pyABC

A framework for distributed likelihood-free inference

Yannik Schälte[1,2] Emmanuel Klinger[1,2,3] Dennis Rickert[1] Jan Hasenauer[1,2]

[1]Helmholtz Zentrum München, Institute of Computational Biology

[2]Technical University Munich, Department of Mathematics

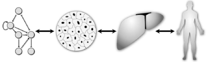
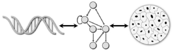
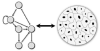
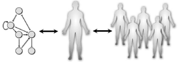
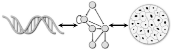
[3]Max Planck Institute for Brain Research Frankfurt, Department of Connectomics

SBMC Bremen, 2018-07-05

HelmholtzZentrum münchen
German Research Center for Environmental Health

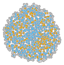
mpibrain research

Multi-scale models

		Boolean / Petri net models	Constraint-based models	Markov chain models	SDE models	ODE models	PDE models	Agent-based models	
whole-heart (tissue and fluid mechanics, signaling)		■				■	■		Hunter and Borg, *Nature Reviews Molecular Cell Biology*, 4(3):237-243, 2003
cancer growth (signaling, cell division and death, angiogenesis, tissue remodeling)					■		■	■	Anderson and Quaranta, *Nature Reviews Cancer*, 8(3):227-234. 2008
liver lobule (cell division and cell death, tissue mechanics)				■	■		■	■	Hoehme et al., *PNAS*, 107(23):10371-10376, 2010
glucose-insulin-glucagon regulation (blood and interstitial flow, organ uptake, signaling)						■	■	■	Schaller et al., *CPT: Pharmacometrics and Systems Pharmacology*, 2:e65, 2013
whole-cell (transcription, translation, DNA replication, metabolism, replication)		■	■	■		■		■	Karr et al., *Cell*, 150(2):389-401, 2012

[1] [Hasenauer, *Data-driven modeling of biological multi-scale processes*, J. Coup. Sys. and Mult. Dyn., 2015]

Example: Multi-scale model of tumor growth[1]



proliferating cells

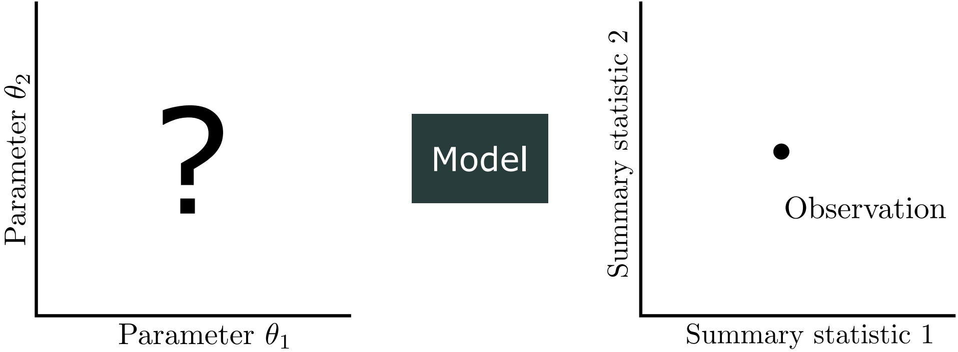
- Hybrid discrete-continuous model
- Cells modeled as interacting agents, dynamics of extracellular substances by reaction-diffusion equations
- Simulate up to 10^6 cancer cells
- 10s - 1h for one forward simulation

[1] [Jagiella et al., *Parallelization and high-performance computing enables automated statistical inference of multi-scale models*, Cell Systems, 2017]

How to do parameter inference for stochastic multi-scale models?

