Recent Advances in Approximate Bayesian Computation Methodology

Application in structural dynamics

A. Ben Abdessalem, K. Worden, N. Dervilis, D. Wagg

E-mail

a.b.abdessalem@sheffield.ac.uk

Dynamics Research Group, University of Sheffield

Bristol, January 9th, 2017

Outline

Aims

The **aims** of this talk can be summarised in the following points:

1. Introduce the ABC as a promising alternative for model selection and parameter estimation.
2. Introduce a new variant of ABC algorithms: **ABC-NS**.
3. Make a comparison between ABC-NS and ABC-SMC.
4. Illustrate the efficiency of the ABC-NS through some examples.

Outline

Model selection

- Strike the right balance

Goodness-of-fit | Complexity of the model

Parsimony principle

- Simple models are preferred if they provide the same prediction capability as the complex ones.

The general Bayesian framework

- Bayes theory:

Given a model $\mathcal{M}$, an observed data $\mathcal{D}_{obs}$ and a set of parameters θ:

$$\pi(\theta|\mathcal{D}_{obs},\mathcal{M}) = \frac{f(\mathcal{D}_{obs}|\theta,\mathcal{M})\pi(\theta|\mathcal{M})}{\pi(\mathcal{D}_{obs}|\mathcal{M})} \propto f(\mathcal{D}_{obs}|\theta,\mathcal{M})\pi(\theta|\mathcal{M})$$

- $f(\mathcal{D}_{obs}|\theta,\mathcal{M})$: is the likelihood function,
- $\pi(\theta|\mathcal{M})$: is the prior distribution,
- $\pi(\mathcal{D}_{obs}|\mathcal{M})$: is a normalisation constant (marginal likelihood).

→ Numerous sampling methods can be used to sample from the posterior distribution (MH, MCMC,...).

Model assessment and comparison

- Information criterions

$$\text{IC} = \underbrace{\bar{D}}_{\text{Deviance = fit of a model}} + \underbrace{p_D}_{\text{Complexity of a model}}$$

$$\bar{D} = -2\max_{\hat{\theta}_\ell} \log \mathcal{L}_N(\mathcal{M}, \hat{\theta}_\ell)$$

Information criterions

Criterion	Penalty Term for model $\mathcal{M}_\ell$
AIC	$d_{\mathcal{M}_l}$
AIC_C	$\frac{N d_{\mathcal{M}_l}}{N - d_{\mathcal{M}_l} - 1}$
BIC	$0.5 \log(N) d_{\mathcal{M}_l}$
BIC_2	$0.5 \log(N) d_{\mathcal{M}_l} - \log(2\pi) d_{\mathcal{M}_l}$

Model assessment and comparison

- ICs offer a way to estimate the generalisation performance of the model → the model with the smallest IC value should be preferred.
- The ICs present some drawbacks:
  - Estimated in a single point, the MLE $\hat{\theta}_\ell$.
  - Cannot be used to compare models with the same number of parameters.

Outline

ABC principle

In the ABC framework, the posterior distribution is given by:

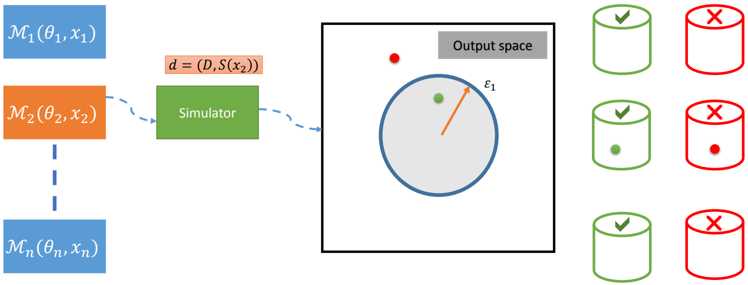
$$\pi_{\text{ABC}}(\theta|\mathcal{D}_{sim}, \mathcal{D}_{obs}, \mathcal{M}) \propto f_w(\mathcal{D}_{obs}|\mathcal{D}_{sim}, \theta, \mathcal{M}) f(\mathcal{D}_{sim}|\theta, \mathcal{M}) \pi(\theta|\mathcal{M})$$

- $\mathcal{D}_{sim}$ is the simulated data.
- $f_w(\mathcal{D}_{obs}|\mathcal{D}_{sim}, \theta, \mathcal{M})$ is the weighting function so-called indicator function.

