



@yannik_schaelte



yannikschaelte

HelmholtzZentrum münchen
German Research Center for Environmental Health

pyABC: Statistical inference without a likelihood?

Yannik Schälte

OpenSourceEconomics Meetup, 2020-11-10



Helmholtz Zentrum München, Institute of Computational Biology
Technische Universität München, Department of Mathematics

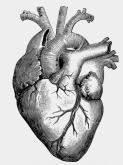
Overview

- Motivation for likelihood-free inference
- Basics of ABC
- pyABC: Features and usage

Why?

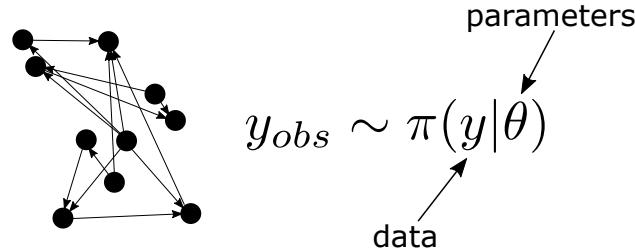
Parameter inference

dynamical system



Parameter inference

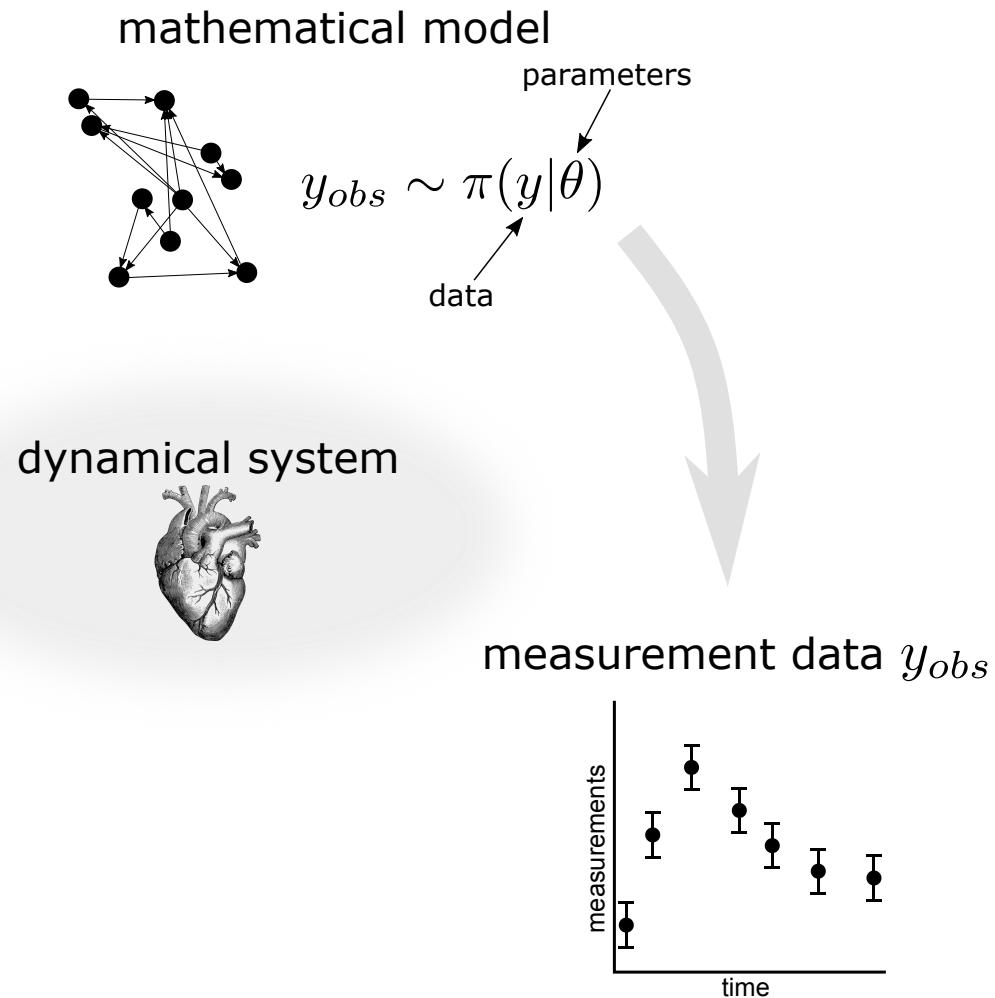
mathematical model



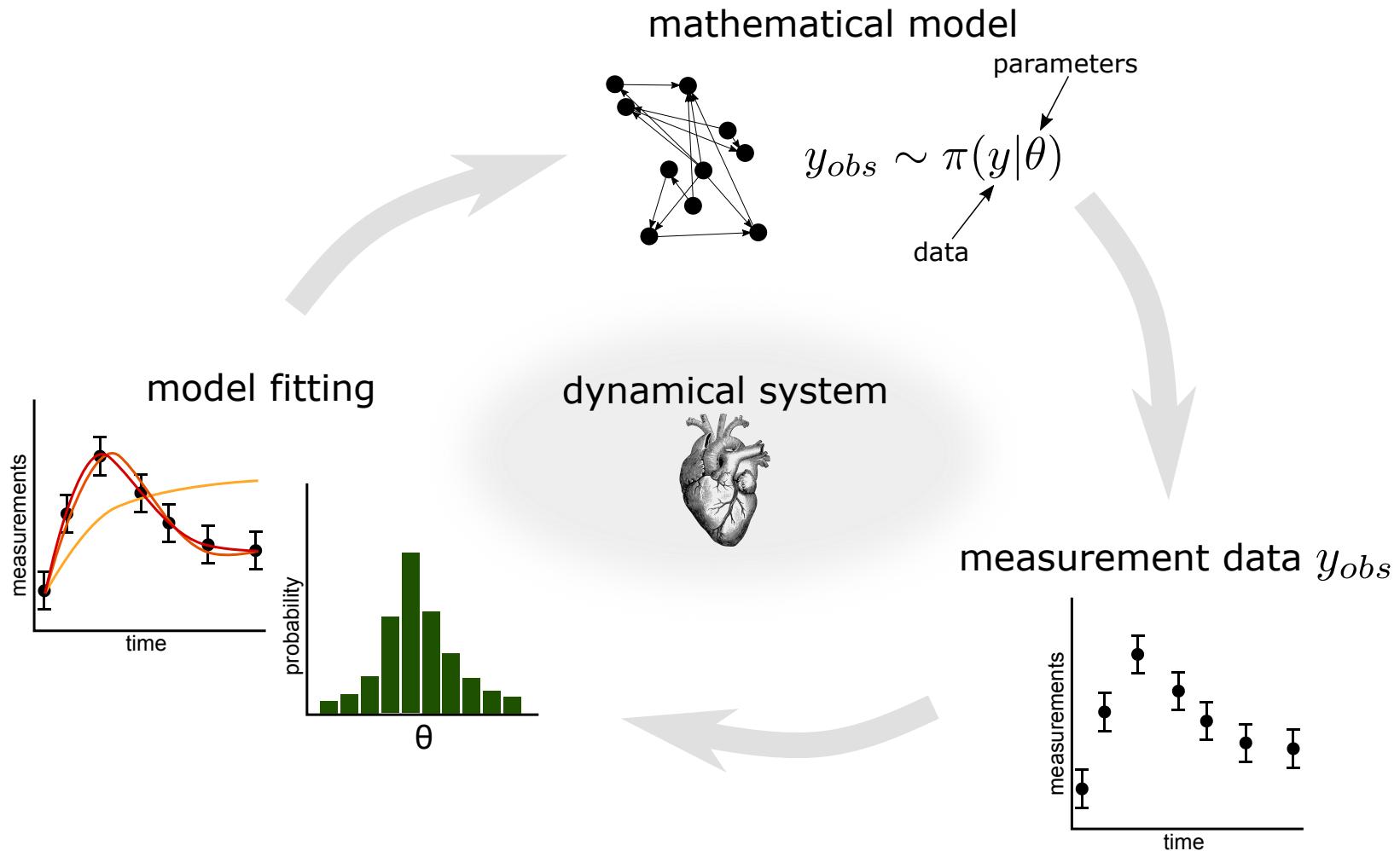
dynamical system



Parameter inference



Parameter inference



(Likelihood-free) Bayesian inference

- goal: infer parameters θ given data y_{obs} , i.e. analyze the posterior distribution

(Likelihood-free) Bayesian inference

$$\pi(\theta|y_{obs}) = \frac{\pi(y_{obs}|\theta)\pi(\theta)}{\pi(y_{obs})}$$

posterior likelihood prior
evidence

- goal: infer parameters θ given data y_{obs} , i.e. analyze the posterior distribution

(Likelihood-free) Bayesian inference

$$\pi(\theta|y_{obs}) = \frac{\pi(y_{obs}|\theta)\pi(\theta)}{\pi(y_{obs})}$$

posterior → ↗ likelihood ← prior
 ↑
 evidence

- goal: infer parameters θ given data y_{obs} , i.e. analyze the posterior distribution

(Likelihood-free) Bayesian inference

$$\pi(\theta|y_{obs}) = \frac{\pi(y_{obs}|\theta)\pi(\theta)}{\pi(y_{obs})}$$

posterior → ↗ likelihood ← prior
 ↑
 evidence

- goal: infer parameters θ given data y_{obs} , i.e. analyze the posterior distribution
- common optimization and sampling methods require the (unnormalized) likelihood

(Likelihood-free) Bayesian inference

$$\pi(\theta|y_{obs}) = \frac{\pi(y_{obs}|\theta)\pi(\theta)}{\pi(y_{obs})}$$

posterior

prior

likelihood

evidence

- goal: infer parameters θ given data y_{obs} , i.e. analyze the posterior distribution
 - common optimization and sampling methods require the (unnormalized) likelihood
 - can happen: numerical evaluation infeasible

(Likelihood-free) Bayesian inference

$$\pi(\theta|y_{obs}) = \frac{\pi(y_{obs}|\theta)\pi(\theta)}{\pi(y_{obs})}$$

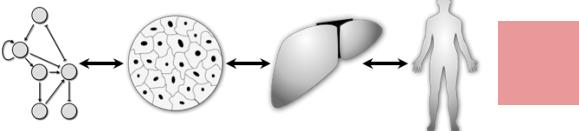
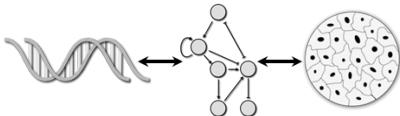
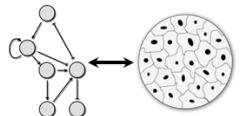
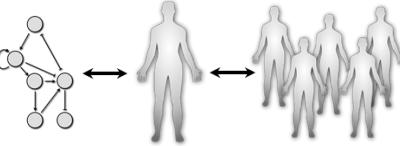
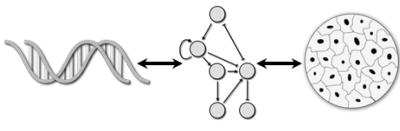
Diagram illustrating the components of Bayes' Theorem:

- posterior**: $\pi(\theta|y_{obs})$
- likelihood**: $\pi(y_{obs}|\theta)$
- prior**: $\pi(\theta)$
- evidence**: $\pi(y_{obs})$

Arrows point from the words to their corresponding terms in the formula.

