

# Approximate Bayesian computation: methods and applications for complex systems

Mark A. Beaumont,  
School of Biological Sciences,  
The University of Bristol,  
Bristol, UK

11 November 2015

# Structure of Talk

1. Motivation
2. Introduction to ABC
3. Example applications
4. ABC methodology
5. Agent-based model example
6. Methods for dealing with 'big data'.
7. Conclusions

# Modelling Complex Systems

What do we want to do?

- ▶ Understand the structure of a model
- ▶ Compare models and choose between them
- ▶ Validate and criticise models
- ▶ Make predictions
- ▶ Make decisions

# Likelihood

- ▶ In a likelihood-based analysis we aim to maximise the use of the information in the data (likelihood principle)
- ▶ Bayesian inference extends likelihood-based modelling to explicitly include a prior probability distribution on parameters.
- ▶ Amenable to robust model criticism (e.g. posterior predictive checking).

# Motivation for ABC

- ▶ We would like to make use of the likelihood function and Bayesian framework.
- ▶ For many problems it is easy to simulate data that 'looks like' real data (e.g. agent-based models), but impossible to write down the likelihood function in practice.
- ▶ However it is still possible to carry out likelihood-based and Bayesian inference via simulations (if only approximately).

# Posterior Distributions Without Likelihoods

If the data are discrete and of low dimension then it is possible to sample from the posterior density of the parameter without an explicit likelihood function, and without approximation, by the following algorithm (Rubin, 1984):

## Monte Carlo Algorithm

Given observation  $y$ , repeat the following until  $N$  points have been accepted:

1. Draw  $\theta_i \sim \pi(\theta)$
2. Simulate  $x_i \sim p(x|\theta_i)$
3. Reject  $\theta_i$  if  $x_i \neq y$

These are sampled from  $p(\theta|y)$ .

# Approximate Bayesian Computation

With high-dimensional data two approximations are useful:

- ▶ Define a distance function  $\rho(\cdot)$
- ▶ Define a summary statistic function  $S(\cdot)$

## ABC Algorithm

Given observation  $y$ , repeat the following until  $N$  points have been accepted:

1. Draw  $\theta_i \sim \pi(\theta)$
2. Simulate  $x_i \sim p(x|\theta_i)$
3. Reject  $\theta_i$  if  $\rho(S(x_i), S(y)) > \epsilon$ .

Here, as with  $x$ ,  $y$ , and  $\theta$ , the function  $S(\cdot)$  may be vector-valued.



## Example Applications of ABC

- ▶ Fitting and comparing models of cultural evolution using archaeological data (Crema *et al*, *J. Archaeol. Sci*, 2014).
- ▶ Modelling epidemiology of bovine tuberculosis in lion populations in Kruger National Park (Kosmala *et al*, *Ecological Applications*, 2015).
- ▶ Inferring cosmological parameters from Type Ia Supernova data (Weyant *et al*, *Astrophysical Journal*, 2013).
- ▶ Parameterising emulators to model climate sensitivity (effect of doubling CO<sub>2</sub> on equilibrium mean global temperature) (Holden *et al*, *Climate Dynamics*, 2010)

# ABC methodology

There are three main computational approaches for ABC:

- ▶ Rejection and regression adjustment (Beaumont *et al*, *Genetics*, 2002):
  - ▶ Use a local-linear regression model to obtain a conditional-density estimate, given the observed data.
  - ▶ Example application: inference of great ape evolutionary history from whole-genome data ( Prado-Martinez, J. *et al*. *Nature*, 2013).
- ▶ Markov chain Monte Carlo (MCMC-ABC) (Marjoram *et al*, PNAS, 2003)
  - ▶ Replace accept-reject Metropolis step in MCMC with accept-reject ABC step.
  - ▶ Example application: agent-based model of social structure dynamics in *Drosophila melanogaster* (Foley *et al*, *Am. Nat.*, 2015)
- ▶ Sequential Monte Carlo (SMC-ABC) (Sisson *et al*, PNAS, 2007)
  - ▶ Iterative refinement where parameter values are initially sampled from the prior, and finally sampled from the approximate posterior.
  - ▶ Example application: Modelling bovine tuberculosis spread among farms in the UK (Brooks-Pollock *et al*, *Nature*, 2014).

## ABC model choice

It is also straightforward in the ABC framework to compare the posterior probabilities of different models.

