Non-linear regression models for Approximate Bayesian Computation (ABC)

Michael Blum   Olivier François
TIMC-IMAG, Faculty of Medicine
U. Grenoble

Accept/Reject ABC methods suffer from the curse of dimensionality

ABC is wasteful. A typical usage may involve millions of simulations, and it retains only a few thousands.

Be more tolerant and recycle the wasted data by using regression-based ABC (Beaumont et al 2002)

Linear regression-based ABC can sometimes be improved

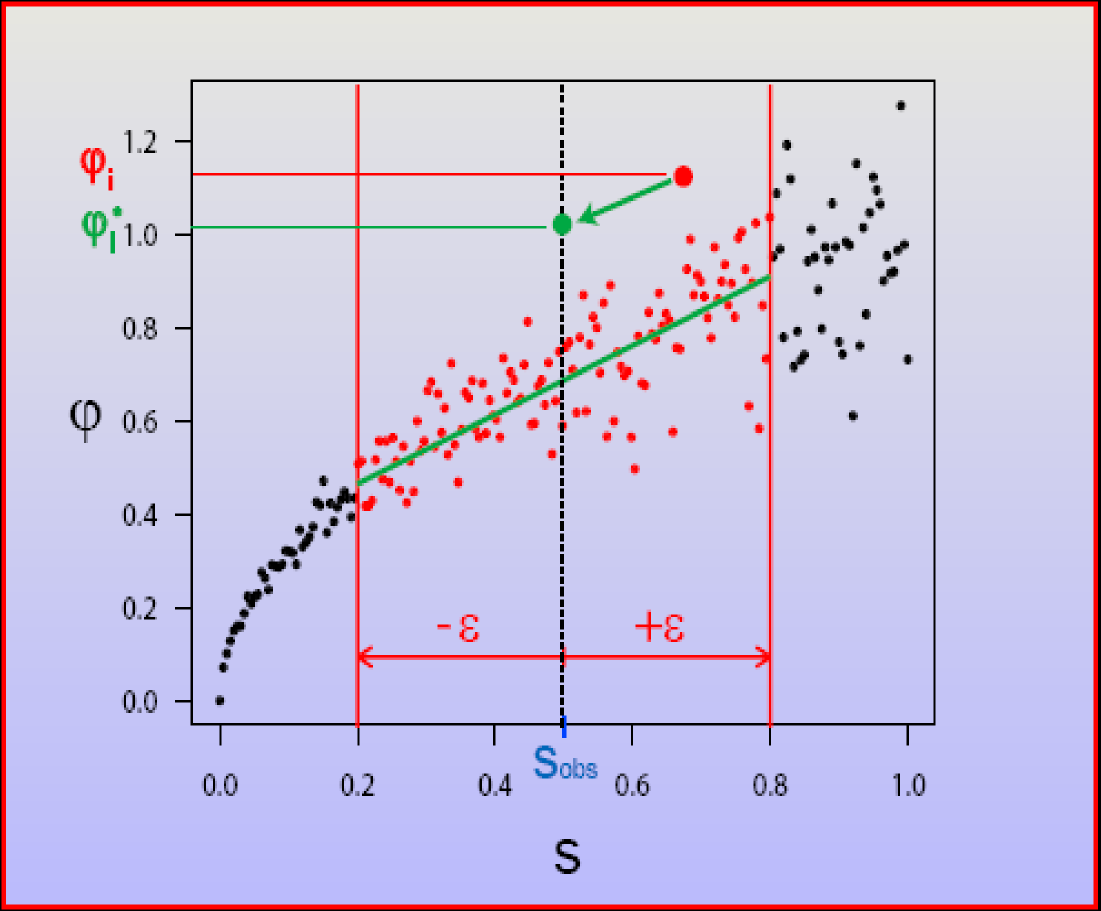
abc of ABC

Using stochastic simulations with known parameters, $\theta \sim \pi(\theta)$, we compute a subset of summary statistics, s, and we compare them to the observed subset, s_{obs}

$$P_{\varepsilon}(\ \theta \mid D\ ) \quad \alpha \quad P(\ \|s - s_{obs}\| < \varepsilon \mid \theta\ )\ \ \pi(\theta)$$