- goal: infer parameters θ given data y_{obs} , i.e. analyze the posterior distribution
 - common optimization and sampling methods require the (unnormalized) likelihood
 - can happen: numerical **evaluation infeasible**
 - ... but **possible to simulate data** $y \sim \pi(y|\theta)$

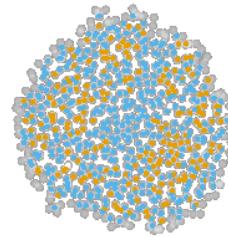
Multi-scale models

	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, <i>Nature Reviews Molecular Cell Biology</i> , 4(3):237-243, 2003	
cancer growth (signaling, cell division and death, angiogenesis, tissue remodeling)							Anderson and Quaranta, <i>Nature Reviews Cancer</i> , 8(3):227-234. 2008	
liver lobule (cell division and cell death, tissue mechanics)								Hoehme et al., <i>PNAS</i> , 107(23):10371-10376, 2010
glucose-insulin-glucagon regulation (blood and interstitial flow, organ uptake, signaling)							Schaller et al., <i>CPT: Pharmacometrics and Systems Pharmacology</i> , 2:e65, 2013	
whole-cell (transcription, translation, DNA replication, metabolism, replication)							Karr et al., <i>Cell</i> , 150(2):389-401, 2012	

based on Hasenauer et al., J. Coup. Sys. and Mult. Dyn. 2015

Example: Modeling tumor growth

based on Jagiella et al., Cell Systems 2017



- cells modeled as interacting stochastic agents, dynamics of extracellular substances by PDEs
- simulate up to 10^6 cells
- 10s - 1h for one forward simulation
- 7-18 parameters

What we tried

- multi-start local methods
 - deterministic gradient descent
 - Levenberg-Marquardt
 - interior-point
 - trust-region
 - stochastic gradient descent
 - Bayesian optimization
- global methods
 - simulated annealing
 - > 20 particle methods
 - enhanced scatter search

Failed

Worked

-
- multi-start local methods
 - deterministic gradient descent
 - Levenberg-Marquardt
 - interior-point
 - trust-region
 - stochastic gradient descent
 - Bayesian optimization
 - global methods
 - simulated annealing
 - > 20 particle methods
 - enhanced scatter search

How to infer parameters for complex stochastic models?