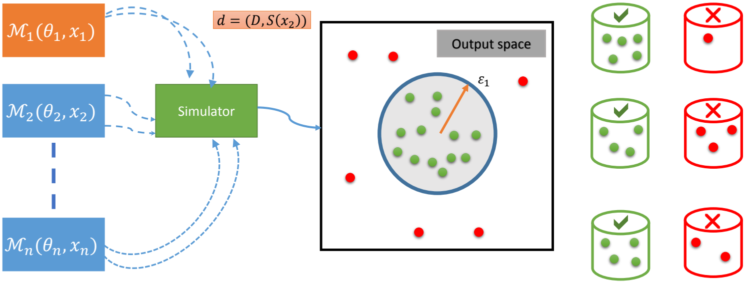
$$f_w(\mathcal{D}_{obs}|\mathcal{D}_{sim}, \theta, \mathcal{M}) \propto \mathbb{I}(d(\mathcal{D}_{obs}, \mathcal{D}_{sim|\theta}) \leq \varepsilon, \mathcal{M})$$

- $d(\mathcal{D}_{obs}, \mathcal{D}_{sim|\theta})$ is a distance measure between the observed data and the simulated data.
- If ε is small enough, $\pi_{\text{ABC}}(\theta|\mathcal{D}_{sim}, \mathcal{D}_{obs}, \mathcal{M})$ is a good approximation of the true posterior distribution.

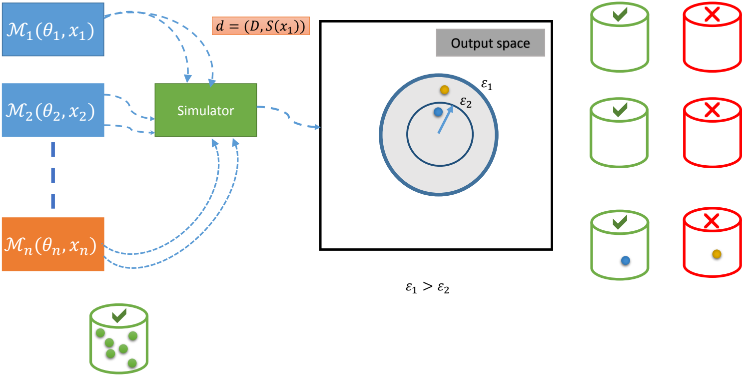
ABC-SMC principle



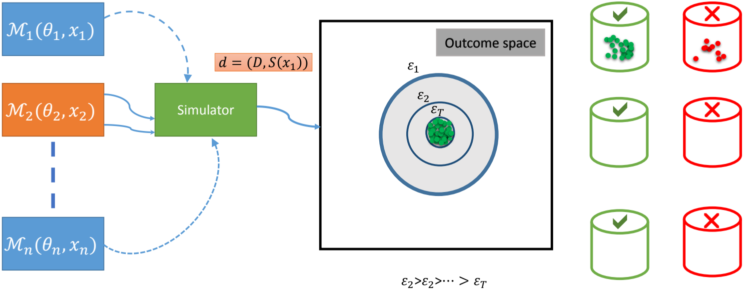
ABC-SMC principle



ABC-SMC principle



ABC-SMC principle



$$\hat{p}_{\mathcal{M}_i} = \frac{\text{Number of particles accepted for } \mathcal{M}_i}{\text{Total number of particles}}$$

ABC-SMC principle

Hyperparameters to be defined:

- $\Delta(u, u^*)$: distance to measure the degree of similarity.
- Sequence of the tolerance thresholds: $\varepsilon_1 > \varepsilon_2 > \ldots > \varepsilon_T$.
- Type of kernels to move particles and models.
- N number of particles per population.
- Stopping criterion.

- The efficiency of the algorithm depends heavily on the selection of the tuning parameters.
- Always the objective is to strike the right balance between computational requirements and acceptable precision.

ABC favours automatically simpler models.