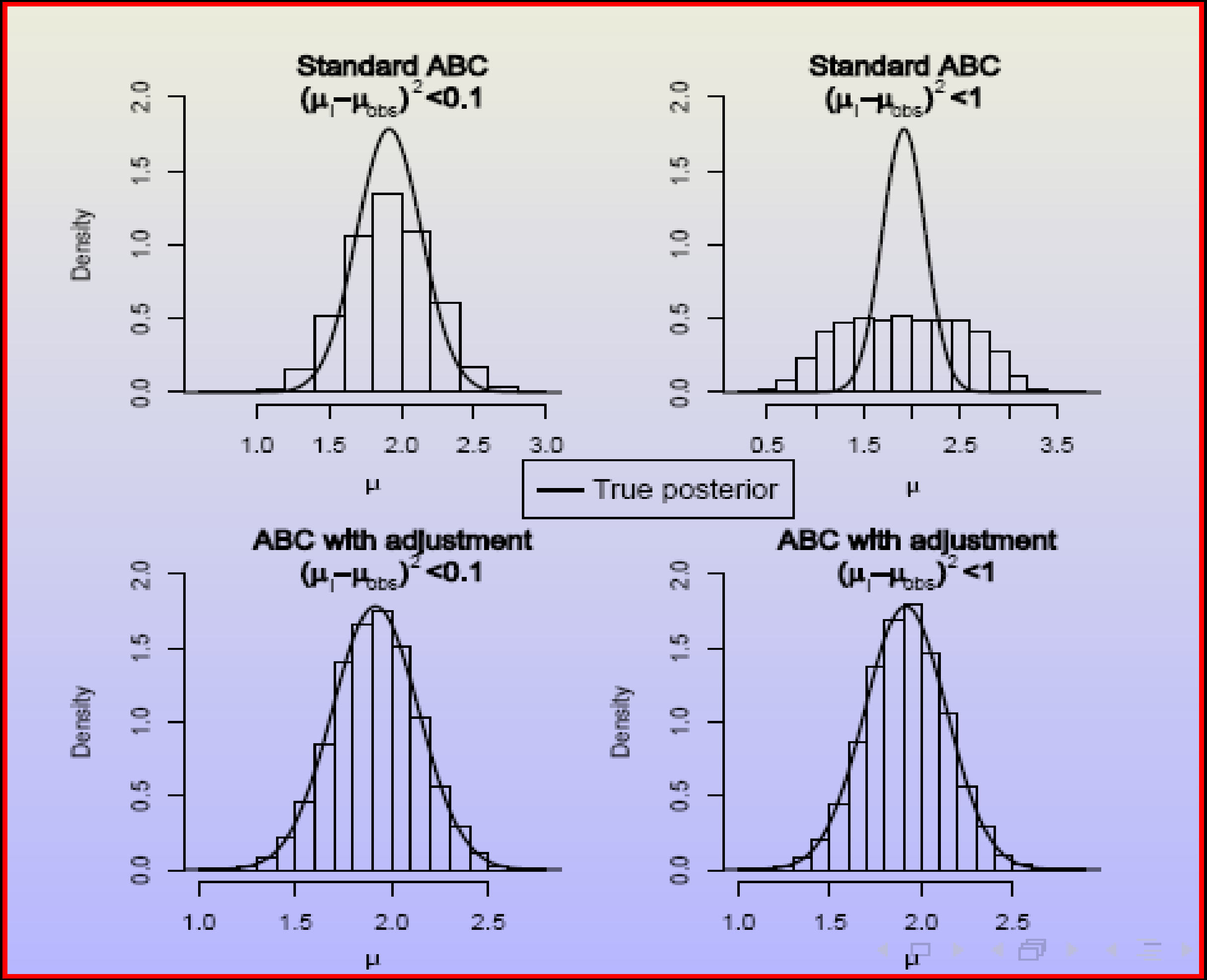
Problematic if s is of large dimension

Beaumont et al (2002) increase the tolerance ε , and correct the posterior distribution by performing linear regression