1. Simulate  $n$  samples from joint distribution  $p(m, S(x))$ :
  - 1.1 Simulate a model index  $m_i \sim \pi(\mathcal{M} = m) \quad m = 1, \dots, M$
  - 1.2 Simulate parameters  $\theta_{m_i} \sim \pi_m(\theta_{m_i})$
  - 1.3 Simulate  $x_i \sim p_{m_i}(x|\theta_{m_i})$
  - 1.4 Reject  $m_i$  if  $\rho(S(x_i), S(y)) > \epsilon$ .
2. Accepted  $m_i$  are sampled approximately from  $p(\mathcal{M} = m|S(y))$

The above basic rejection approach has counterparts in rejection-ABC, MCMC-ABC, and SMC-ABC.

# Summary Statistics and the Curse of Dimensionality

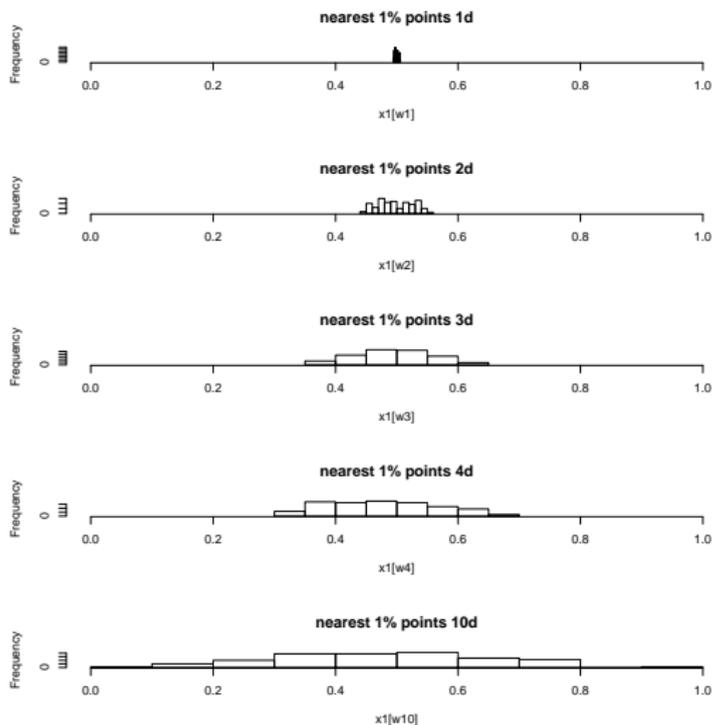
- ▶ According to the Pitman-Koopman-Darmois theorem only in the case of probability distributions in the Exponential Family can the number of sufficient statistics reach an upper bound with increasing sample size.
- ▶ *I.e.* generally we do not expect to capture all the information in the data with any predefined set of summary statistics.
- ▶ Yet ideally we would like to be as close to sufficiency as possible.
- ▶ But the more statistics we have the less likely we will be able to closely match them for any given tolerance interval  $\epsilon$
- ▶ This is an example of the 'curse of dimensionality'.

# Toy example

Sample  $X_i \sim U^d(0, 1)$   
 $i = 1, \dots, 10,000$ ;

compute  $\| X_i - 0.5 \|^2$ ;

for closest 1% of points plot  
marginal distribution for  
 $X[1]_{i=1, \dots, 10000}$ .



## Semi-automatic ABC

One approach to address the curse of dimensionality is that of Fearnhead and Prangle (RSSB, 2012)

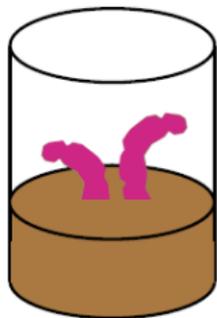
- ▶ Motivation for the method is a proof that if the posterior mean for a parameter is used as a summary statistic, this minimizes mean square error.
- ▶ Regression gives an estimate of  $E(\theta|S(x))$ .
- ▶ So we can use the linear predictor from the regression as a projection to map the vector of summary statistics to a scalar  $\hat{E}(\theta|S(x))$  for each parameter and use this projection in place of the original summary statistics.

## Example use of ABC for an agent-based model

van der Vaart, E., Beaumont, M. A., Johnston, A. S., & Sibly, R. M. (2015). Calibration and evaluation of individual-based models using Approximate Bayesian Computation. *Ecological Modelling*, 312, 182-190.