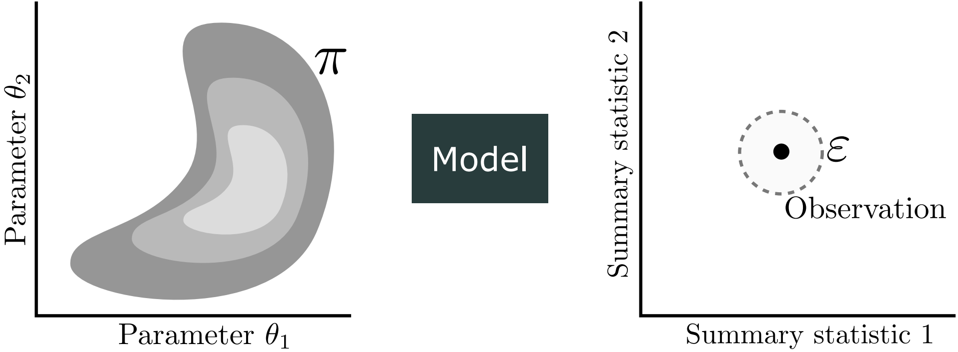
Rejection ABC

Principle of Approximate Bayesian Computation



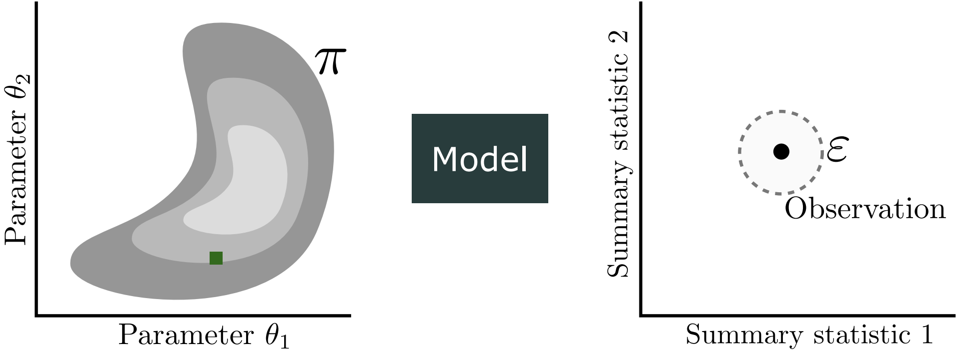
Rejection ABC

Principle of Approximate Bayesian Computation



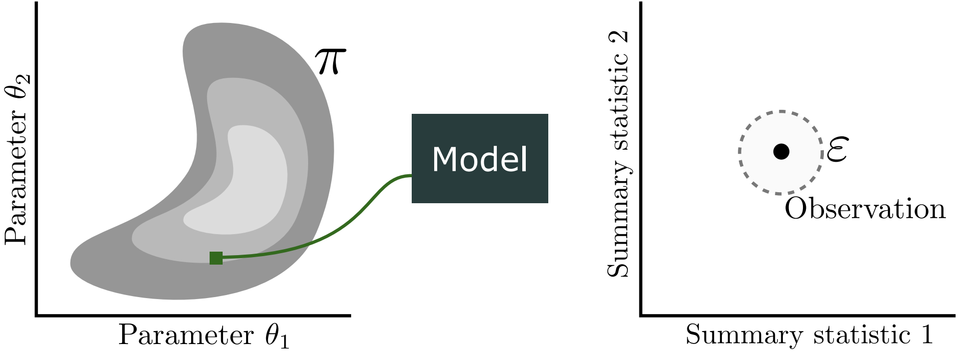
Rejection ABC

Principle of Approximate Bayesian Computation



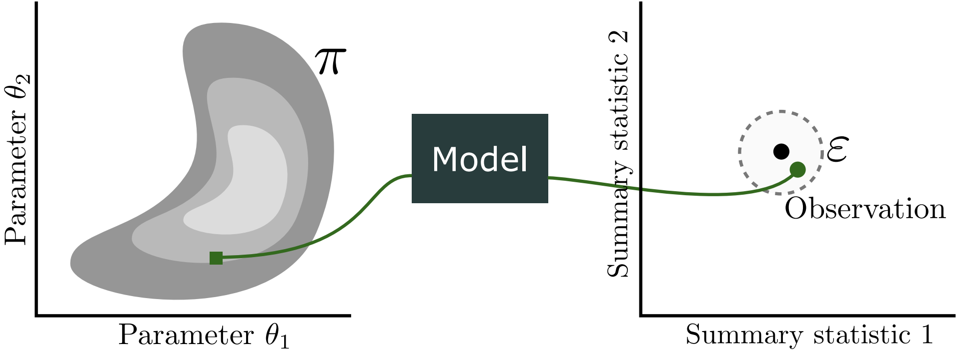
Rejection ABC

Principle of Approximate Bayesian Computation



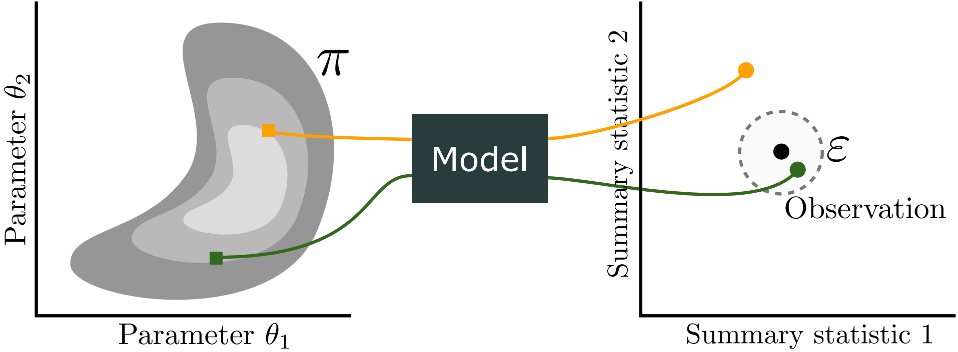
Rejection ABC