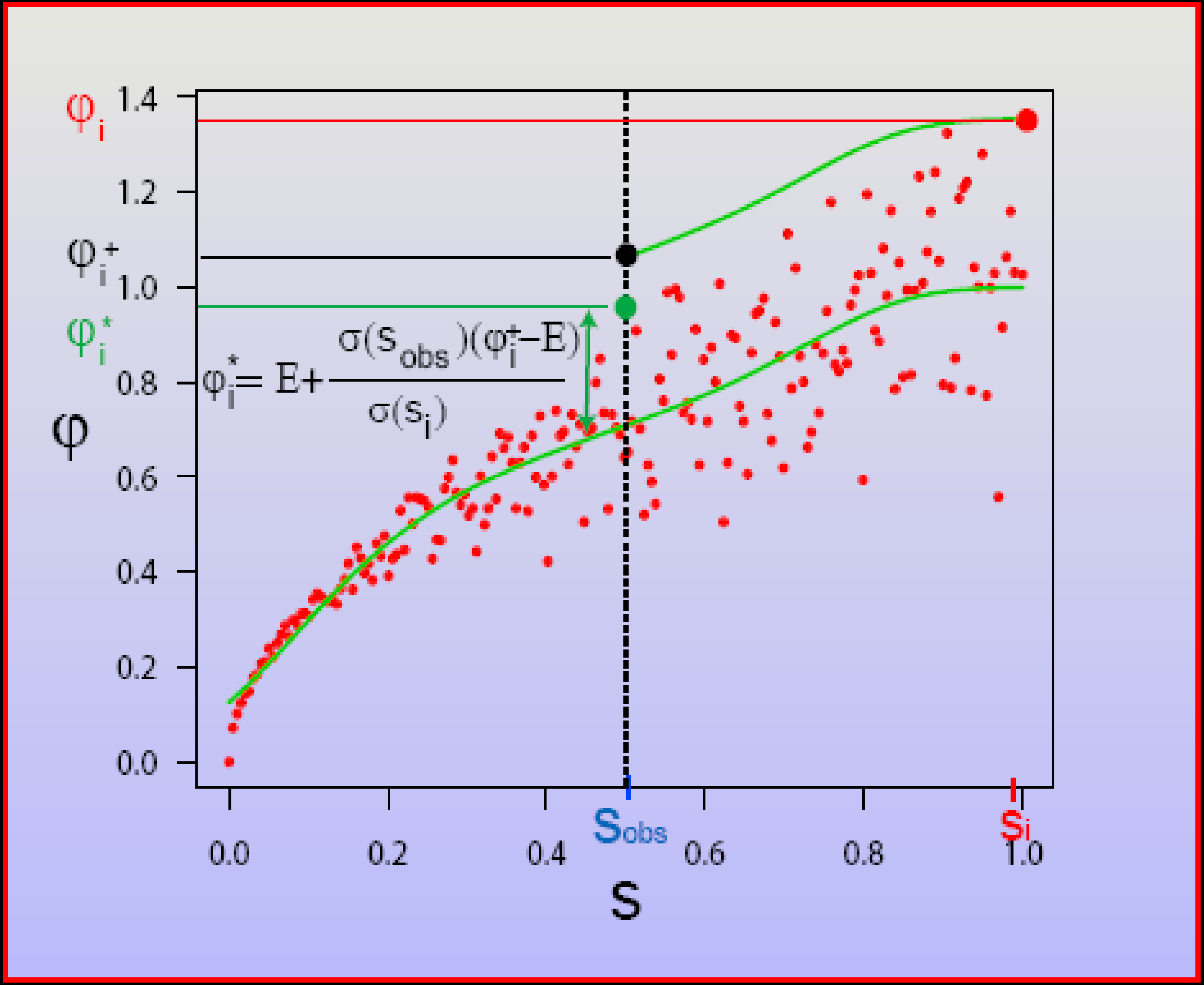


φ = model parameter, s = summary statistic (here in one dimension)

Linear regression improves ABC



Blum and OF (2009) suggest the use of non-linear conditional heteroscedastic regression models



One example in population genetics

Given n = 100 DNA sequences at a particular locus, and having observed 10 polymorphic sites at this locus,

Estimate θ the scaled mutation rate (or the population size).