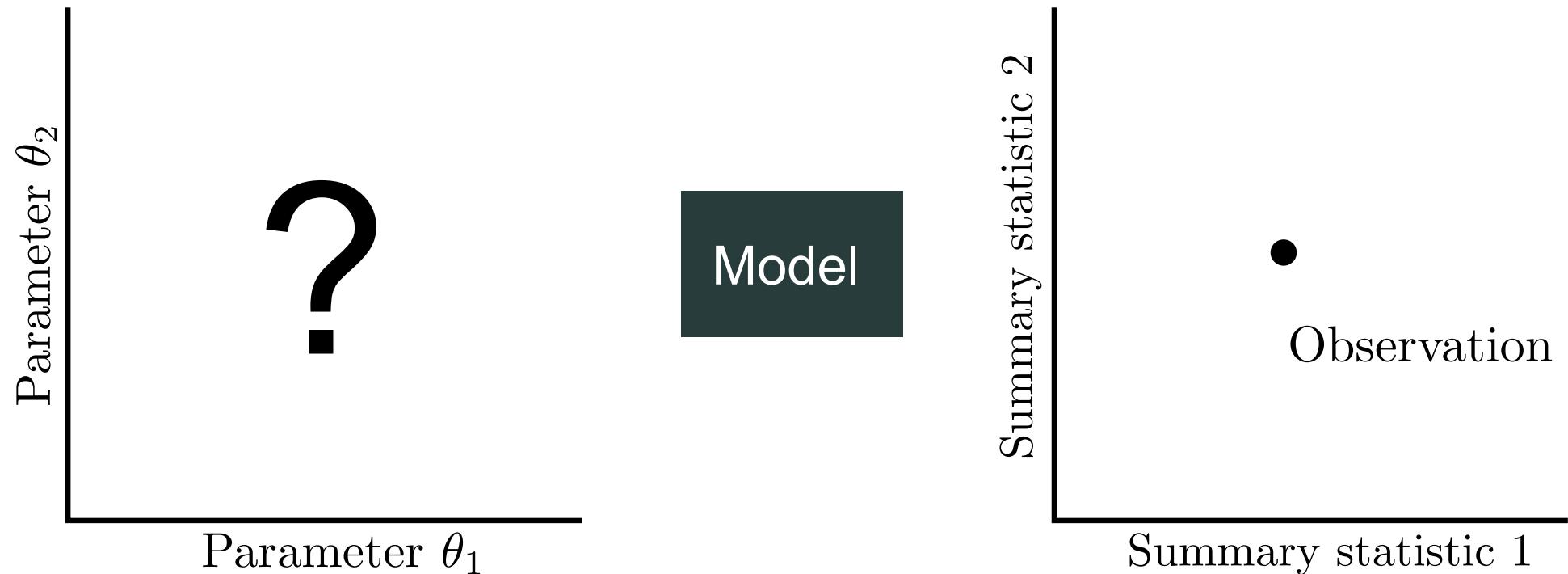
ABC

Approximate Bayesian Computation

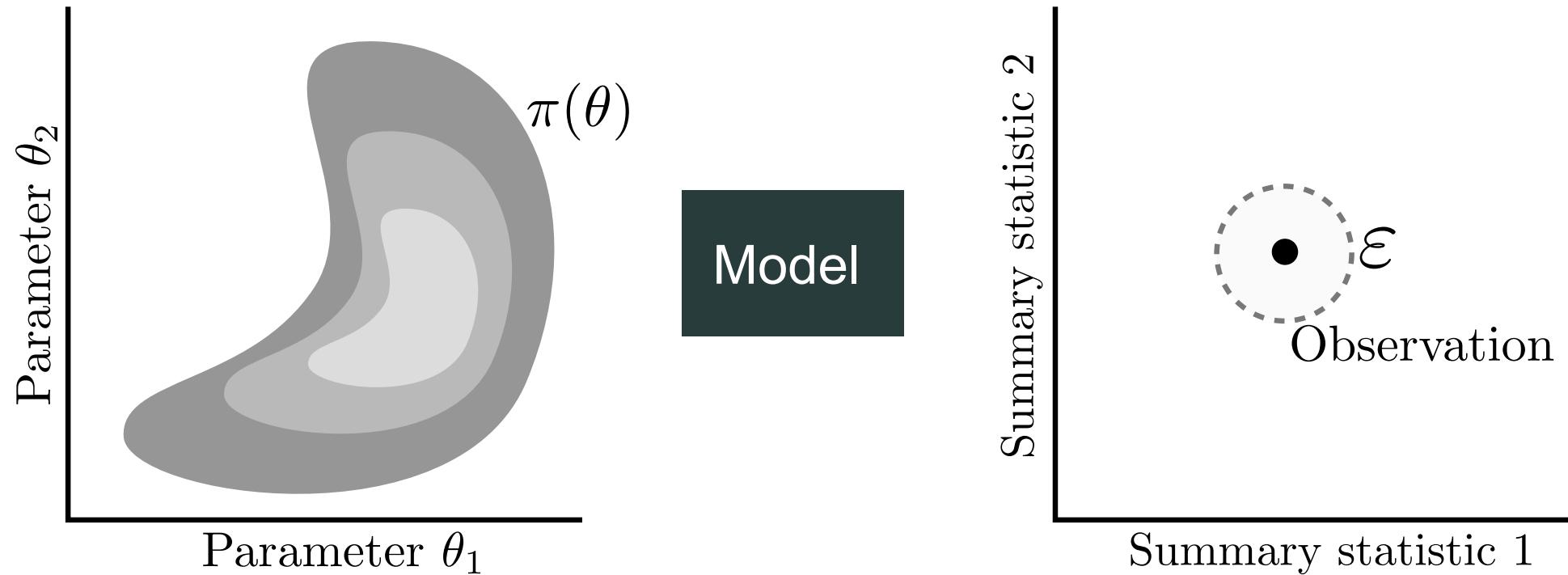
Rejection ABC

Model

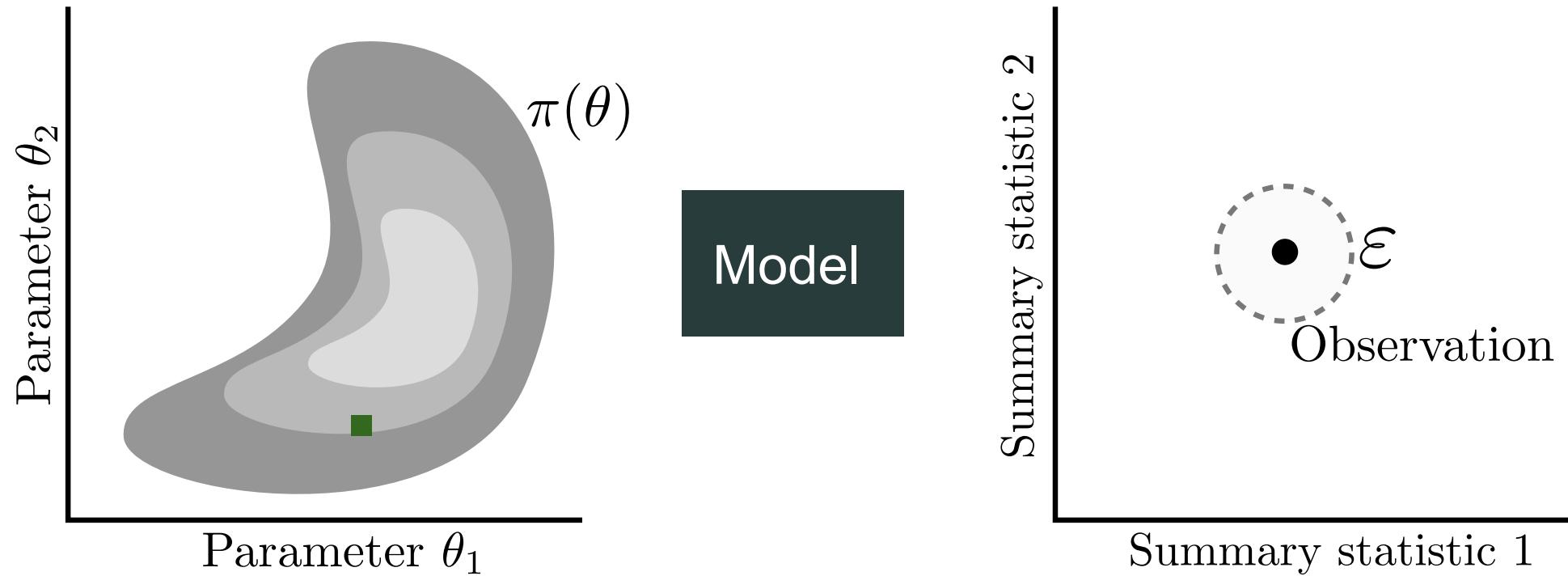
Rejection ABC



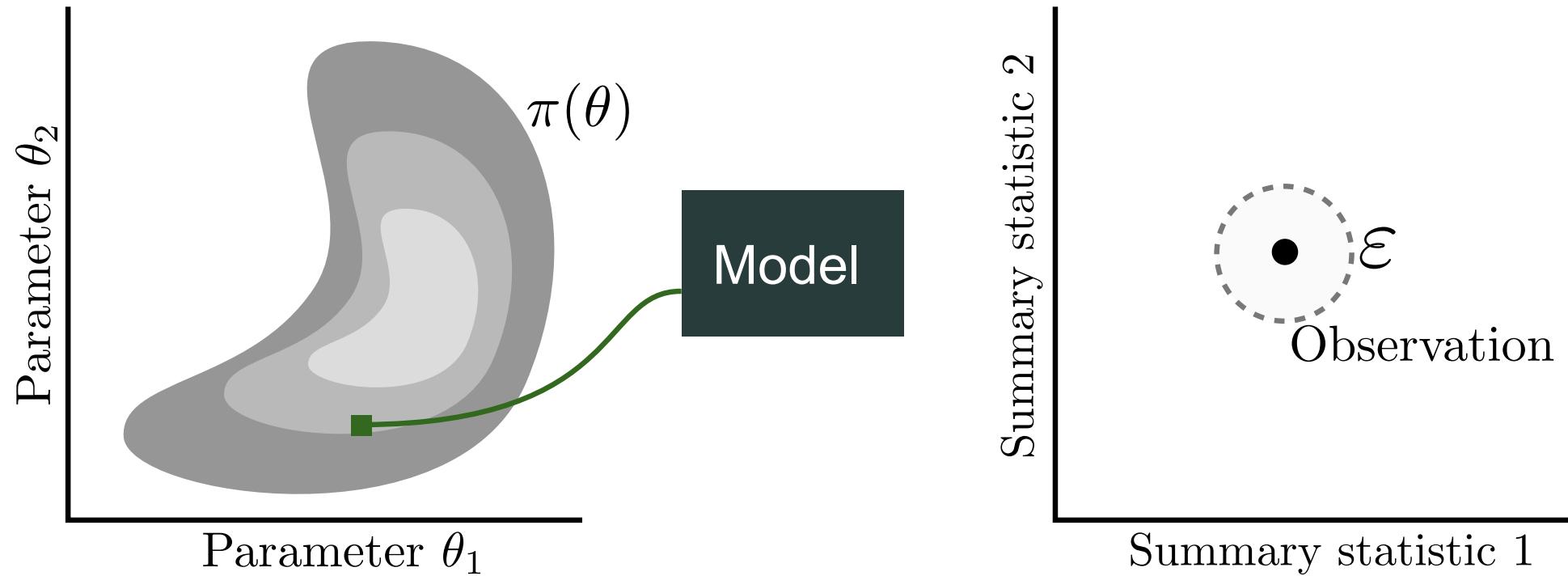
Rejection ABC



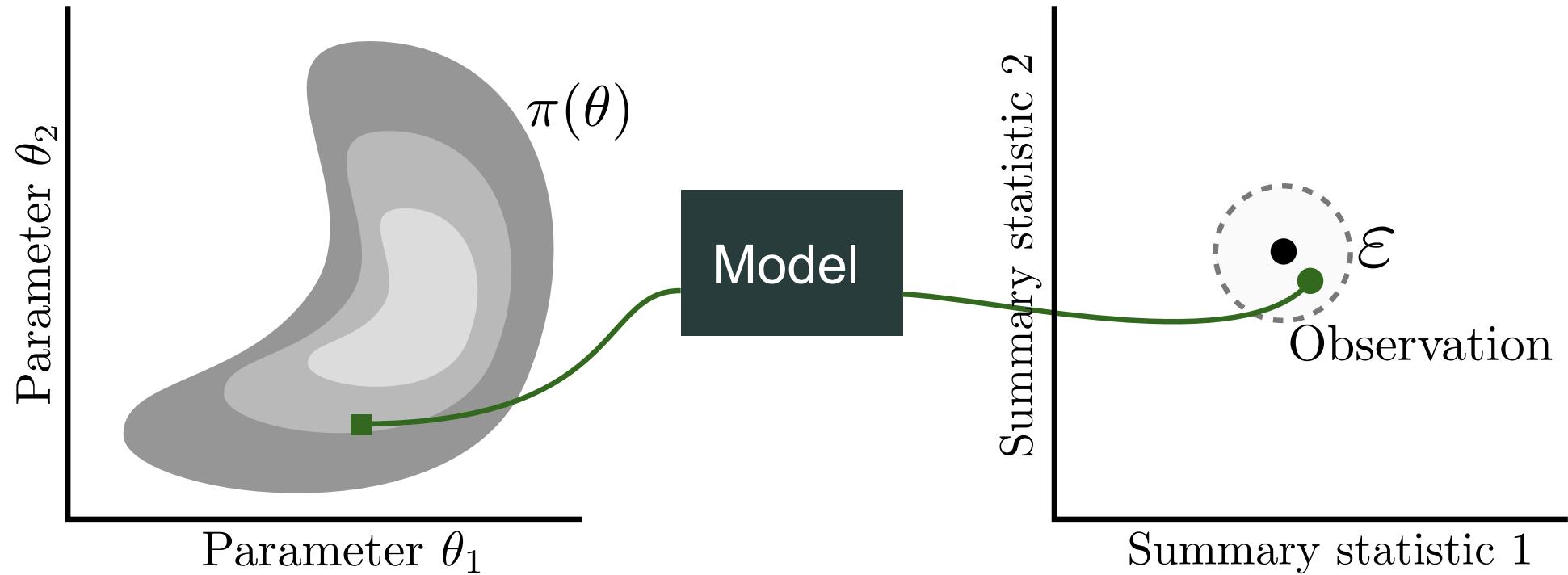
Rejection ABC



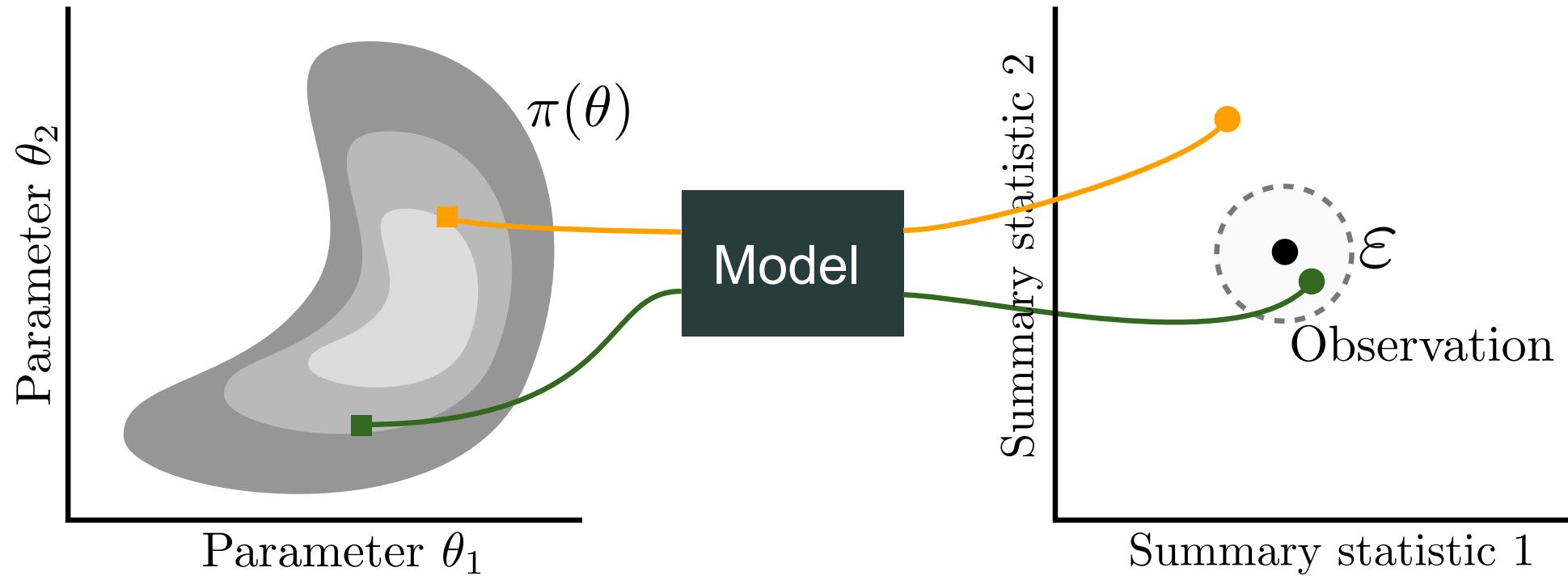
Rejection ABC



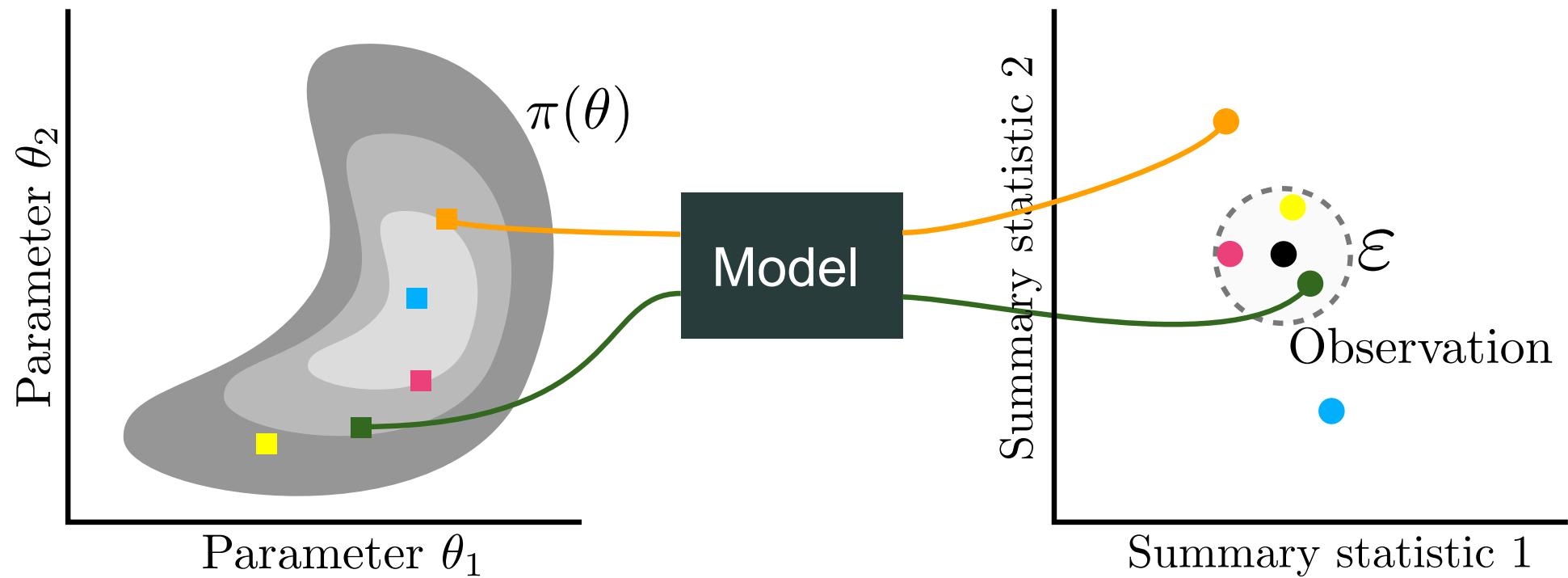
Rejection ABC



Rejection ABC



Rejection ABC



Rejection ABC

With distance d , threshold $\varepsilon > 0$, summary statistics s :

until N acceptances:

1. sample $\theta \sim g(\theta)$
2. simulate data $y \sim \pi(y|\theta)$
3. accept θ if $d(s(y), s(y_{\text{obs}})) \leq \varepsilon$

A "derivation"

Rejection sampling

Background: Want to sample from f , but can only sample from $g \gg f$.

until N acceptances:

1. sample $\theta \sim g(\theta)$
2. accept θ with probability $\propto \frac{f(\theta)}{g(\theta)}$

Accepted θ are independent samples from $f(\theta)$.