Example 1: Cubic and CQ models

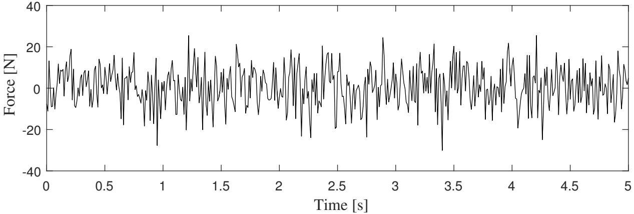
→ Two competing models: each model is characterised by a set of parameters Θ.

$$\begin{aligned}\mathcal{M}_1 &: \quad m\ddot{z} + c\dot{z} + kz + k_3 z^3 = f(t),\\ \mathcal{M}_2 &: \quad m\ddot{z} + c\dot{z} + kz + k_3 z^3 + k_5 z^5 = f(t).\end{aligned}$$

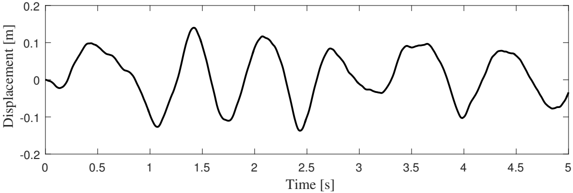
Parameter	True value	Lower bound	Upper bound
m	1	0.1	10
c	0.05	0.005	0.5
k	50	5	500
k_3	10^3	10^2	10^4
k_5	10^5	10^4	10^6

Example 1: Cubic and CQ models

- Input force: Gaussian with $\mu = 0, \sigma = 10$.



- Training data from the CQ model (noise of 1% r.m.s was added).



Example 1: Cubic and CQ models

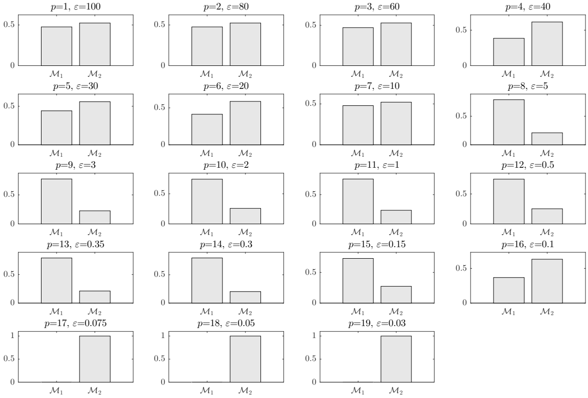
→ Selection of the ABC-SMC[Ref. 1] hyperparameters:

- Number of particles N=1000.
- Equal prior probabilities: $p(\mathcal{M}_1) = p(\mathcal{M}_2) = \frac{1}{2}$.
- Uniform priors for model parameters.
- Uniform kernels are used to move particles.
- The tolerance thresholds sequence is set successively to be $\varepsilon_{i=1:19} = (100, 80, \ldots, 0.05, 0.03)$.
- The MSE is considered to measure the level of agreement:

$$\boxed{\Delta(u^{\text{obs}}, \check{u}^{\text{sim}}) = \frac{100}{p\sigma_u^2} \sum_{\ell=1}^{p} \left(u_\ell^{\text{obs}} - \check{u}_\ell^{\text{sim}}\right)^2}$$

Example 1: Cubic and CQ models

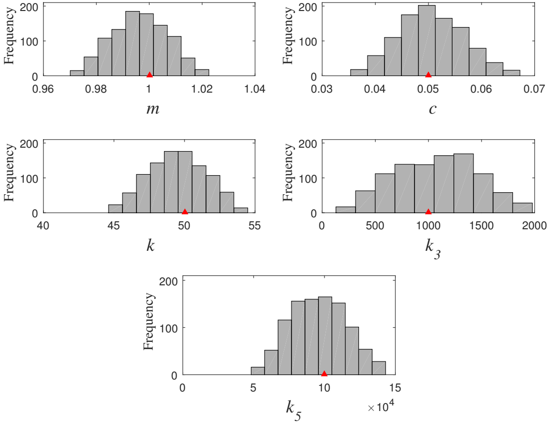
- Model posterior probabilities:



- The ABC-SMC algorithm tends first to converge to the cubic model (the model with less parameters).

Example 1: Cubic and CQ models

- Histograms of the model parameters.



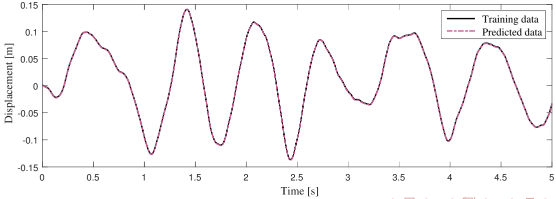
- The true parameter values are marked by a triangle.

