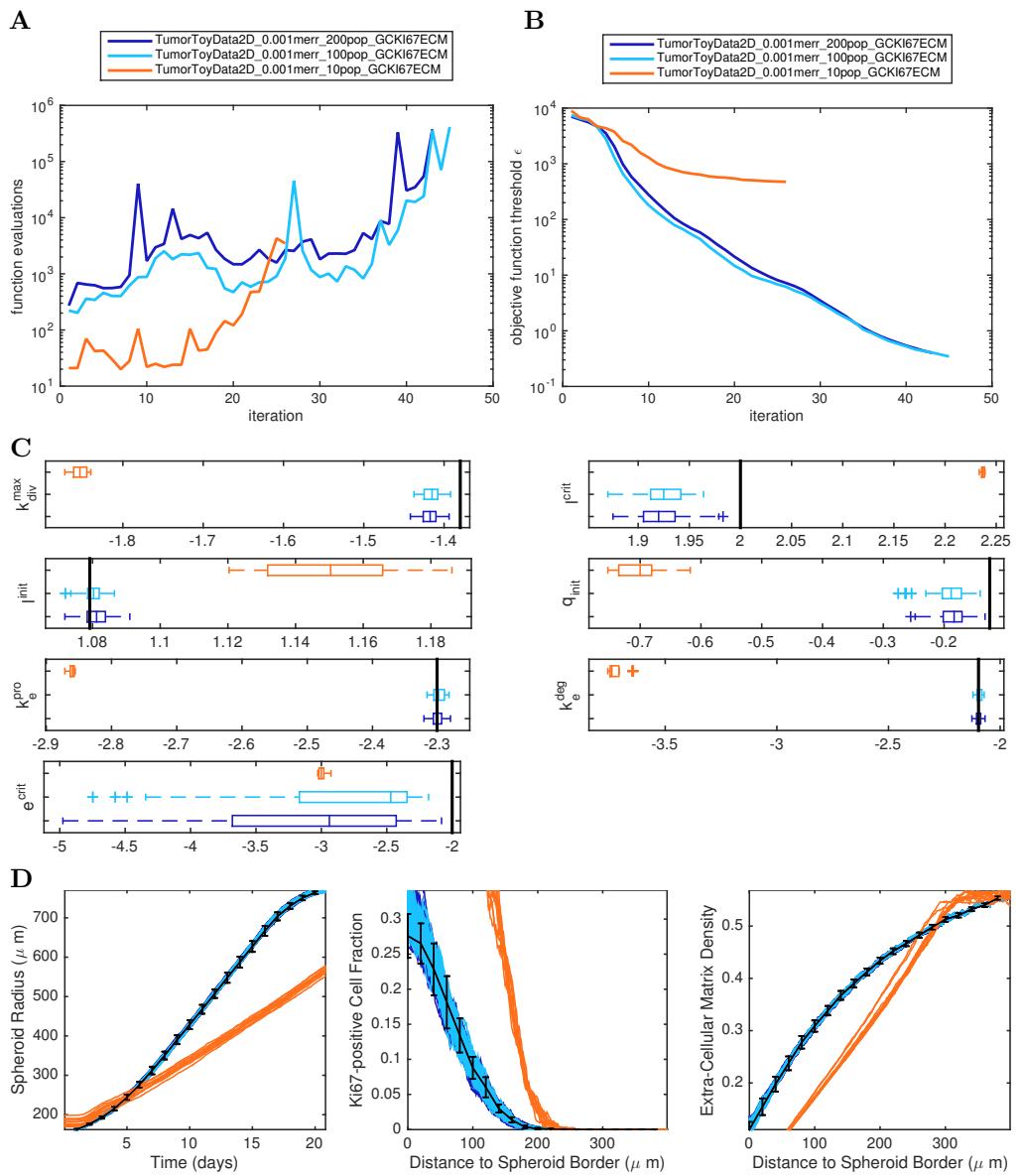


Figure 4. Sensitivity to Population/Sample Size. **A** number of function evaluations over iteration; **B** threshold over iteration; **C** box plot of final sample for different population sizes. **D** final fits.