Let $f = \pi(\theta|y_{\text{obs}})$, $g = \pi(\theta) \Rightarrow \frac{\pi(\theta|y_{\text{obs}})}{\pi(\theta)} \propto \pi(y_{\text{obs}}|\theta)$

- not available
- idea: **circumvent likelihood evaluation by simulating data** and matching them to the observed data

Likelihood-free rejection sampling

until N acceptances:

1. sample $\theta \sim \pi(\theta)$
2. simulate data $y \sim \pi(y|\theta)$
3. accept θ if $y = y_{\text{obs}}$

- Acceptance probability: $\mathbb{P}[y_{\text{obs}}]$
- can be small in particular for continuous data
- idea: accept simulations that are **similar** to y_{obs}

Rejection ABC

With distance d , threshold $\varepsilon > 0$:

until N acceptances:

1. sample $\theta \sim \pi(\theta)$
2. simulate data $y \sim \pi(y|\theta)$
3. accept θ if $d(y, y_{\text{obs}}) \leq \varepsilon$

- **curse of dimensionality:** if the data are too high-dimensional, the probability of simulating similar data sets is small
- idea: create an informative lower-dimensional representation via **summary statistics**
- ideally minimal sufficient statistics

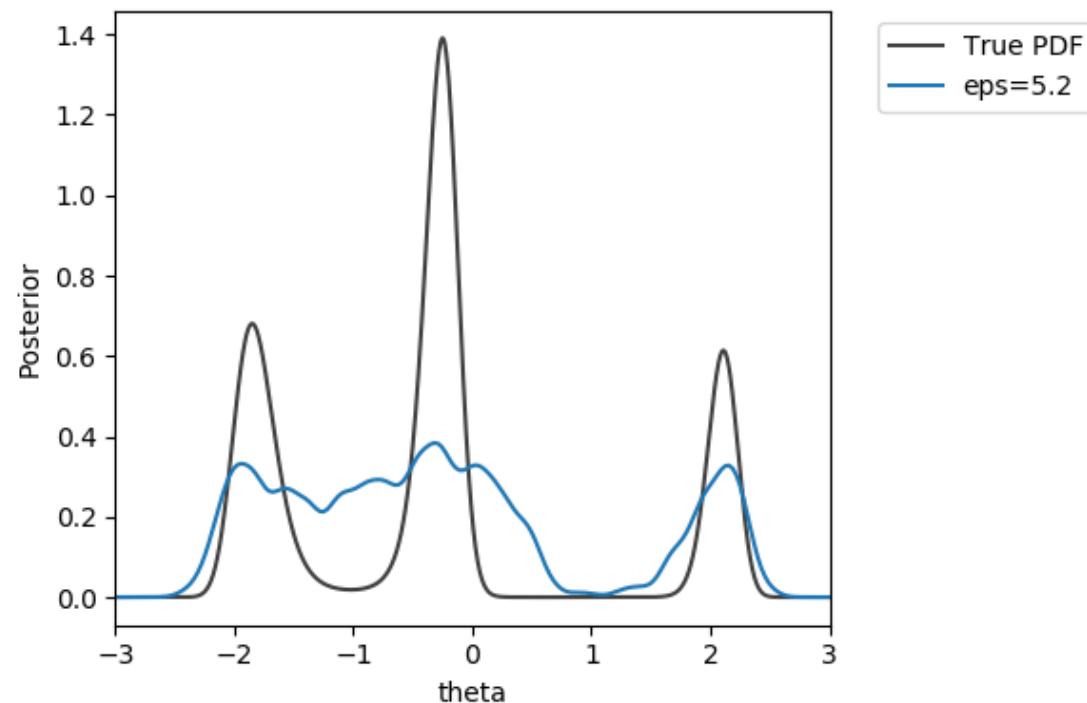
Toy example

$$y \sim \mathcal{N}(2(\theta - 2)\theta(\theta + 2), 1 + \theta^2), y_{\text{obs}} = 2$$

$$\pi(\theta) = U[-3, 3]$$

$$d = |\cdot|_1$$

$N = 1000$ acceptances



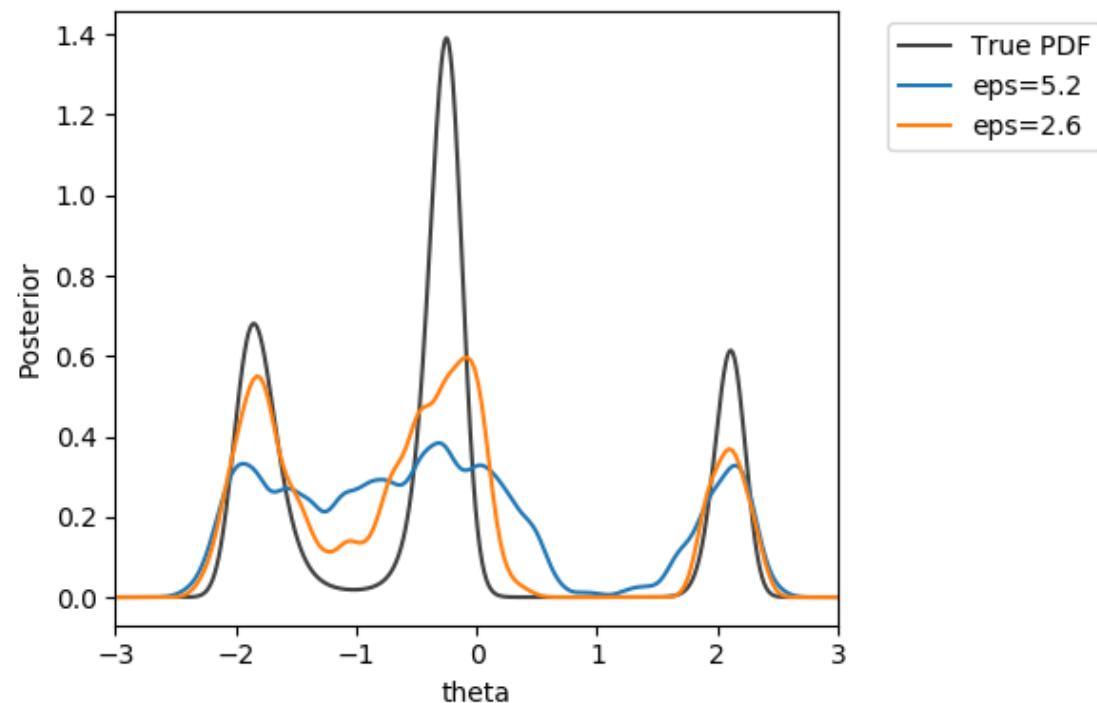
Toy example

$$y \sim \mathcal{N}(2(\theta - 2)\theta(\theta + 2), 1 + \theta^2), y_{\text{obs}} = 2$$

$$\pi(\theta) = U[-3, 3]$$

$$d = |\cdot|_1$$

$N = 1000$ acceptances



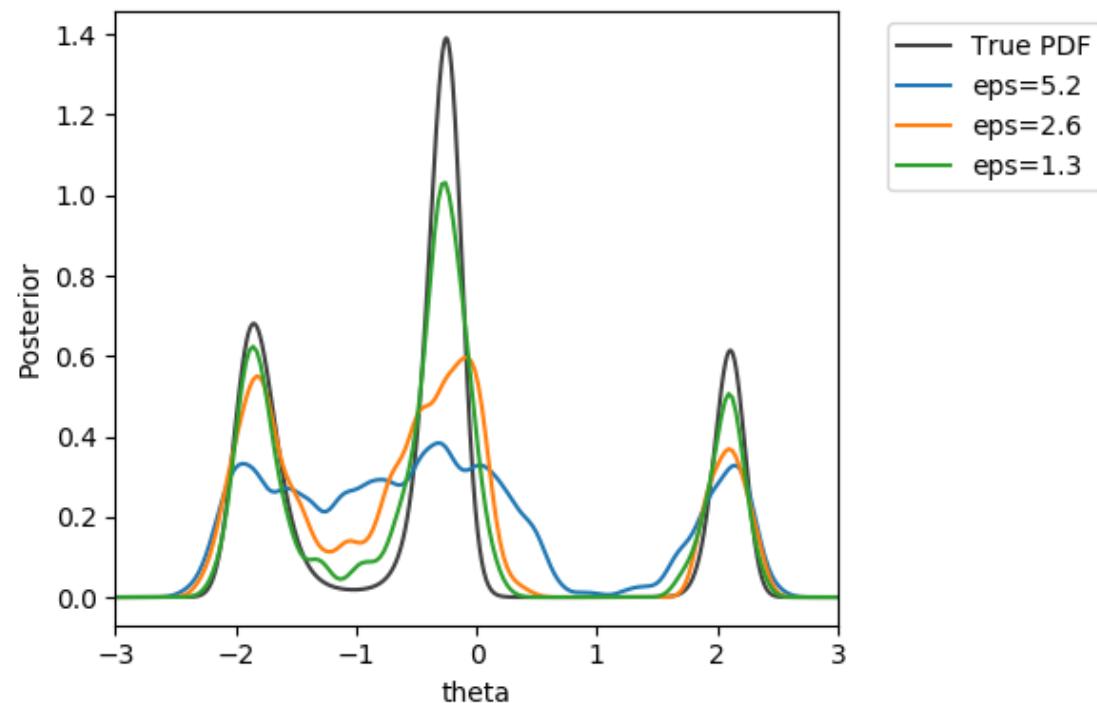
Toy example

$$y \sim \mathcal{N}(2(\theta - 2)\theta(\theta + 2), 1 + \theta^2), y_{\text{obs}} = 2$$

$$\pi(\theta) = U[-3, 3]$$

$$d = |\cdot|_1$$

$N = 1000$ acceptances



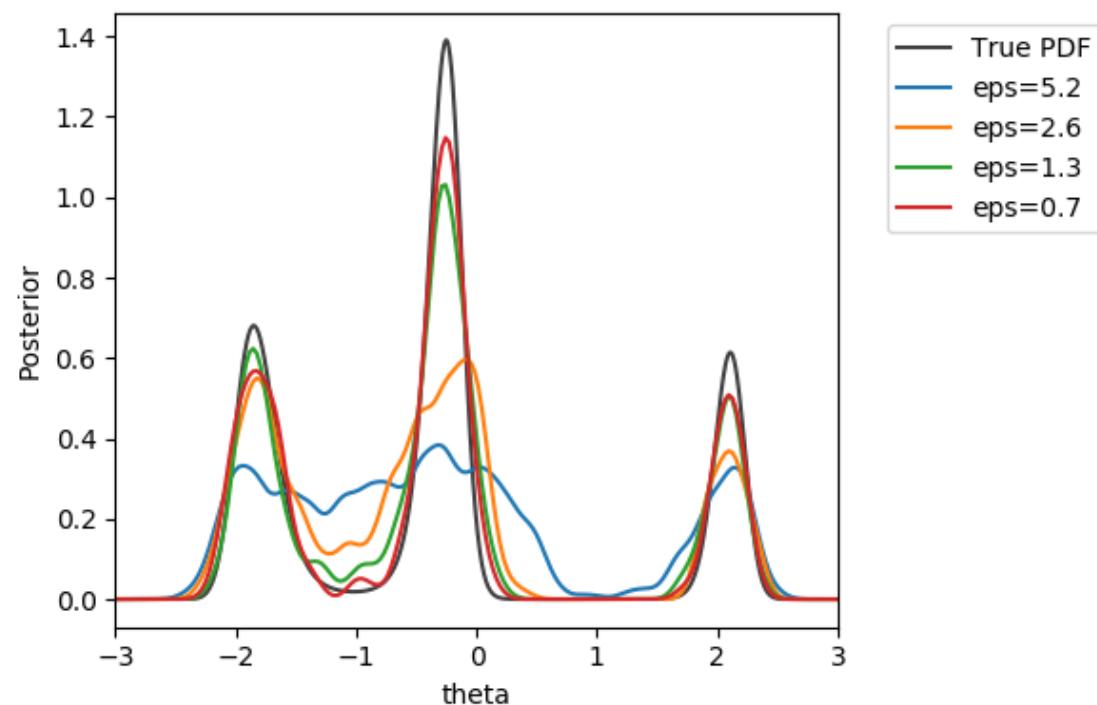
Toy example

$$y \sim \mathcal{N}(2(\theta - 2)\theta(\theta + 2), 1 + \theta^2), y_{\text{obs}} = 2$$