Principle of Approximate Bayesian Computation



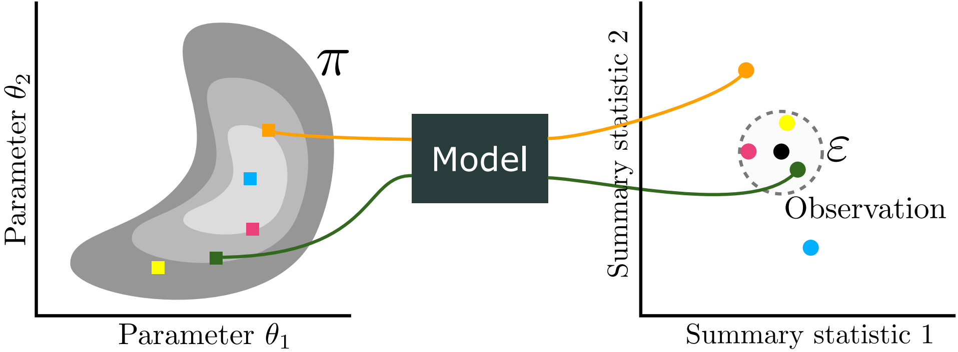
Rejection ABC

Principle of Approximate Bayesian Computation



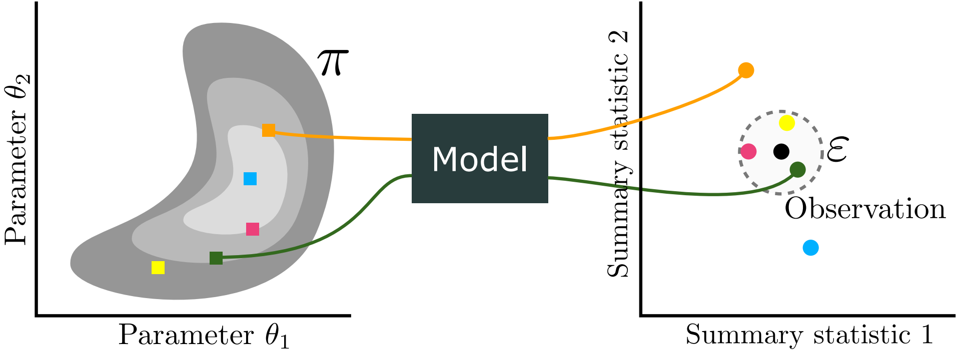
Rejection ABC

Principle of Approximate Bayesian Computation



Rejection ABC

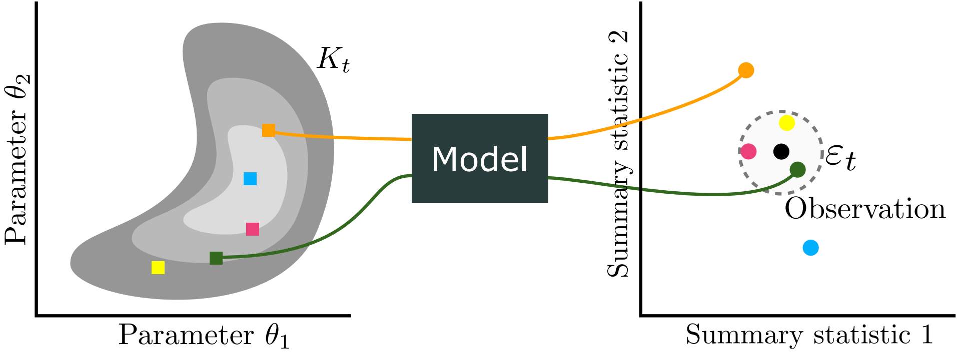
Principle of Approximate Bayesian Computation



- Approximation error
- Inefficient

ABC-SMC[1]

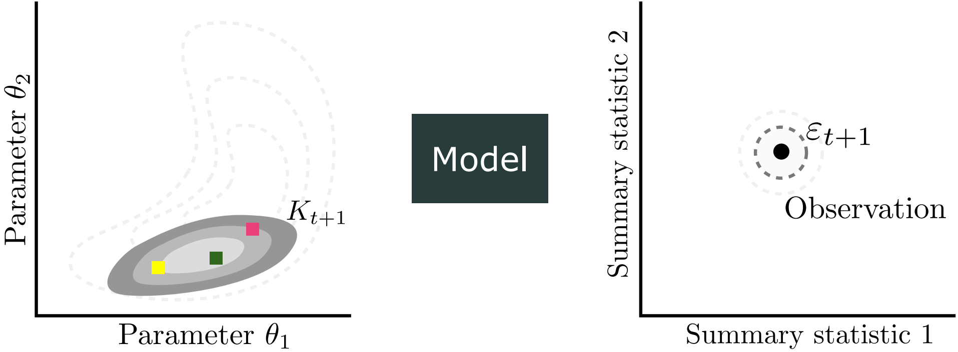
Combine with a Sequential Monte Carlo scheme



1 [Toni, Stumpf, *ABC scheme for parameter inference and model selection in dynamical systems*, J. R. Soc., 2009]

ABC-SMC[1]

Combine with a Sequential Monte Carlo scheme



[1] [Toni, Stumpf, *ABC scheme for parameter inference and model selection in dynamical systems*, J. R. Soc., 2009]

pyABC[1]

https://github.com/icb-dcm/pyabc

user-friendly

scalable

flexible

[1] [Klinger, Rickert, Hasenauer, *pyABC: distributed, likelihood-free inference*, Bioinformatics, 2018]