Generating model for s

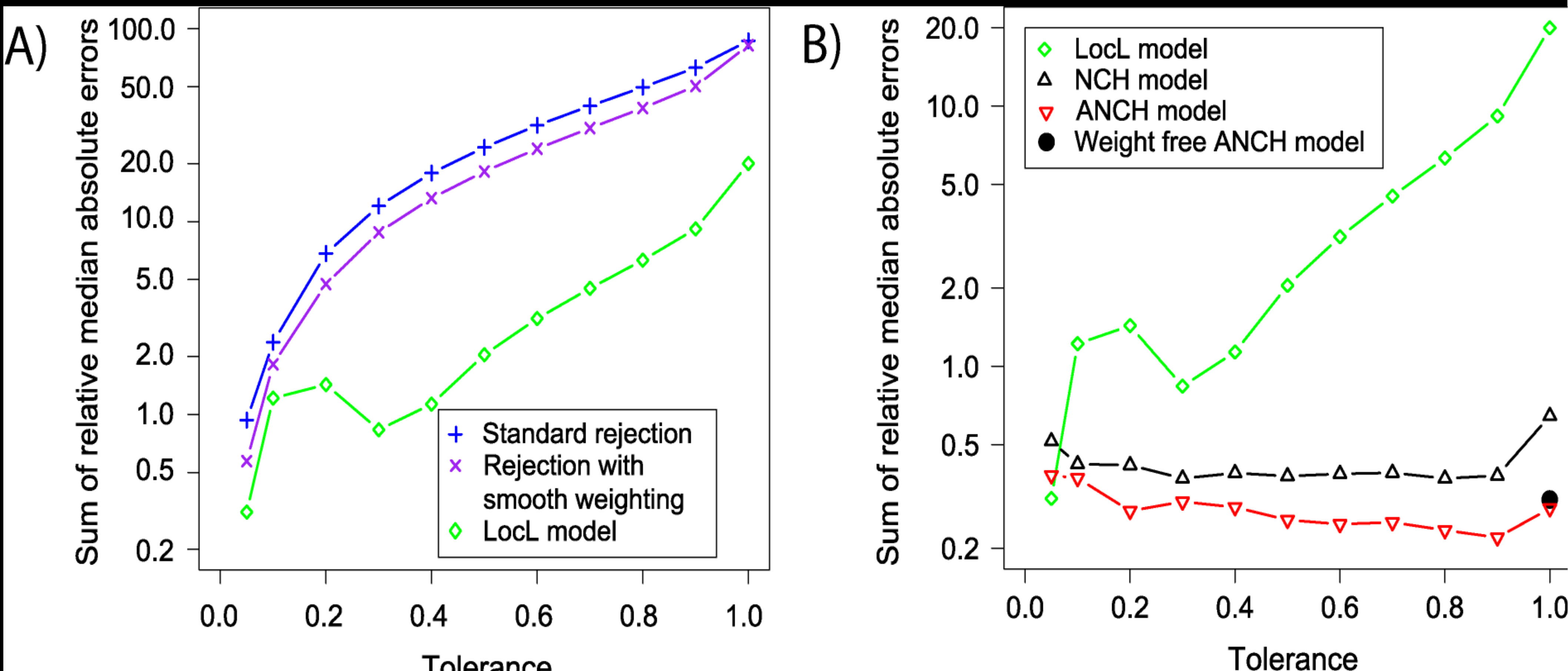
1. Draw θ from a vague prior

2. Simulate L_n the length of the sample genealogy

3. Generate $s \sim \text{Poisson}(\theta L_n/2)$

Replicated 2,000 times

Results : Distance to the (exact) posterior distribution



Reducing the dimensionality of the set of summary statistics

This may be achieved by many regression models, like ridge regression, gam's or projection pursuit.

We choose feedforward neural networks (Ripley 1996; Bishop 2006)

Why neural nets?

Their first layer allows for a nonlinear projection on a subspace of much lower dimensionality

Regression is performed using the reduced number of projection variables

Automatic choice of (new) summary statistics

Neural nets

Build H linear combinations of summary statistics ($H < D$ = dimension of s)

$$S_h = w_{0h} + \sum_{k=1...D} w_{kh}\, s^k \,, \qquad\qquad h = 1,...,H$$

Perform non-linear regression on the H projections

$$\begin{aligned} \theta &= w_{00} + \sum_{h=1...H} w_{h0}\, \varphi(S_h) + \text{error} \\ &= g_w(s) + \text{error} \end{aligned}$$