$$\pi(\theta) = U[-3, 3]$$

$$d = |\cdot|_1$$

$N = 1000$ acceptances



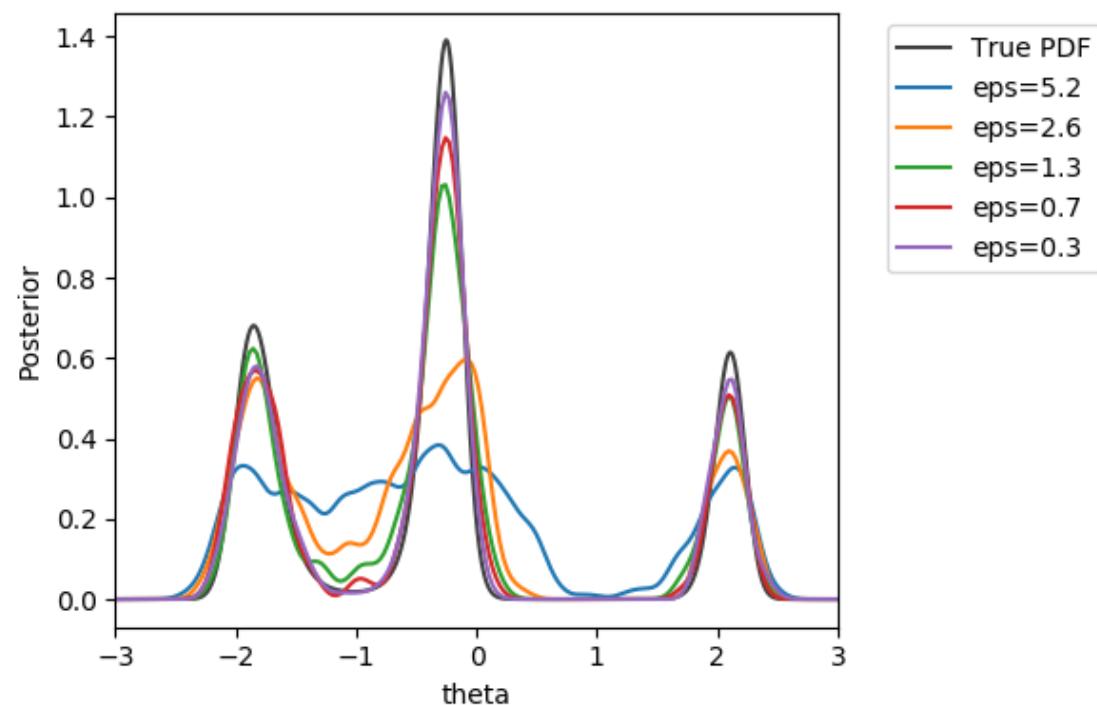
Toy example

$$y \sim \mathcal{N}(2(\theta - 2)\theta(\theta + 2), 1 + \theta^2), y_{\text{obs}} = 2$$

$$\pi(\theta) = U[-3, 3]$$

$$d = |\cdot|_1$$

$N = 1000$ acceptances



Approximate Bayesian Posterior

We want:

$$\pi(\theta|y_{\text{obs}}) \propto p(y_{\text{obs}}|\theta)\pi(\theta)$$

Approximate Bayesian Posterior

We want:

$$\pi(\theta|y_{\text{obs}}) \propto p(y_{\text{obs}}|\theta)\pi(\theta)$$

We get:

$$\pi_{ABC}(\theta|s(y_{\text{obs}})) \propto \int I(\{d(s(y), s(y_{\text{obs}})) \leq \varepsilon\})p(y|\theta) \, dy \, \pi(\theta) \approx \frac{1}{N} \sum_{i=1}^N \delta_{\theta^{(i)}}($$

with distance d , threshold $\varepsilon > 0$, and summary statistics s

Sources of approximation errors in ABC

- model error (as for every model of reality)
- Monte-Carlo error (as for sampling in general)
- summary statistics
- epsilon threshold

Sources of approximation errors in ABC

- model error (as for every model of reality)
- Monte-Carlo error (as for sampling in general)
- summary statistics
- epsilon threshold

Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise.

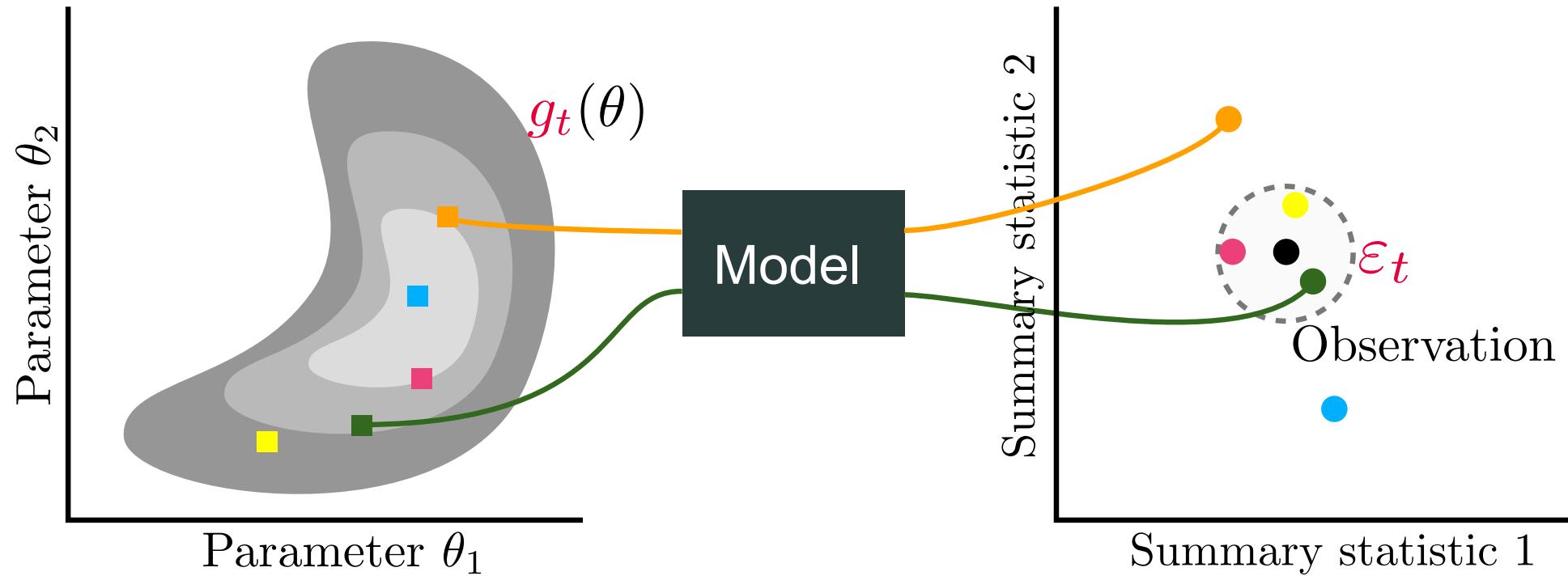
John Tukey 1962

Efficient samplers

- Rejection ABC, the basic ABC algorithm, can be **inefficient** due to repeatedly sampling from the prior
- smaller ε leads to lower acceptance rates
- many (likelihood-based) algorithms like IS, MCMC, SMC, SA have **ABC-fied** versions

ABC-SMC

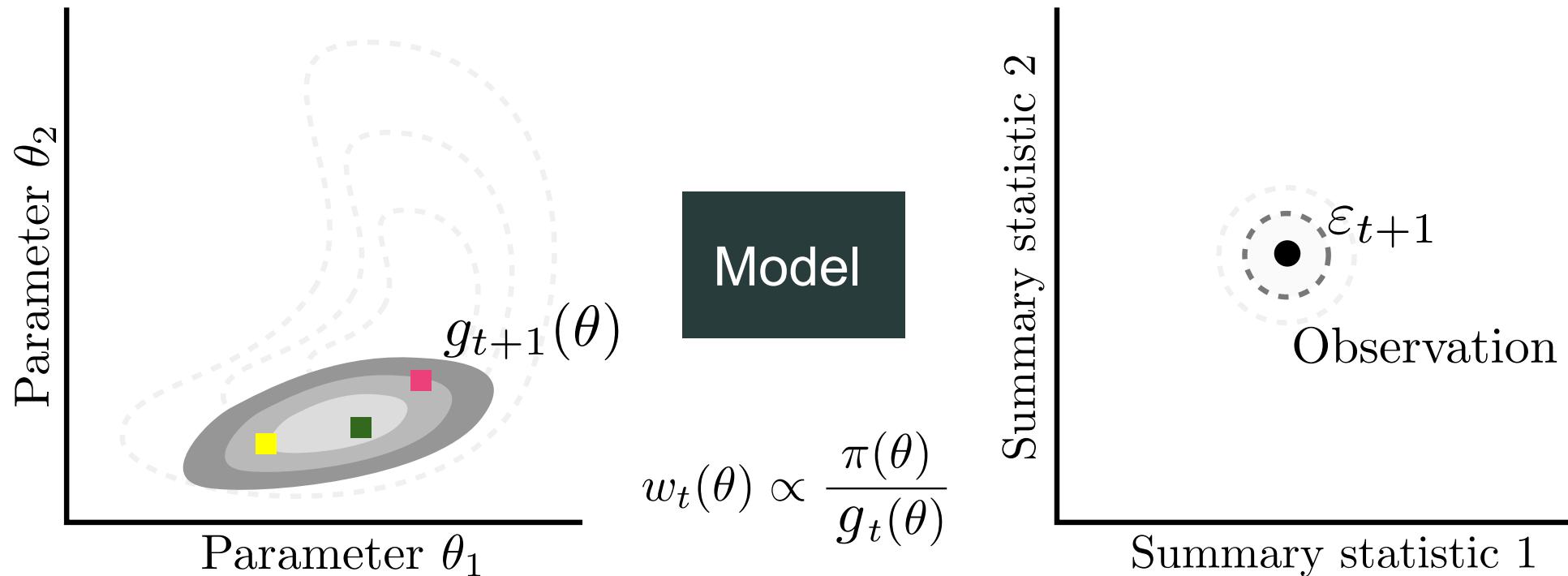
combine with a sequential Monte-Carlo scheme