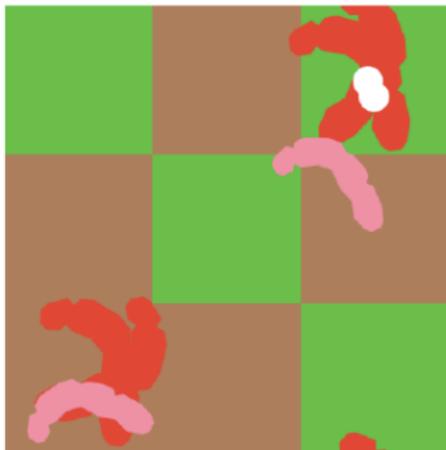
- ▶ Part of an ongoing study in collaboration with Richard Sibly and Elske van der Vaart, University of Reading (NERC Grant NE/K006088/1).
- ▶ Richard Sibly's group are developing a number of IBMs for different systems, typically with an applied ecology aim (major partner is Syngenta).

## The Earthworm Experiments



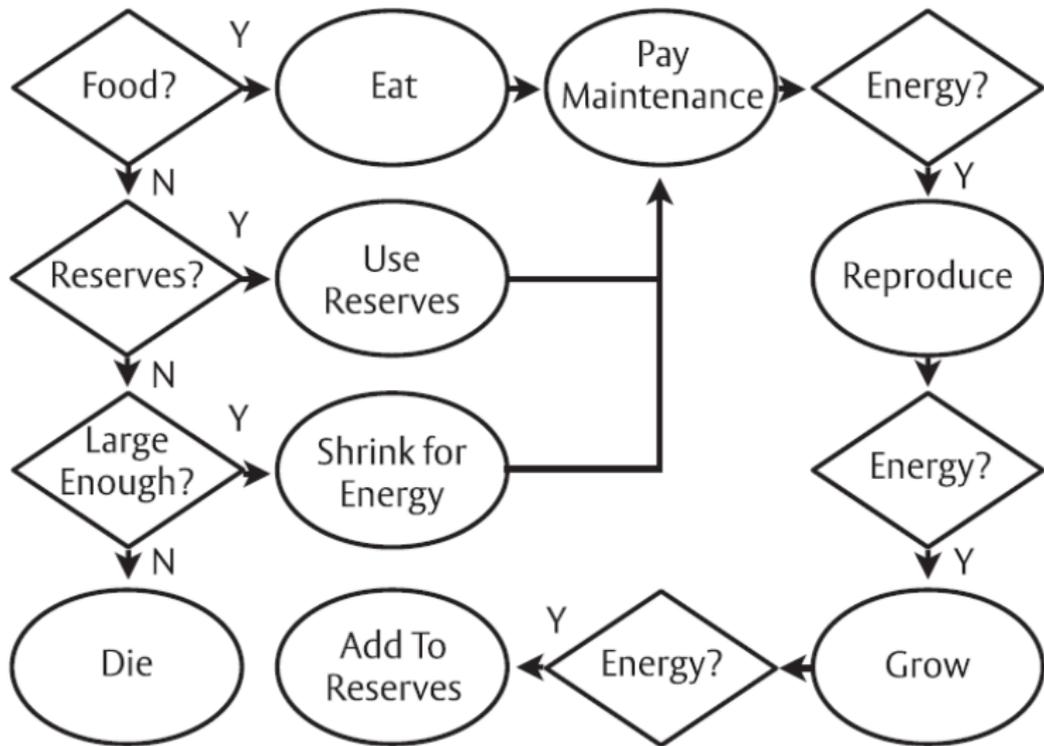
- 5 - 10 earthworms are introduced to a cylinder of manure
- the manure is gradually depleted and replenished
- the earthworms' weight and cocoon production is measured
- the objective is to quantify the effects of pesticides

## The Earthworm Model



- developed by Alice Johnston and Richard Sibly
- simulates earthworms in the lab or in the field
- each earthworm has its own energy budget
- the objective is to quantify the effects of pesticides

(Johnston et al., in press)