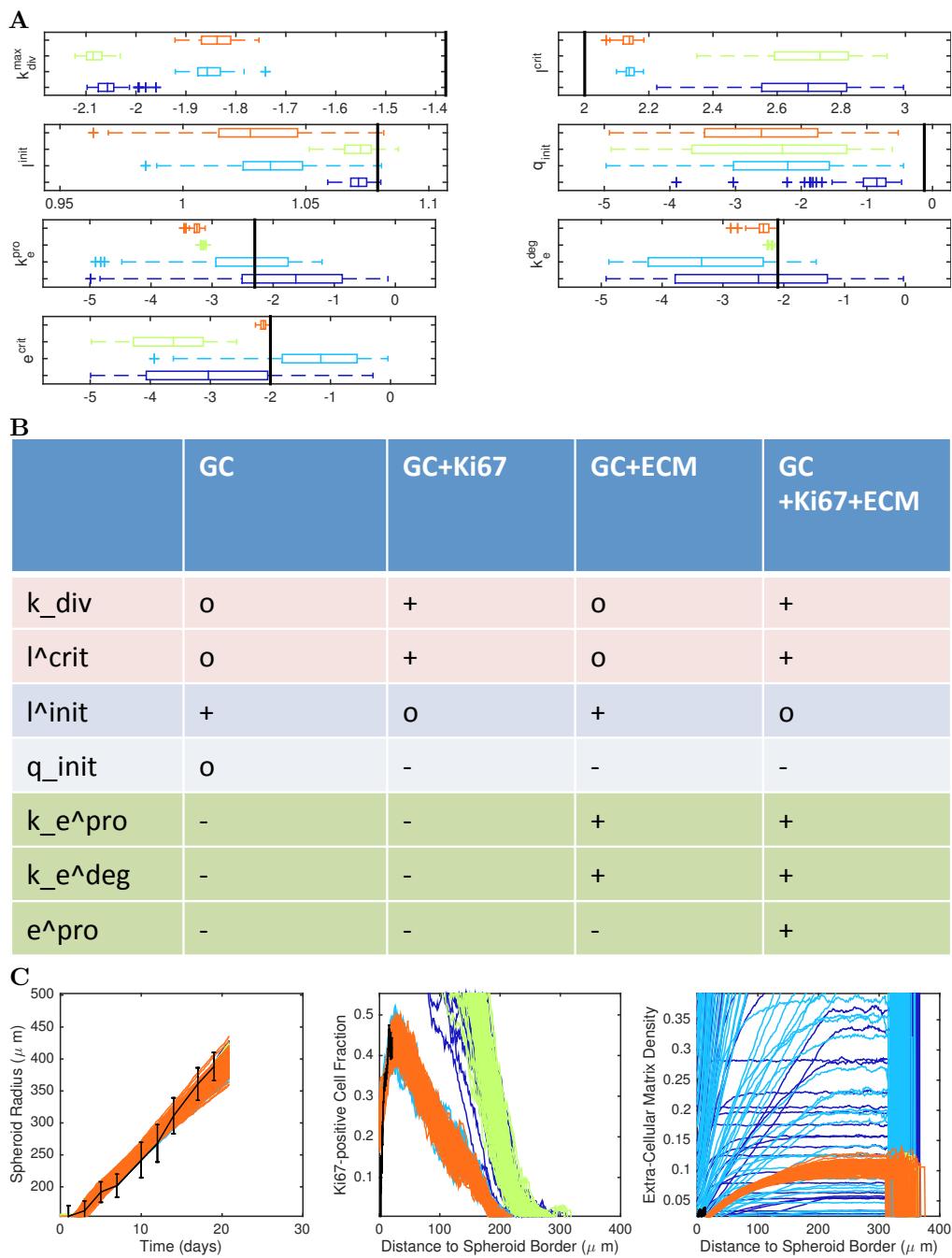


Figure 5. Different combinations of experimental data sets. **A** box plot for different combinations of data sets; **B** identifiability table (+ identifiable, o large uncertainty, - unidentifiable); **C** final fits and scenarios.