Example 1: Cubic and CQ models

- Model parameter estimates

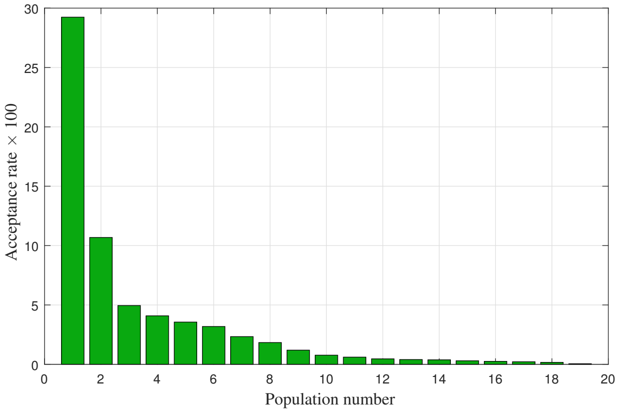
Parameter	Θ_{True}	μ	σ	95% CI
m	1	0.9965	0.0105	[0.97, 1.014]
c	0.05	0.0505	0.0062	[0.0406, 0.0611]
k	50	49.5090	2.0055	[46.25, 52.84]
k_3	10^3	1.069×10^3	393.23	$[430.38,\ 1.72\times10^3]$
k_5	10^5	9.58×10^4	1.96×10^4	$[6.49\times10^4,\ 1.29\times10^5]$

- Model prediction: excellent agreement



Example 1: Cubic and CQ models

- Significant decrease of the acceptance rate:



Outline

ABC based on Nested Sampling

- The **ABC-NS** algorithm is based on the concept of the *nested sampling* (NS) algorithm proposed by Skilling[Ref. 2] and the ellipsoidal sampling technique proposed by Mukherjee et al.[Ref. 3].

The implementation of the algorithm requires the selection of a number of hyperparameters, (the values used here are given):

- β_0: proportion of the alive particles ($\beta_0 = 0.6$).
- α_0: proportion of the died particles ($\alpha_0 = 0.3$).
- f_0: enlargement factor ($f_0 = 1.1$).

The first iteration of the ABC-NS algorithm is described in the pseudocode given in the next slide. For the next iteration, $(1 - \beta_0)N$ particles are sampled (subjected to the constraint $\Delta(u, u*) < \varepsilon_2$) inside the ellipse defined by its covariance matrix and mean value estimated based on the alive particles obtained previously. For the next iterations, the same procedure is repeated until the convergence criterion is met.

ABC based on Nested Sampling

- Principle of the ABC-NS algorithm:

Algorithm 1 ABC-NS SAMPLER

Require: u: Observed data, $\mathcal{M}(\cdot)$, ε_1, N

```
while i ≤ N do
    repeat
        Sample θ* from the prior distributions π(·)
        Simulate u* using the model M(·)
    until Δ(u, u*) ≤ ε₁
    set Θ_i = θ*, e_i = Δ(u, u*)
    set i = i + 1
end while
```

Sort e on ascending order, $\varepsilon_2 = e(\underline{\alpha_0}N)$, $\alpha_0 \in [0.2, 0.3]$

$\omega_i \propto \frac{1}{\varepsilon_1}\left(1 - (\frac{e_i}{\varepsilon_1})^2\right)$

Remove dead particles, $\Delta(u, u^*) \geq \varepsilon_2$, $\omega_{j=1:\alpha_0 N} = 0$

Normalise the weights such that $\sum_{i=1}^{(1-\alpha_0)N} \omega_i = 1$

Select $\beta_0 N$ alive particles, $\beta_0 \in [0.5, 1-\alpha_0]$

Define the ellipse by its covariance matrix and mean value $\mathcal{E}_1 = \{B_1, \mu_1\}$

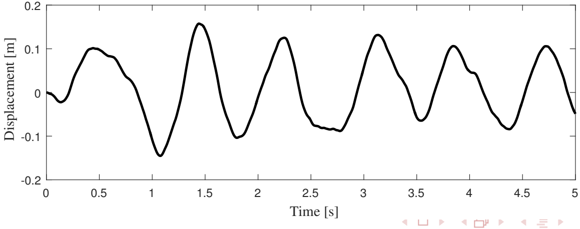
Example 2: ABC-NS for parameter estimation