user-friendly

scalable

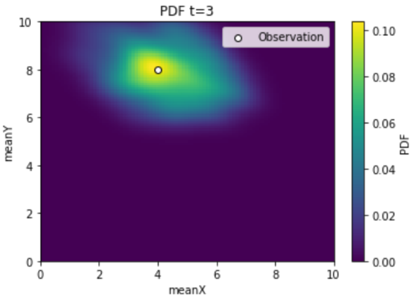
flexible

Three lines get you started

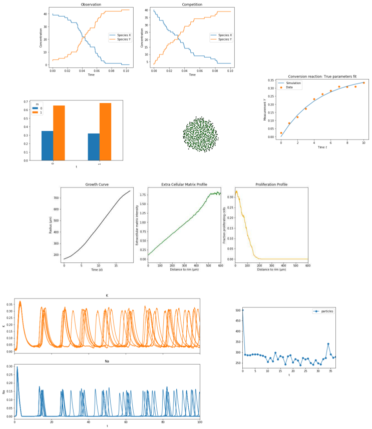
```
# model data
abc = pyabc.ABCSMC(model, prior, distance)
```

```
# observations
abc.new("sqlite:///database.db", observations)
```

```
# run the abc
abc.run(minimum_epsilon=1e-2, max_nr_populations=30)
```



If you need help ...



... 10+ application examples

... extensive documentation

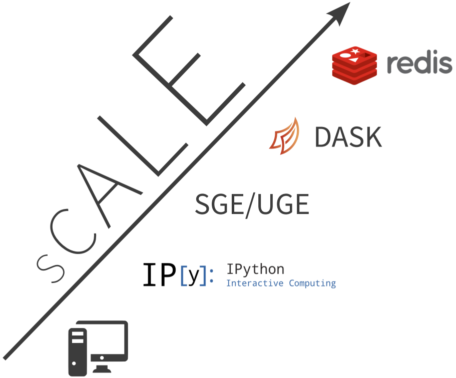
... ask us

user-friendly

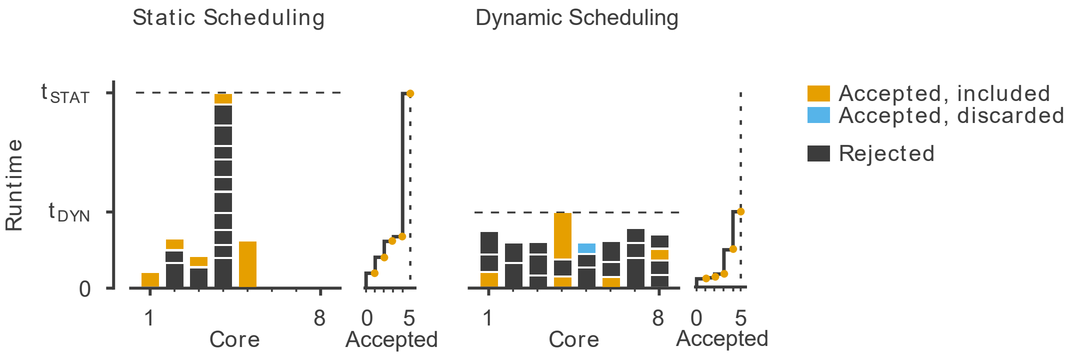
scalable

flexible

Parallel backends: 1 to 1000s of cores



Parallelization strategies[1]



1 [Klinger, Hasenauer, *A scheme for adaptive selection of population sizes in ABC-SMC*, CMSB Proceedings, 2017]

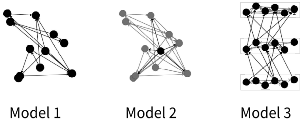
user-friendly

scalable

flexible

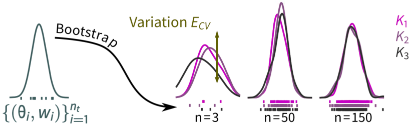
Many customization possibilities

- Model selection



Many customization possibilities

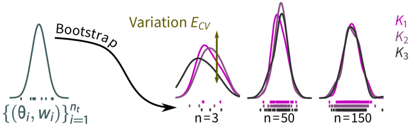
- Model selection
- Adaptive population sizes



$$E_{CV} = \sum_{i=1}^{n_t} w_i \cdot \text{Coefficient of Variation}(\{K_1(\theta_i), K_2(\theta_i), K_3(\theta_i), \ldots\})$$