Fitting NN

Weighted (ie local) penalized least-squares criterion

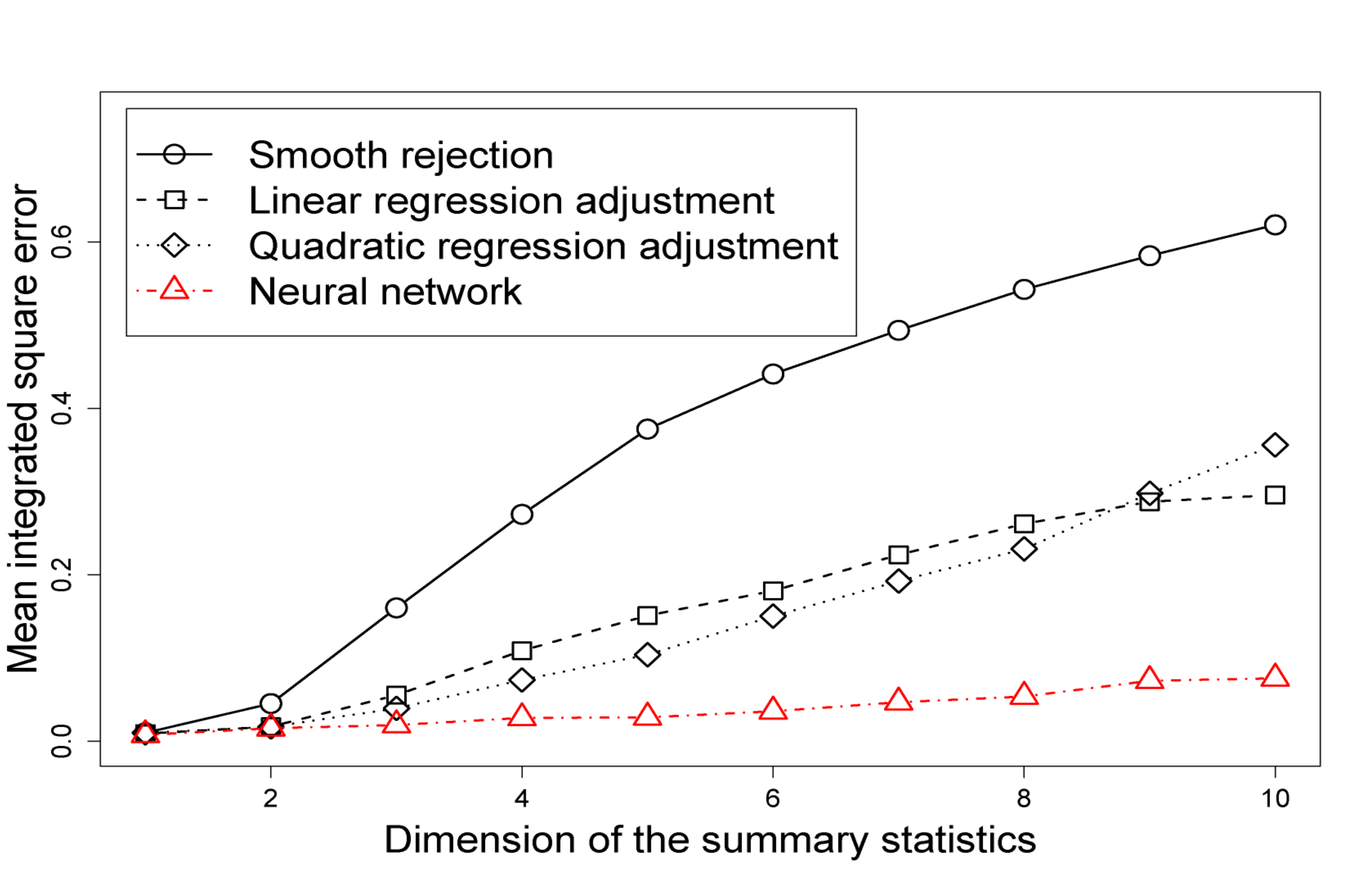
$$\sum_{i=1...n} \ (\theta_i - g_w(s_i))^2 \ K_\varepsilon(||s_i - s_{obs}||) \ + \lambda \ ||w||^2$$

Weight decay, λ, is determined by cross-validation

Posterior sampling correction:

$$\theta^*_i = \theta_i + g_w(s_{obs}) - g_w(s_i), \qquad i = 1,\ldots,M$$

Michael's example (adding useless statistics)



A second example (Heggland and Frigessi, 2004)

M/G/1 queue: single server model

Arrival times ~ Unif(θ_1, θ_2)

Service time ~ Exp(θ_3)