1. The equation of motion is given by:

$$m\ddot{y} + c\dot{y} + \underline{k}y + \underline{k_3}y^3 = f(t)$$

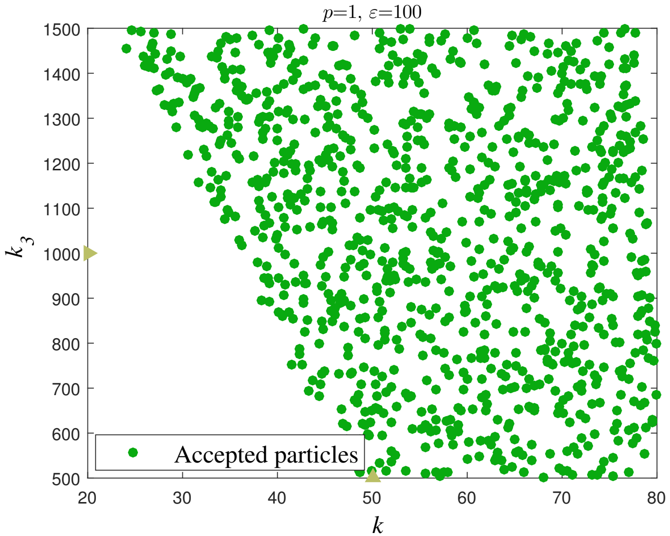
Objective: identify the unknown parameters $\Theta = (k, k_3)$ from the training data.

- $m = 1$.
- $c = 0.05$.
- $k \sim \mathcal{U}(20, 80)$.
- $k_3 \sim \mathcal{U}(500, 1500)$.
- Training data (free-of-noise).



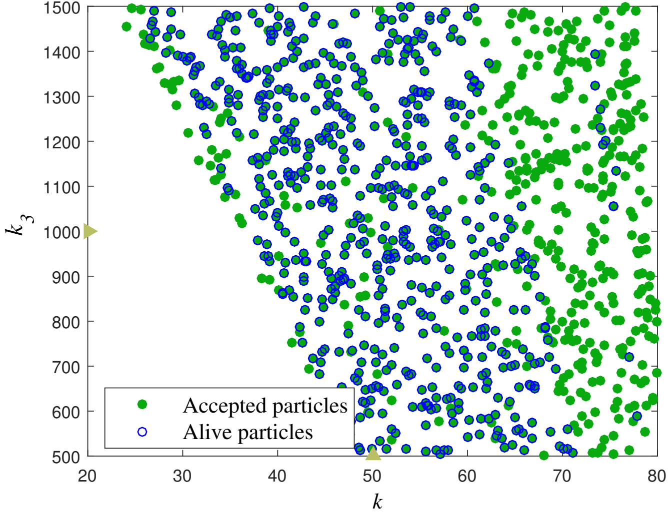
Example 2: ABC-NS for parameter estimation

- Illustration of the ABC-NS algorithm.



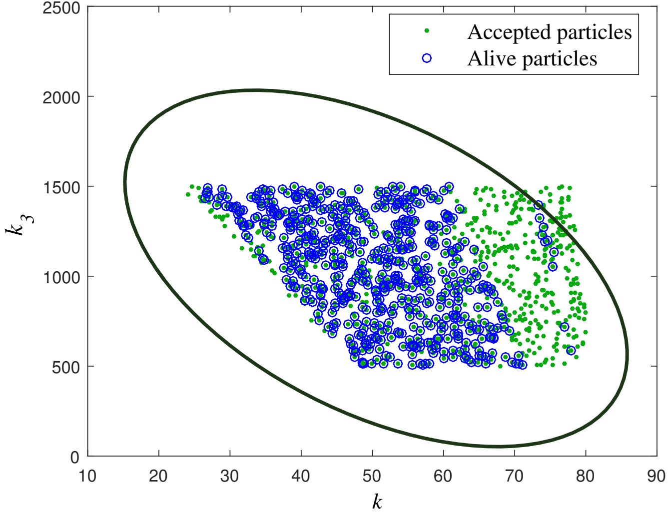
Example 2: ABC-NS for parameter estimation

- Illustration of the ABC-NS algorithm.



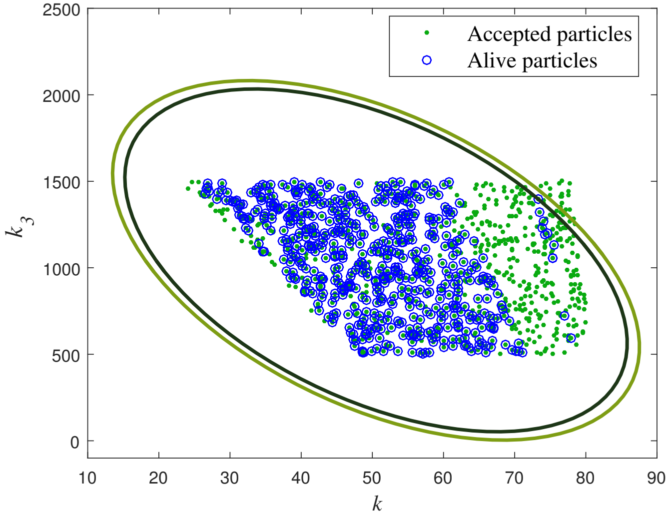
Example 2: ABC-NS for parameter estimation

- Illustration of the ABC-NS algorithm.