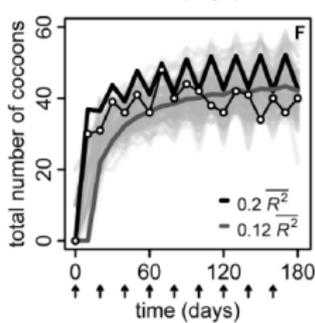
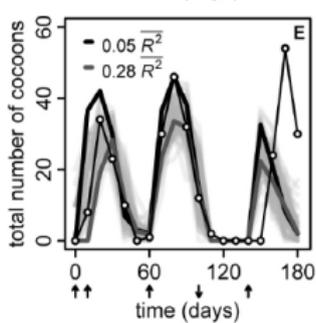
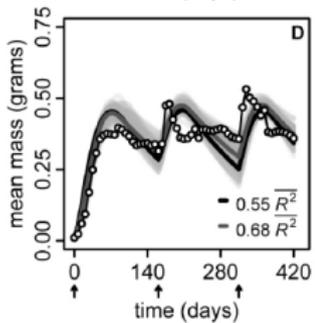
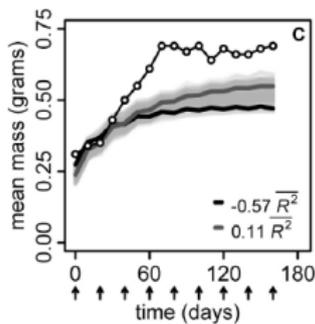
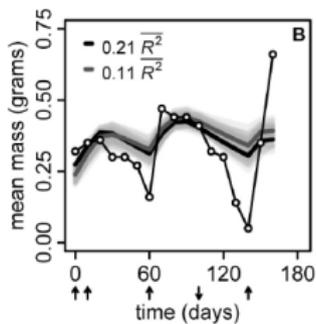
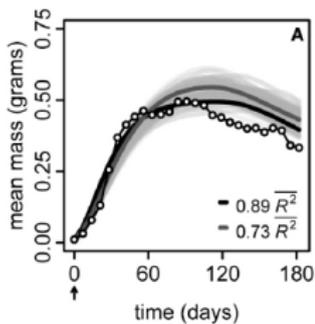
# Model Parameters

Symbol	Definition	Literature value	Unit
$B_o$	Taxon-specific normalisation constant	967	kJ/day
$E$	Activation energy	0.25	eV
$E_c$	Energy cost of tissue	3.6	kJ/g
$E_f$	Energy from food	10.6	kJ/g
$E_s$	Energy cost of synthesis	3.6	kJ/g
$h$	Half saturation coefficient	3.5	g/0.01 g
$IG_m$	Maximum ingestion rate	0.70	g/day/g
$M_b$	Mass at birth	0.011	g
$M_c$	Mass of cocoon	0.015	g
$M_m$	Maximum asymptotic mass	0.5	g
$M_p$	Mass at sexual maturity	0.25	g
$r_B$	Growth constant	0.177	day <sup>-1</sup>
$r_m$	Maximum energy to reproduction	0.182	kJ/g/day
$s$	Movement speed	0.004	m/day

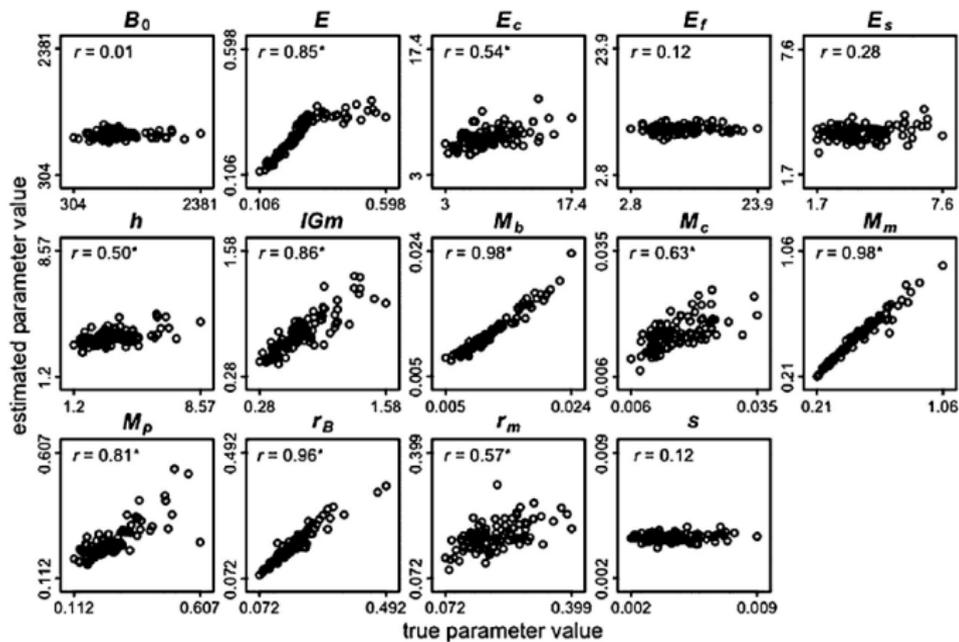
## Application of ABC

- ▶ Utilise massively parallel capability of basic rejection-ABC to run jobs on supercomputer (many tens of thousands of cores).
- ▶  $10^6$  simulations performed, 100 closest points accepted.
- ▶ We use all 143 summary statistics.