Many customization possibilities

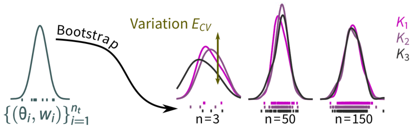
- Model selection
- Adaptive population sizes
- Adaptive acceptance thresholds



$$E_{CV} = \sum_{i=1}^{n_t} w_i \cdot \text{Coefficient of Variation}(\{K_1(\theta_i), K_2(\theta_i), K_3(\theta_i), \dots\})$$

Many customization possibilities

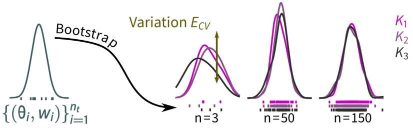
- Model selection
- Adaptive population sizes
- Adaptive acceptance thresholds
- Adaptive distance functions



$$E_{CV} = \sum_{i=1}^{n_t} w_i \cdot \text{Coefficient of Variation}(\{K_1(\theta_i), K_2(\theta_i), K_3(\theta_i), \ldots\})$$

Many customization possibilities

- Model selection
- Adaptive population sizes
- Adaptive acceptance thresholds
- Adaptive distance functions
- Global and local transition kernels
- Early rejection
- Noise assessment

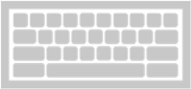


$$E_{CV} = \sum_{i=1}^{n_t} w_i \cdot \text{Coefficient of Variation}(\{K_1(\theta_i), K_2(\theta_i), K_3(\theta_i), \ldots\})$$

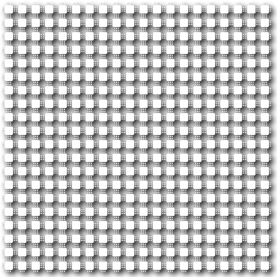
Many customization possibilities

- Model selection
- Adaptive population sizes
- Adaptive acceptance thresholds
- Adaptive distance functions
- Global and local transition kernels
- Early rejection
- Noise assessment