similar to Toni et al., J. Royal Society 2009

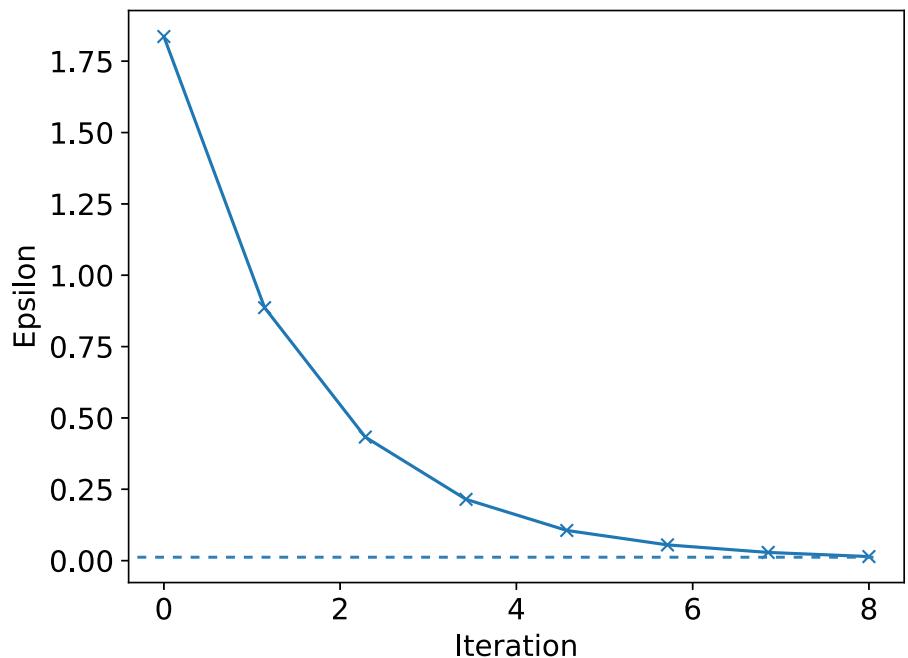
ABC-SMC

combine with a sequential Monte-Carlo scheme

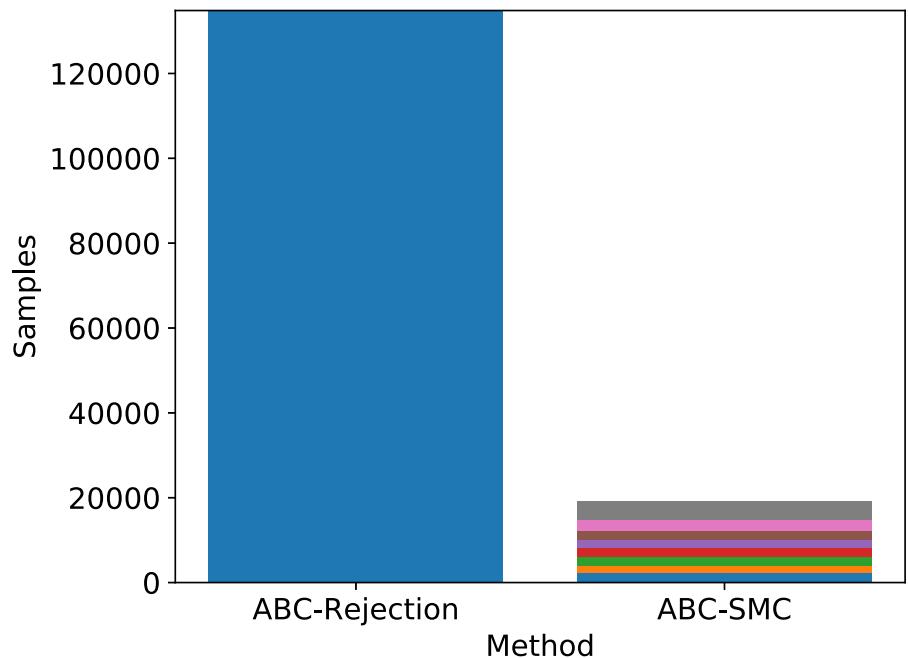


similar to Toni et al., J. Royal Society 2009

ABC-SMC epsilon reduction scheme



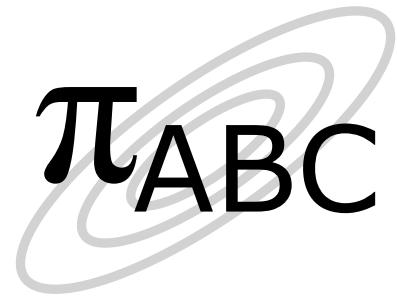
Overall number of samples





github.com/icb-dcm/pyabc

Klinger et al., Bioinformatics 2018



flexible

configure
extend



user-friendly

extensive documentation
inference in 3 lines
self-tuning



scalable

1 to 1,000s cores
seconds to weeks runtime

Easy to use

```
1 # define problem
2 abc = pyabc.ABCSMC(model, prior, distance)
3
4 # pass data
5 abc.new(database, observation)
6
7 # run it
8 abc.run()
```

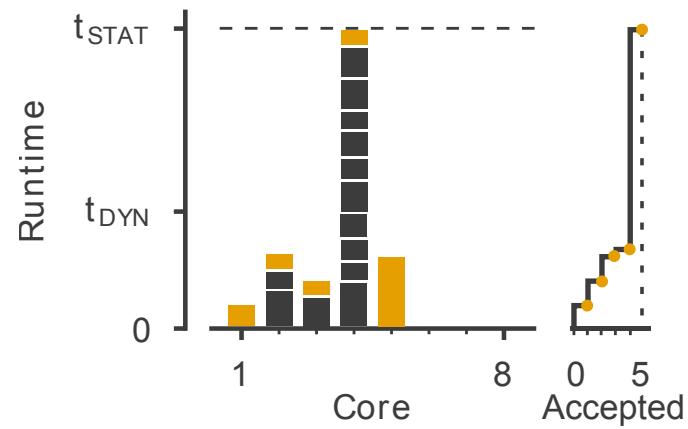
Parallel backends: 1 to 1,000s cores



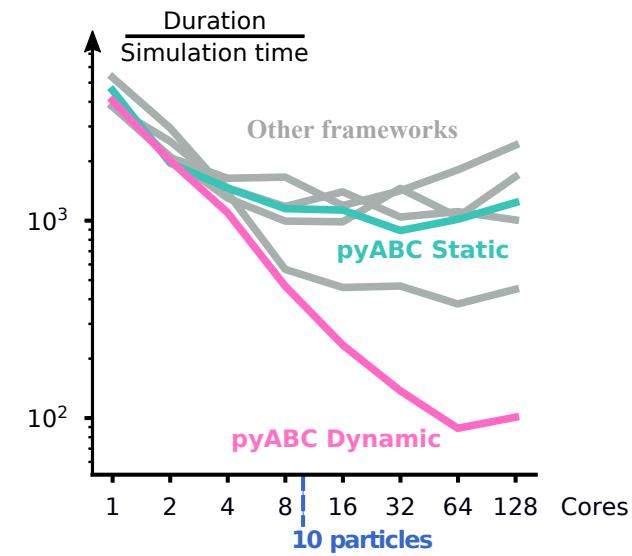
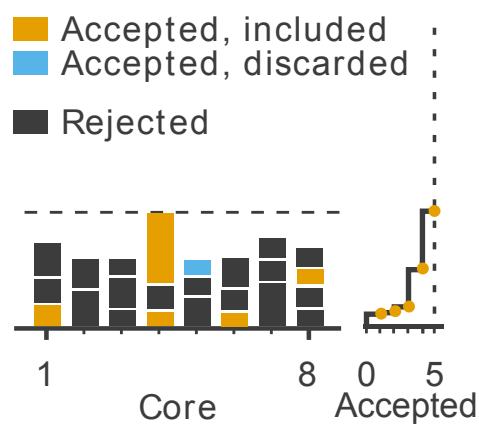
Parallelization strategies