# Results



# Ability to recover parameters



# Summary

ABC has been a useful tool to explore the model:

- ▶ We can see which features of the data are/are not captured by the model.
- ▶ A number of parameters are non-identifiable in the model.

This follows a hypothetico-deductive framework (Gelman and Shalizi, *B. J. Math Stat Psych.*, 2012: Philosophy and practice of Bayesian statistics) in which we constructed and then falsified a model for the earthworms.

## ABC and 'Big Data'

- ▶ In genomics ABC has been used to tackle very large data sets (e.g. whole genome data ( $3 \times 10^9$  base positions across many individuals)).
- ▶ Typically this involves computing summary statistics that are averages across units, which loses a lot of information.
- ▶ There is interest in finding more accurate ways of modelling the data (for example, Hierarchical-ABC, Bazin *et al.*, Genetics, 2010).

# Expectation Propagation and ABC

- ▶ EP is a variational method used in Bayesian computation that aims to approximate the posterior distribution and avoid potentially slow Monte Carlo solutions.
- ▶ Recent developments suggest that it may have a wider utility in conjunction with Monte Carlo.

## Expectation Propagation: outline

- ▶ Assume that the likelihood can be factorised:

$$p(x|\theta) = \prod_{i=1}^n p(x_i|\theta)$$

So:

$$p(\theta|x) \propto \pi(\theta) \prod_{i=1}^n p(x_i|\theta)$$

- ▶ The idea is to find a solution in terms of matching factors:

$$p(\theta|x) = \prod_{i=0}^n g_i(\theta)$$

where typically the prior is given by  $g_0(\theta)$

## Expectation Propagation: outline

The parameters of the  $g_i(\theta)$  are fitted in a series of sweeps through the data.

For the  $i$ th item of data (a 'site'):

- ▶ The 'cavity distribution' is formed:

$$g_{-i} = \prod_{j=0}^n g_j(\theta) / g_i(\theta)$$

- ▶ and the 'tilted distribution':

$$\propto g_{-i}(\theta) p(x_i | \theta)$$

- ▶ A new  $g'_i(\theta)$  is found that minimises the Kullback-Leibler divergence between the tilted distribution and

$$g'(\theta) = g_{-i}(\theta) g'_i(\theta)$$

- ▶ If the  $g_i(\theta)$  are from the Exponential Family this can be achieved by matching the moments between the tilted distribution and  $g_{-i}(\theta) g'_i(\theta)$ .

This is repeatedly applied to all  $n$  sites until convergence.

## Expectation Propagation: outline

Typically the  $g_i(\theta)$  are multivariate normal distributions, with natural parameters  $\mu_i \Sigma_i^{-1}$ , the scaled mean vector, and the precision matrix  $\Sigma_i^{-1}$ .

The nice property of using the natural parameters is that

$$g\left(\theta; \sum_{i=0}^n \mu_i \Sigma_i^{-1}, \Sigma_i^{-1}\right) = \prod_{i=0}^n g(\theta; \mu_i \Sigma_i^{-1}, \Sigma_i^{-1})$$

## Expectation Propagation: algorithm of Barthelmé and Chopin

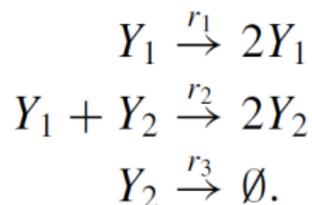
Typically, the integration of the tilted distribution is done analytically or by quadrature. The problem is usually structured so that only low-dimensional integration is necessary.

Barthelmé and Chopin (2013; JASA) proposed an approximate Bayesian computation (ABC) algorithm for cases where the integration was problematic. In this case