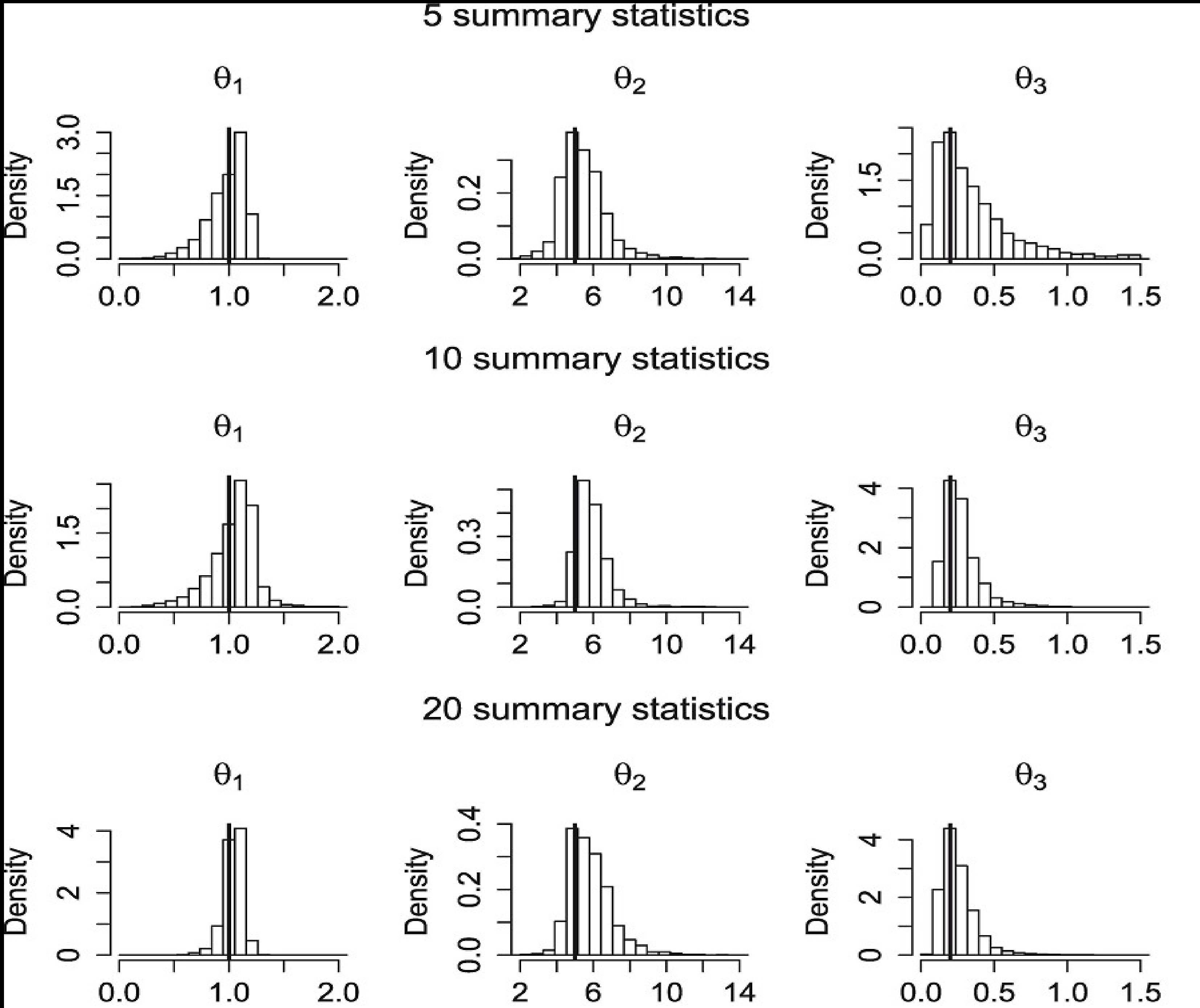
Observation: n = 50 inter-departure times, y