Klinger et al., CMSB Proceedings 2017

Static Scheduling



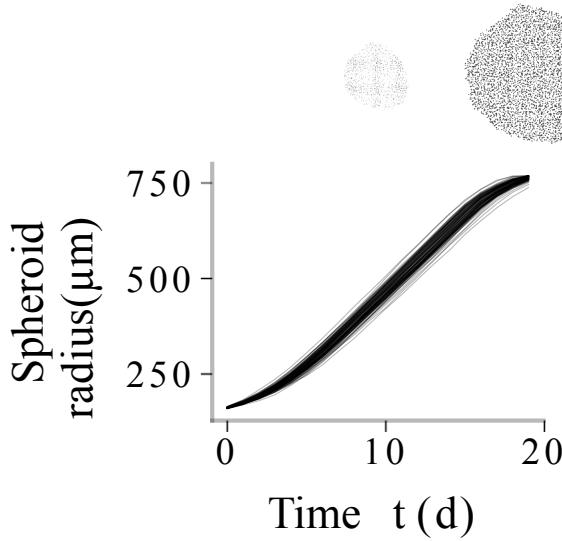
Dynamic Scheduling



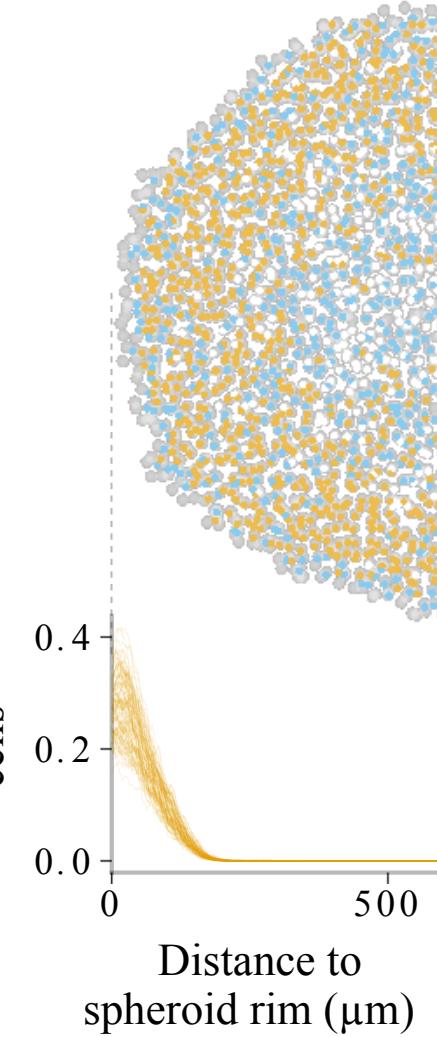
Example: Tumor growth model

based on Jagiella et al., Cell Systems 2017

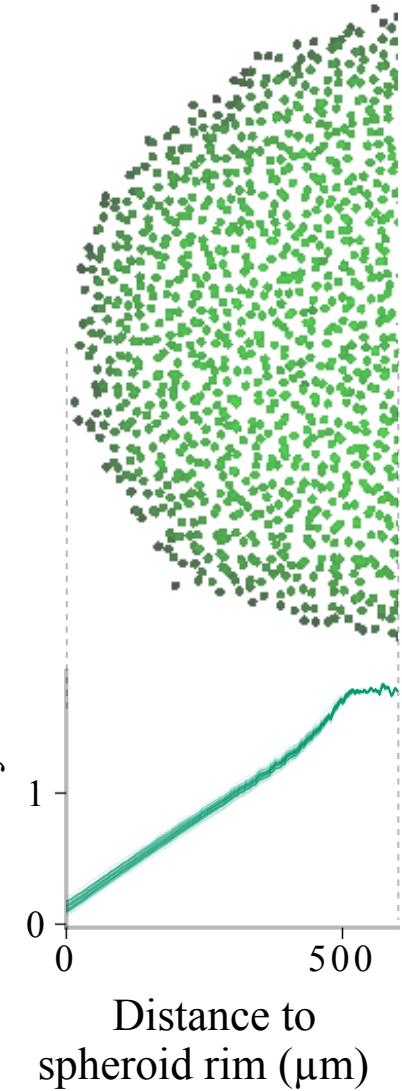
Define summary statistics



Fraction of
proliferating
cells

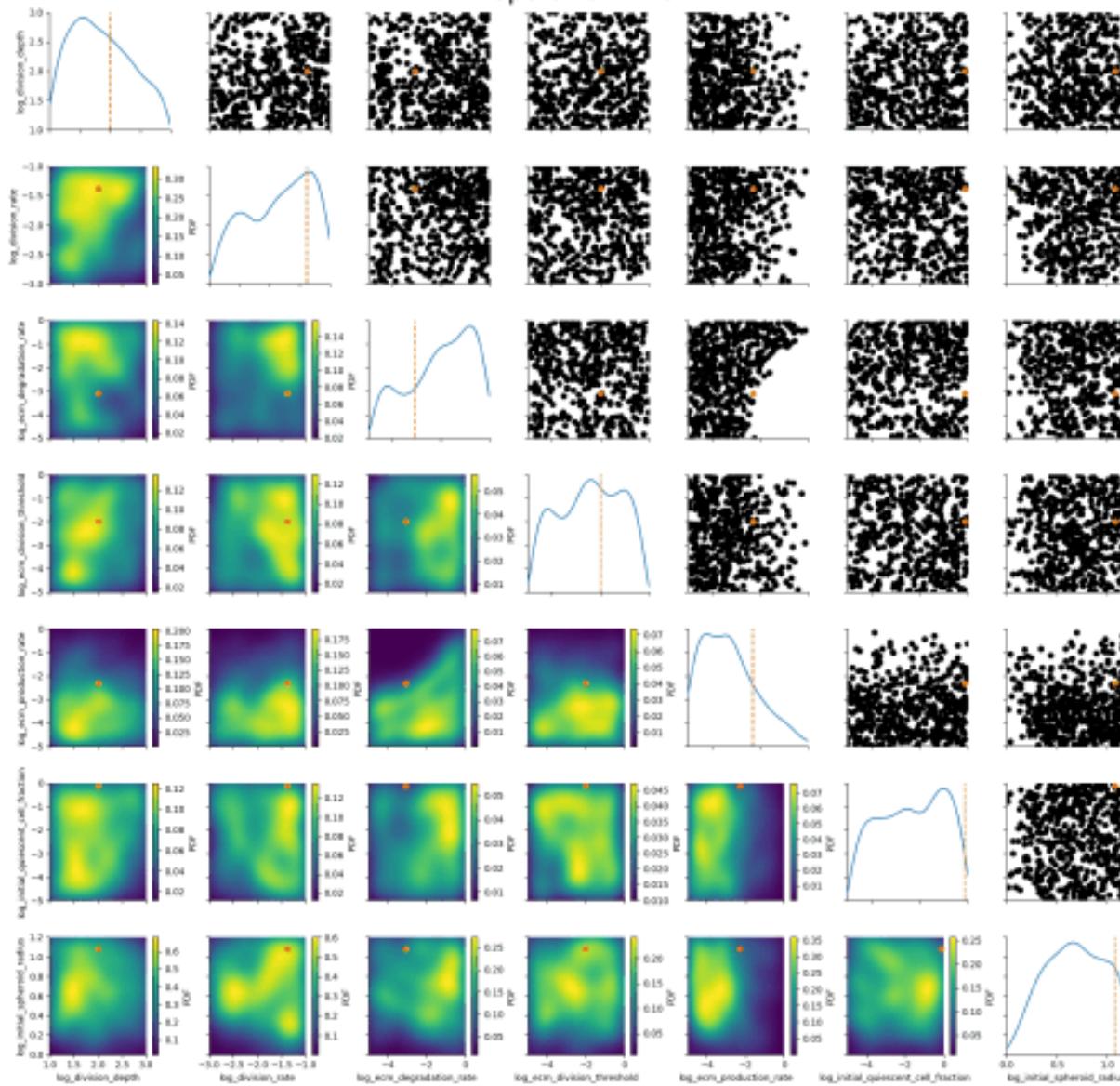


Extracellular
matrix
density



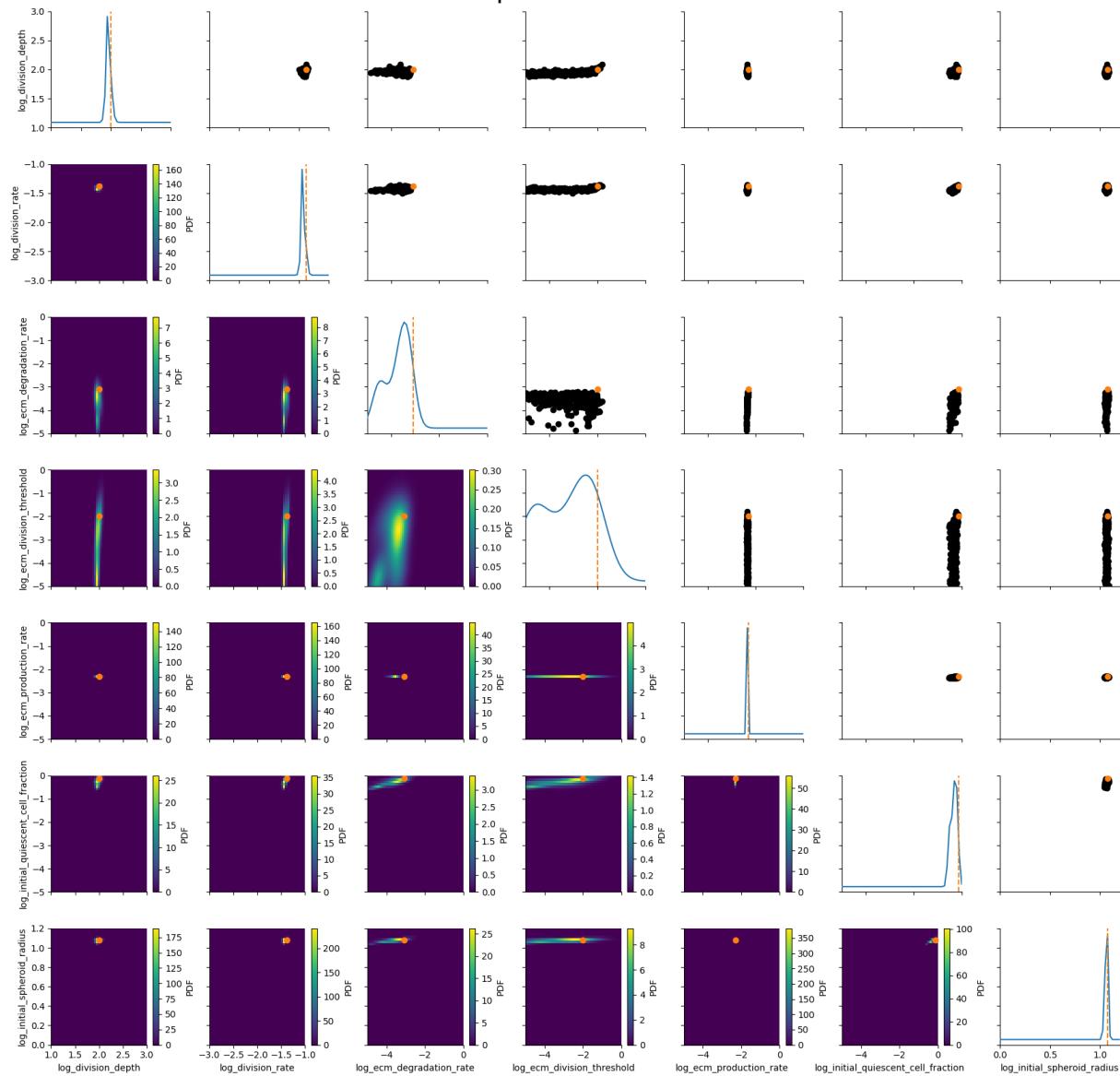
- 400 cores
- 2 days
- 1.8e6 simulations

Population t=0

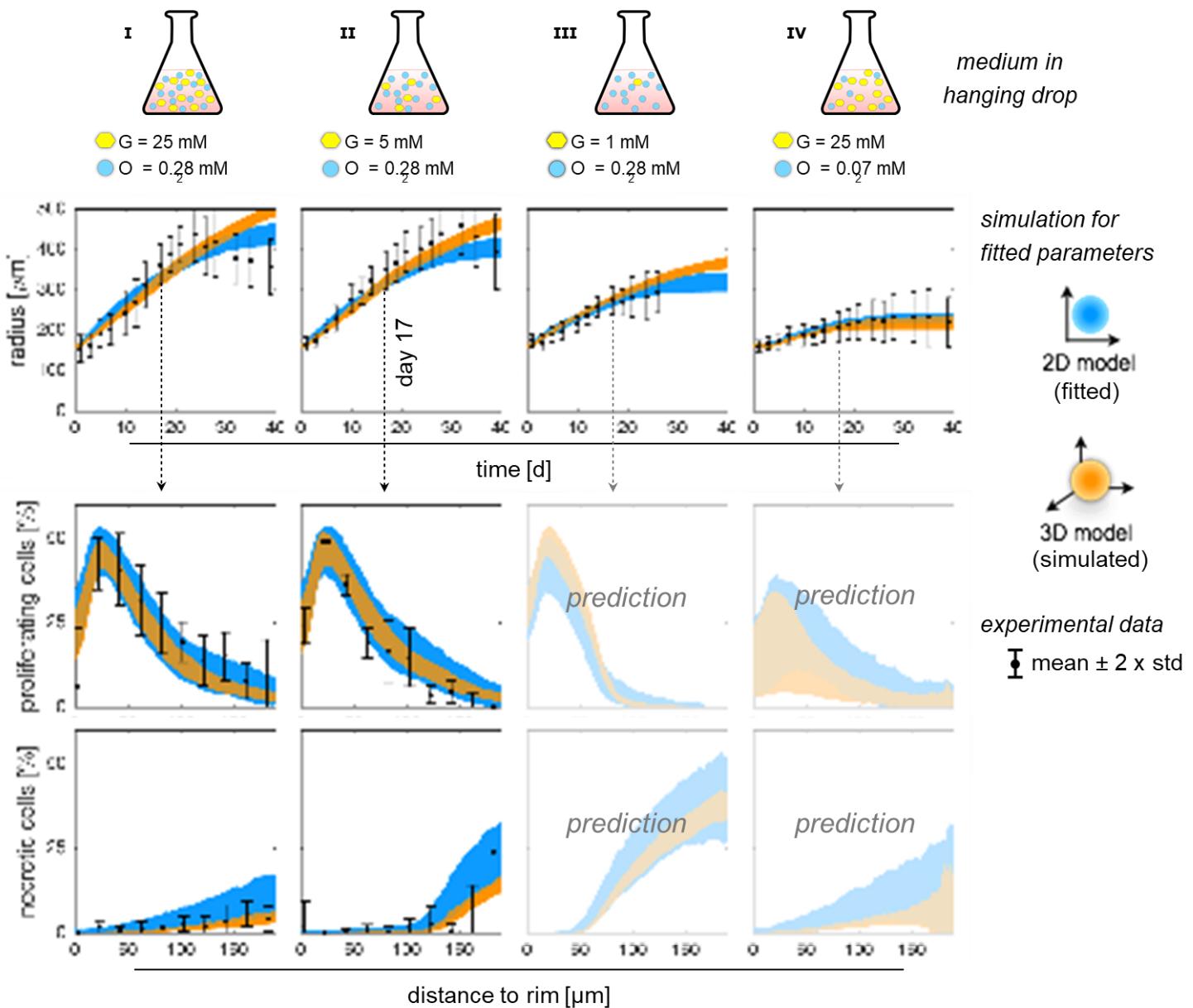


ABC worked where many other methods had failed.

Population t=37



ABC worked where many other methods had failed.

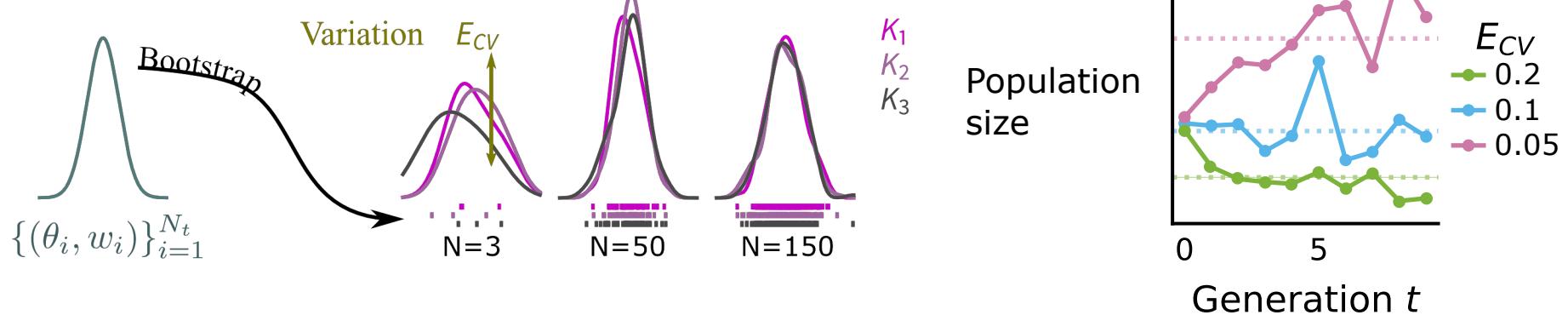


Uncertainty-aware predictions, easy data integration.