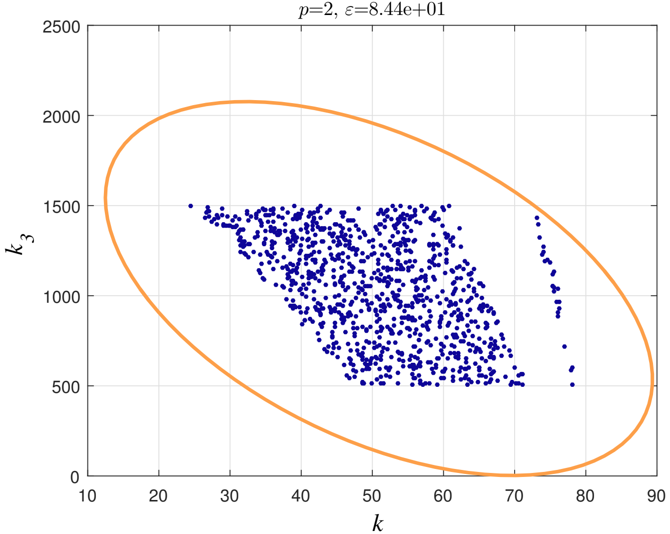
Example 2: ABC-NS for parameter estimation

- Illustration of the ABC-NS algorithm.



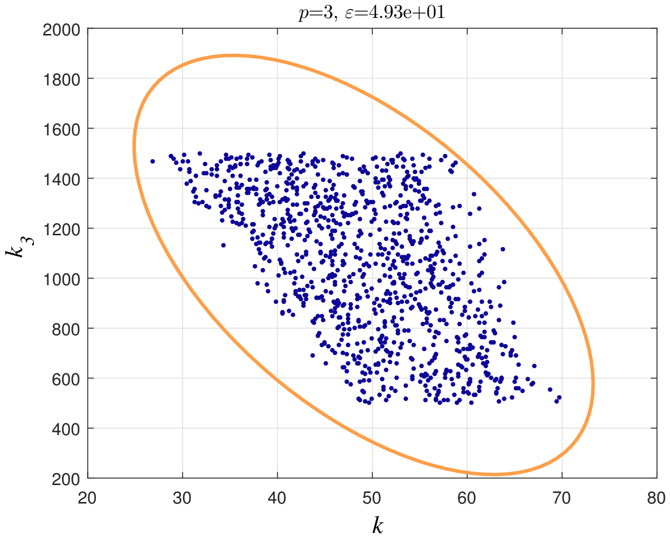
Example 2: ABC-NS for parameter estimation

- Illustration of the ABC-NS algorithm.



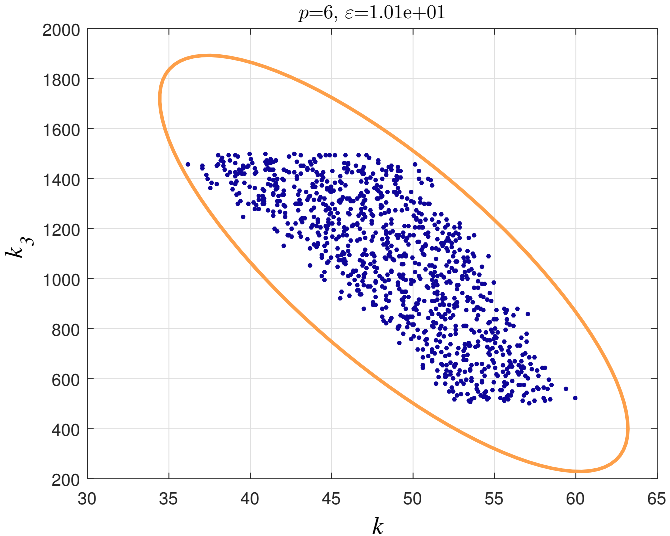
Example 2: ABC-NS for parameter estimation

- Illustration of the ABC-NS algorithm.