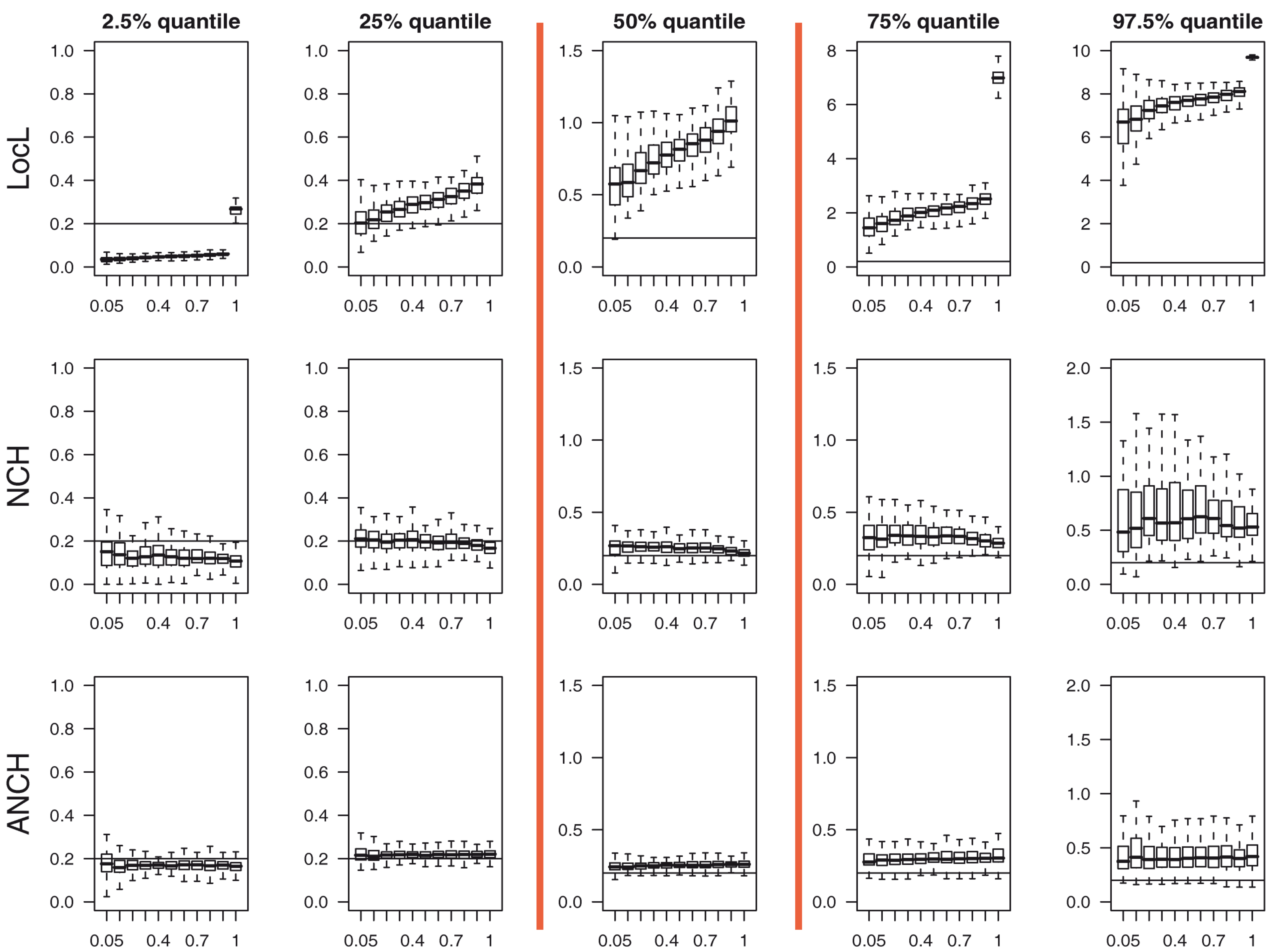
Goal: Estimate the service rate θ_3

Summary statistics: 20%, 10% and 5% quantiles (percentiles) of y

Dimension of s = 5, 10, 20

Tolerance rate = 50%


5 summary statistics
10 summary statistics
20 summary statistics
θ_1
θ_2
θ_3
Density



2.5% quantile
25% quantile
50% quantile
75% quantile
97.5% quantile
LocL
NCH
ANCH
0.05
0.4
0.7
1

Model choice

Computing posterior probabilities for *K* models $M_1, \dots, M_K$

Neural nets extend the multinomial logistic regression approach of Beaumont (2007)

Logistic outputs + entropy criterion

Other extensions

Adaptive NCH model (Blum and OF 2009)

Based on the estimation of support of the posterior density (with SVM) and iterated IS

Conclusions

Regression-based ABC is a post-simulation approach

Easy to implement and pragmatic

R functions available (CRAN package soon)

Acknowledgments

ANR MAEV Modeles Aleatoires pour l'EVolution

Grant UJF-ABC Approches Bayesiennes pour les modeles Complexes