- Implement your own

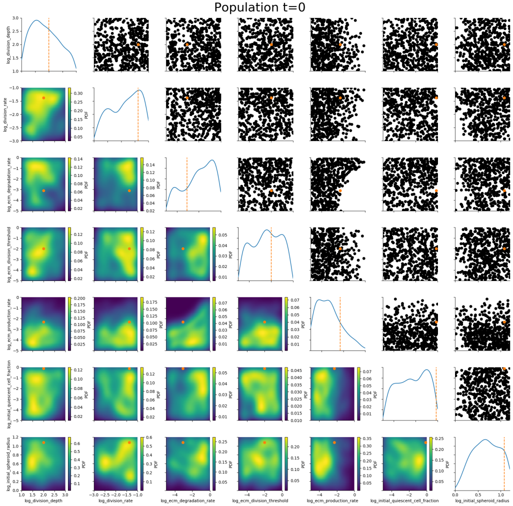


Example: Multi-scale model of tumor growth

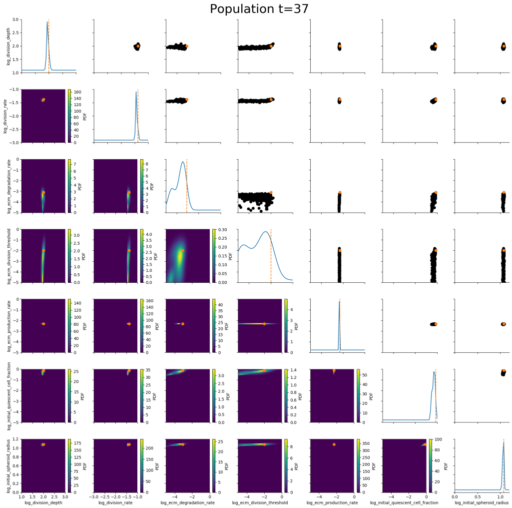


- 400 cores
- 3 days
- 1.883.664 simulations

Example: Multi-scale model of tumor growth



Example: Multi-scale model of tumor growth



Summary

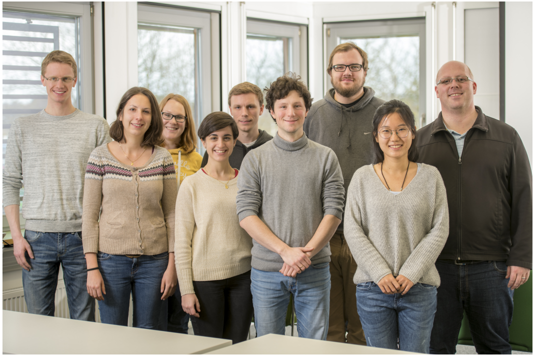
- Parameter estimation when we cannot evaluate the likelihood is challenging
- ABC allows for reliable statistical inference
- pyABC provides a user-friendly, scalable, and flexible framework
- Already used in multiple applications

`https://github.com/icb-dcm/pyabc`

Acknowledgments

Thanks to:

- Elba Raimúndez Álvarez
- Emmanuel Klinger
- Dennis Rickert
- Jan Hasenauer
- Rest of the ICB-DCM group



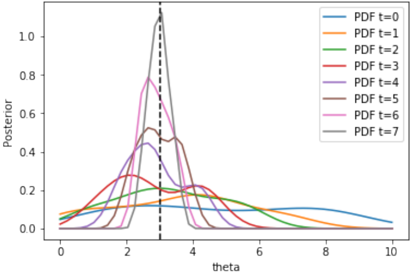
HelmholtzZentrum münchen
German Research Center for Environmental Health

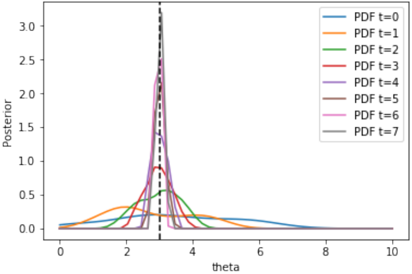
Bundesministerium
für Bildung
und Forschung

e:Med
SYSTEMS MEDICINE

And ...

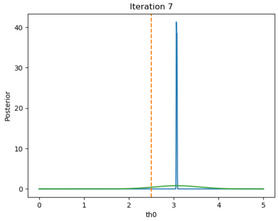
... adaptive distance functions

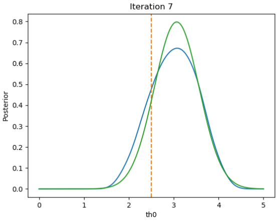


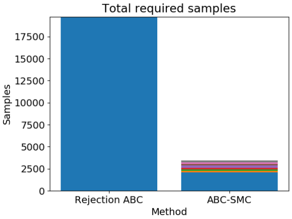


And ...

... noise assessment

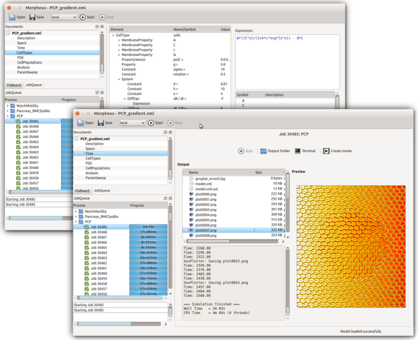






And ...

... model construction using Morpheus[1]



[1] [Staruss et al., *Morpheus: a user-friendly modeling environment for multiscale and multicellular systems biology*, Bioinformatics, 2014]

And ...

... likelihood-free Bayesian inference

$$\pi(\theta|D) = \frac{\pi(D|\theta)\pi(\theta)}{\pi(D)}$$