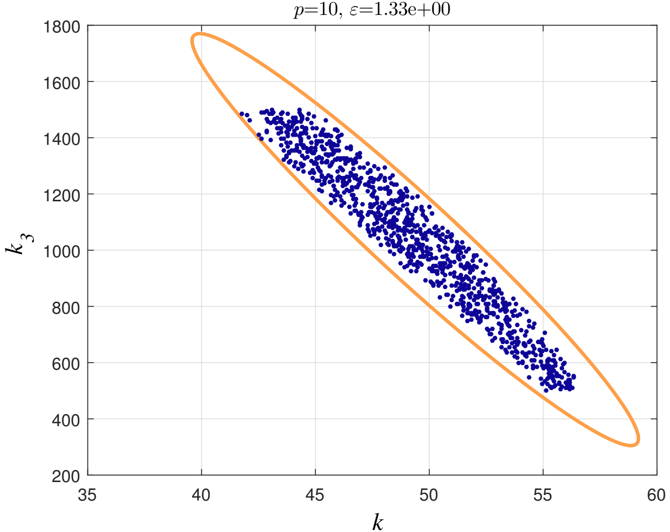
Example 2: ABC-NS for parameter estimation

- Illustration of the ABC-NS algorithm.



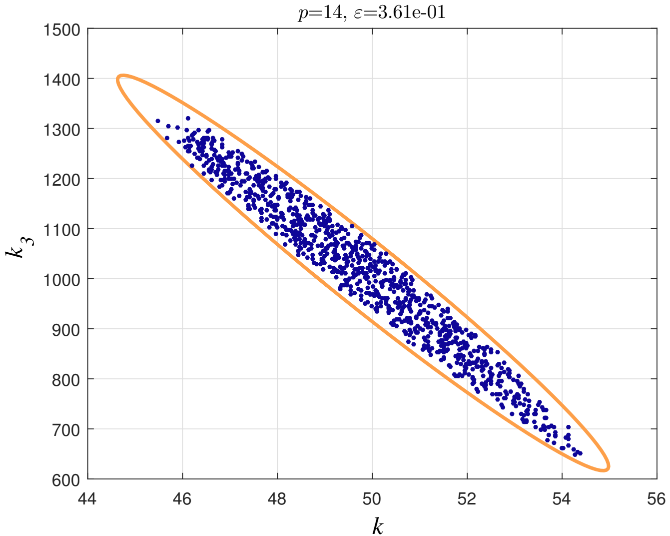
Example 2: ABC-NS for parameter estimation

- Illustration of the ABC-NS algorithm.



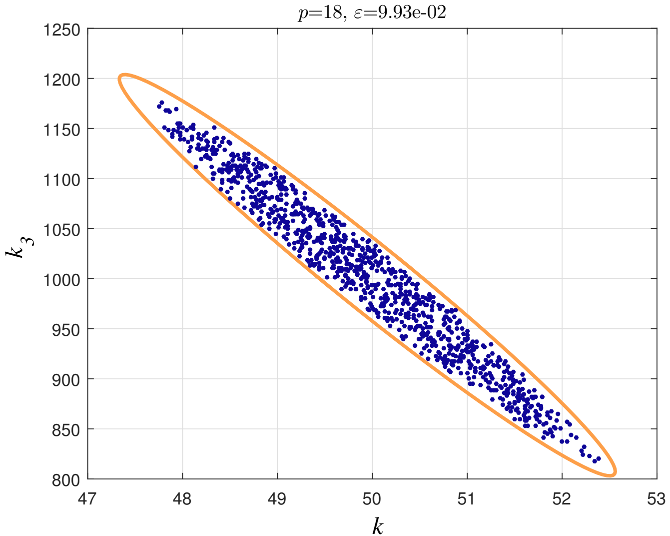
Example 2: ABC-NS for parameter estimation

- Illustration of the ABC-NS algorithm.



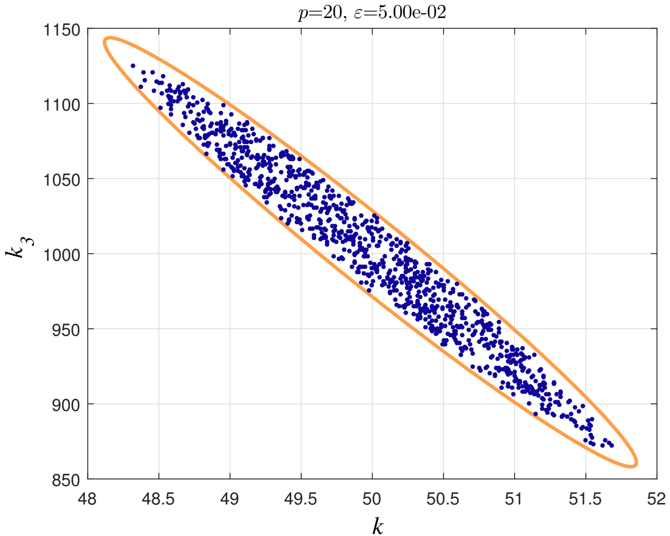
Example 2: ABC-NS for parameter estimation

- Illustration of the ABC-NS algorithm.