Discussion

Materials and Methods

Model

Tempo-spatial multi-scale model to simulate in-vitro tutor growth. Hybrid-model of individual-based model for tutor cells and a continuum model describing the molecular kinetics of extra-cellular matrix (e) and the key metabolites, glucose (g), oxygen (o) and ATP (a).
 37
 38
 39
 40

Cell model: Cells a) populate a static unstructured lattice (max. 1 cell per lattice site), b) can be either viable or necrotic, and c) can perform a birth and death process if viable or a lysis process if necrotic. The respective transition rates are k_{div} , k_{nec} , k_{lys} and depend on local molecular concentrations
 37
 38
 39
 40

$$k_{div} = k_{div}^{max} H(e - e^{crit}) H(a) H(l - l^{crit}) \quad (1)$$

$$k_{nec} = k_{nec}^{max} H(-a). \quad (2)$$

Molecular model: The molecular dynamics is described by system of partial differential equations as follows
 41
 42
 43

$$\partial_t e = D\nabla e + k_{pro}^e c - k_{deg}^e e \quad (3)$$

$$\partial_t g = D\nabla g - k_{con}^g c \quad (4)$$

$$\partial_t o = D\nabla o - k_{con}^o c \quad (5)$$

$$\partial_t a = k_{pro}^a c - k_{con}^a c, k_{pro}^a = 2k_{con}^g c + 17/3k_{con}^o. \quad (6)$$

Initial & boundary conditions: The initial cell population occupies all lattice sites within a sphere of radius l^{init} . A fraction of those cells q_{init} is quiescent, while the rest enters the cell cycle.
 41
 42
 43

Parameters Resulting model is stochastic and only numerically solvable. Here we use the Gillespie algorithm and solve the steady state problem for the molecular system after each update. The parameters that are subject to optimisation are indicated in red.
 44
 45
 46

Data

We used two types of data: the growth curves of multi-cellular spheroids over time and histological information on the spatial distribution of proliferating cells and extra-cellular matrix at a certain moment of the experiment.
 47
 48
 49
 50

ABC

The ABC SMC method used in this paper is based on Toni et al. [1]. The idea can be briefly summarised as the following:
 51
 52
 53

S1) initialize: $t = 1, p_1(\theta) = \pi(\theta), \epsilon_1 = \infty$

S2.0) set $i = 1$ #new generation

S2.1) draw θ_t^i from proposal distribution p_t #sample

S2.2) draw $y_t^i = f(\hat{x}, \theta_t^i)$ #simulate

S2.3) if $d(y_t^i, \hat{y}) \geq \epsilon_t$ go to S2.1. #reject 58

S2.4) if $i = n$ set $t = t + 1$ and go to S2.0. #next generation 59

Each iteration $t \geq 1$ we sample candidate parameter vectors θ^* from a prior distribution $p^{(t)}(\theta)$ and evaluate the model $y^* = f(x, \theta^*)$. If the corresponding objective function value is $\delta(y^*, y) < \varepsilon^{(t)}$, then θ^* will be accepted and added to $\Theta^{(t)}$. If a minimal number of n accepted parameter vectors is reached, then the algorithm will proceed to the next iteration, $t = t + 1$. The main extension of ABC SMC compared to ABC is to iteratively adapt both, the prior distribution $p^{(t)}(\theta)$ and the acceptance threshold $\varepsilon^{(t)}$. 60
61
62
63
64
65

Choice of Prior Distribution $p^{(t)}(\theta)$ Here we chose a Gaussian perturbation kernel: 66
67

$$p^{(t)}(\theta) = \sum_i \frac{1}{w_i^{(t)}} (2\pi)^{-k/2} |\Sigma|^{-1/2} e^{-1/2(x-\mu)^T \Sigma^{-1}(x-\mu)}, \quad (7)$$

where μ is the mean and Σ is the standard deviation of the current population t . 68

Choice of ε -thresholds The threshold distance ϵ for accepting a candidate parameter is chosen to be the median among the actual population 69
70

$$\epsilon^{(t)} = \text{median}\{\theta^{(t)}\} \quad (8)$$

(Surrogate Approximation)

Supporting Information

Acknowledgments

References

1. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of The Royal Society Interface. 2009;6(31):187–202.