Algorithmic details

Adaptive population sizes

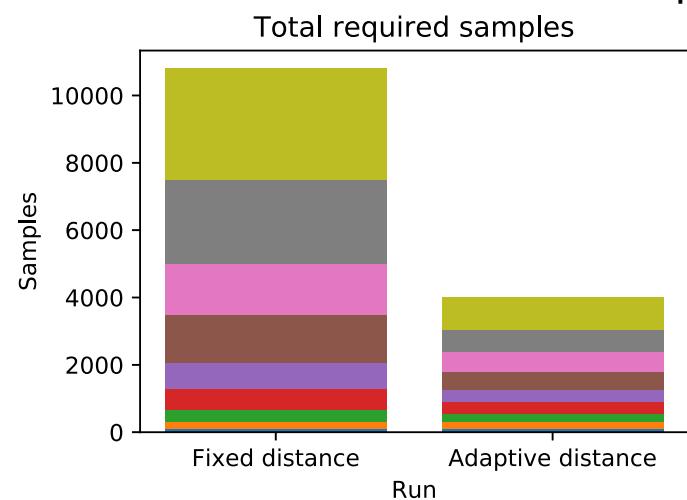
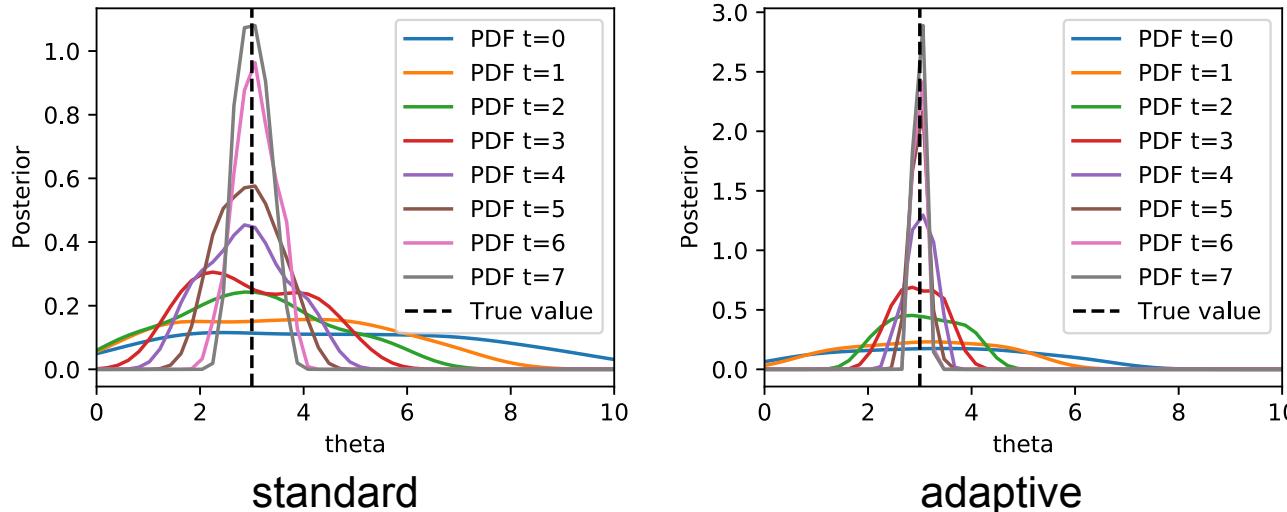
Klinger et al., CMSB Proceedings 2017



idea: adapt population size trying to match a target accuracy

Self-tuning distance functions

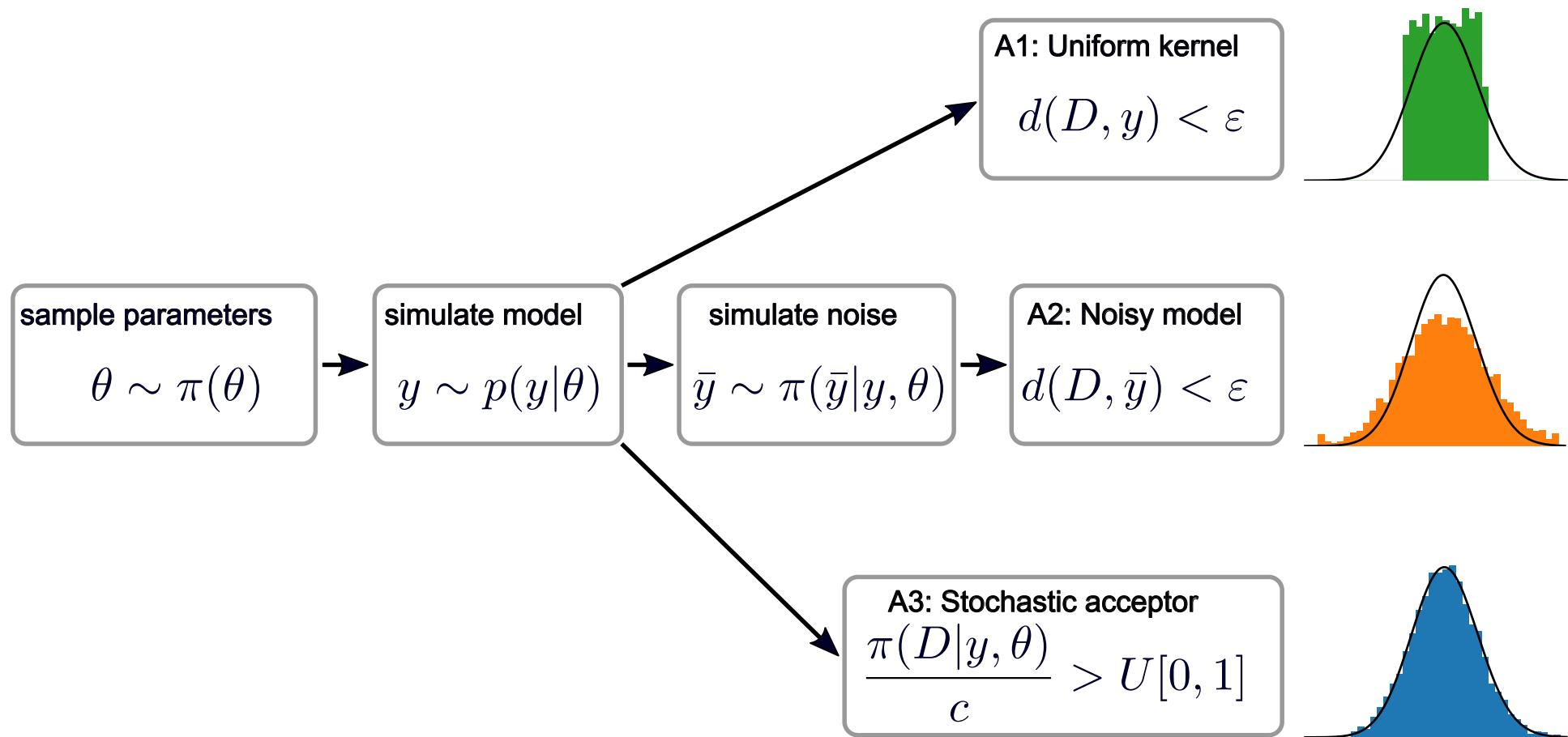
based on Prangle, Bayesian Analysis 2015



Measurement noise

Schälte et al., Bioinformatics 2020

How to efficiently account for measurement noise in ABC?



And ...

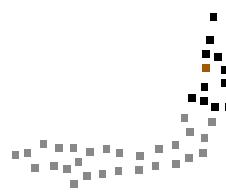
quantile-based epsilons



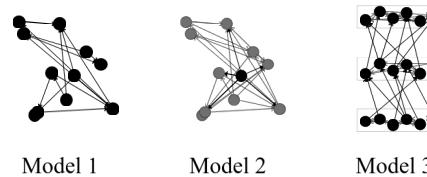
early rejection

$$\sum_i d_i < \varepsilon$$

adaptive proposal kernels



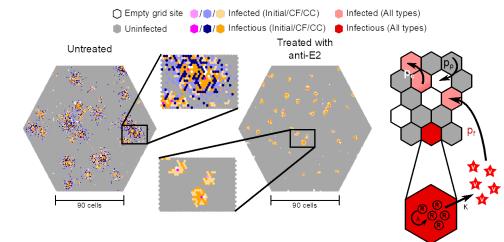
model selection



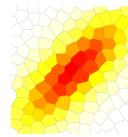
speaks python and R



other applications

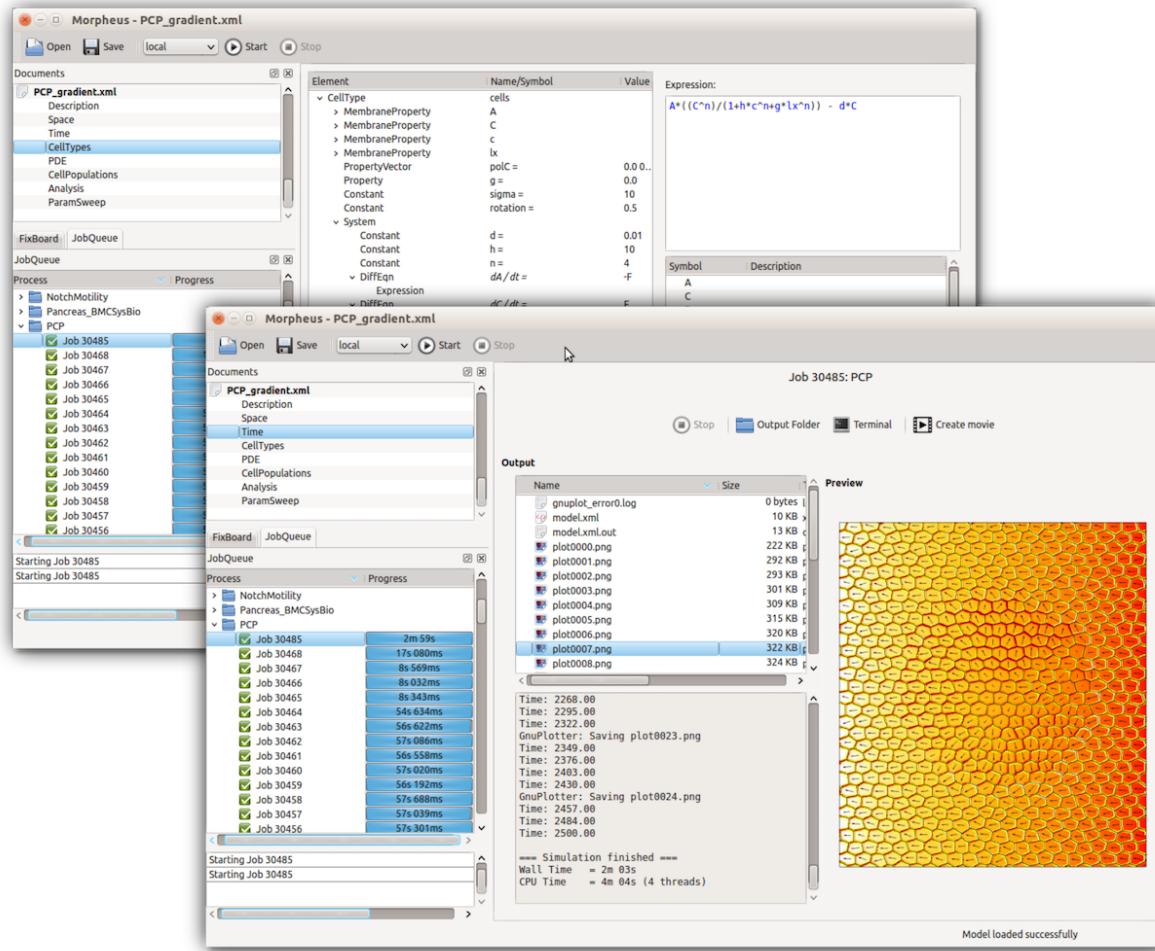


• • •



FitMultiCell

Joint initiative to perform inference for multi-cellular models



Morpheus toolbox: Staruß et al., Bioinformatics 2014

Summary

Summary

- parameter estimation when we cannot evaluate the likelihood is challenging
- ABC allows for reliable statistical inference
- pyABC provides an easy-to-use framework

Summary

- parameter estimation when we cannot evaluate the likelihood is challenging
- ABC allows for reliable statistical inference
- pyABC provides an easy-to-use framework



Not everything is a nail.

Literature



- Sisson, Scott A., Yanan Fan, and Mark Beaumont, eds. "Handbook of approximate Bayesian computation." CRC Press, 2018.
- Beaumont, Mark A. "Approximate Bayesian computation in evolution and ecology." *Annual review of ecology, evolution, and systematics* 41 (2010): 379-406.
- Martin, Gael M., David T. Frazier, and Christian P. Robert. "Computing Bayes: Bayesian Computation from 1763 to the 21st Century." *arXiv preprint arXiv:2004.06425* (2020).

Acknowledgments

Thanks to: Jan Hasenauer (and the whole lab), Elba Raimández-Álvarez, Emad Alamoudi, the FitMultiCell team, ...



Bundesministerium
für Bildung
und Forschung

HelmholtzZentrum münchen
German Research Center for Environmental Health



@yannik_schaelte



yannikschaelte



yannik.schaelte@helmholtz-muenchen.de



Thanks! Questions?