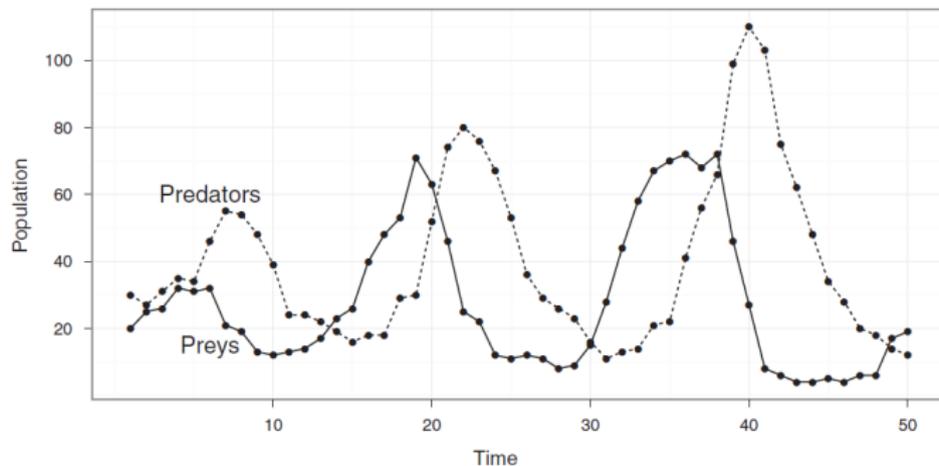
- ▶ Form the cavity distribution in the normal way with parameters  $\mu_{-i}\Sigma_{-i}^{-1}$  and  $\Sigma_{-i}$
- ▶ Simulate random draws (indexed by  $[m]$ ) of  $\theta^{[m]}$  from the cavity distribution.
- ▶ Simulate  $x_i^{[m]} \sim p(x_i|\theta)$ .
- ▶ Accept  $\theta^{[m]}$  giving  $x_i^{[m]}$  that are within some tolerance interval  $\epsilon$  of the data  $x_i^*$ .
- ▶ Compute moments from these to get  $\mu'$  and  $\Sigma'$ , and hence update  $g_i(\theta)$  and  $g(\theta)$

# Stochastic Lotka-Volterra Example

Abundances of prey  $Y_1$  and predator  $Y_2$  follow these dynamics:



SDE easily simulated by Gillespie's method. Test data below simulated with  $r_1 = 0.4$ ,  $r_2 = 0.01$ ,  $r_3 = 0.3$ .



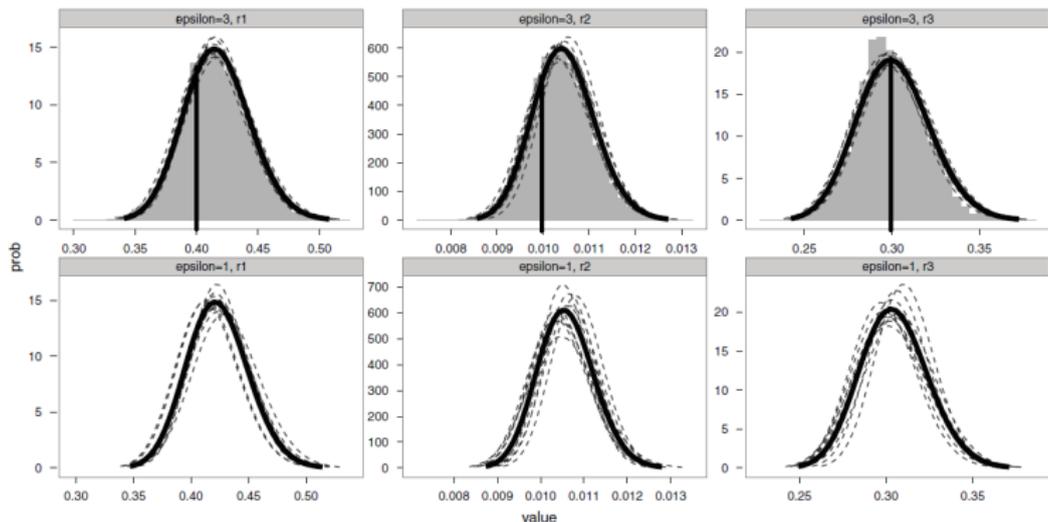
(Barthelmé and Chopin, 2013)

# Lotka-Volterra Example

Dotted lines: single runs of EP-ABC (2 minutes CPU)

Black line: average of 10 runs

Grey histogram: PMCMC-ABC (2 days CPU)



(Barthelmé and Chopin, 2013)

# Conclusions

- ▶ The main strength of ABC is that it allows the full Bayesian machinery of model-fitting and model-criticism to be applied to (almost) arbitrarily complicated systems.
- ▶ ABC is still undergoing a lot of development.
- ▶ Current areas of research:
  - ▶ Using machine learning to approximate the joint distribution of parameters and summary statistics.
  - ▶ Coping with model discrepancy.
  - ▶ Dimension reduction for summary statistics
  - ▶ Expectation propagation