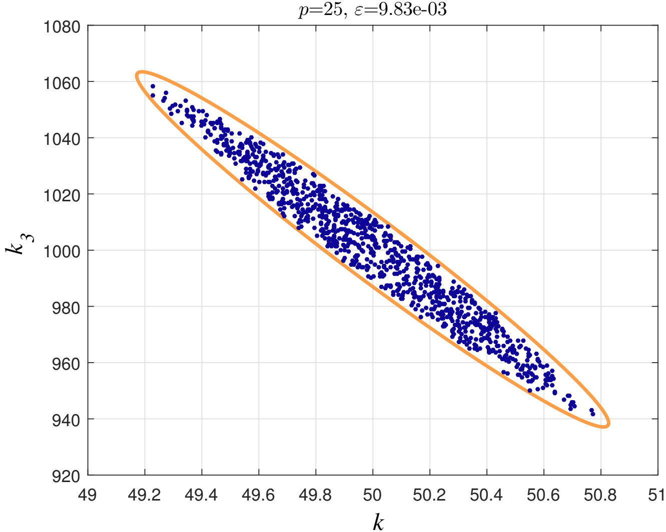
Example 2: ABC-NS for parameter estimation

- Illustration of the ABC-NS algorithm.



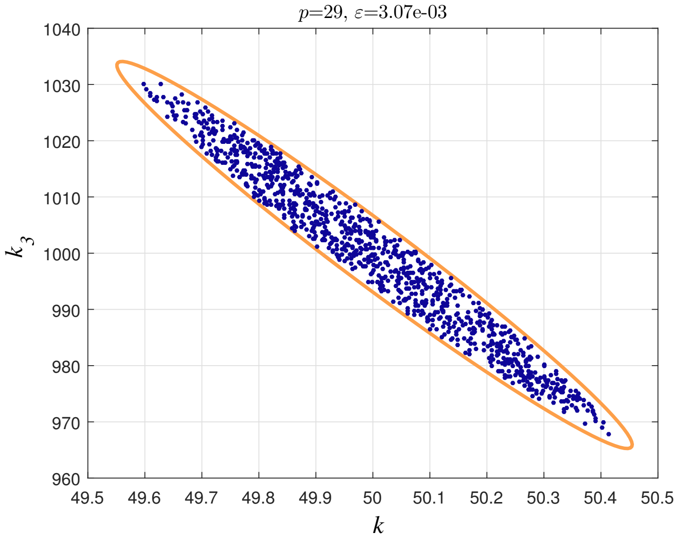
Example 2: ABC-NS for parameter estimation

- Illustration of the ABC-NS algorithm.



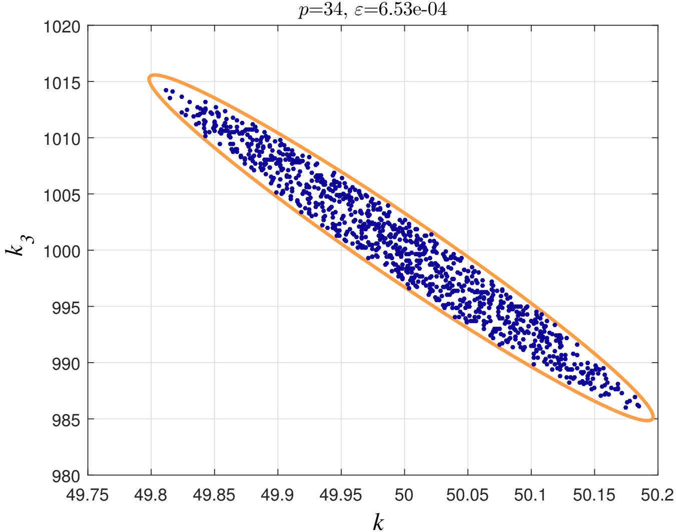
Example 2: ABC-NS for parameter estimation

- Illustration of the ABC-NS algorithm.



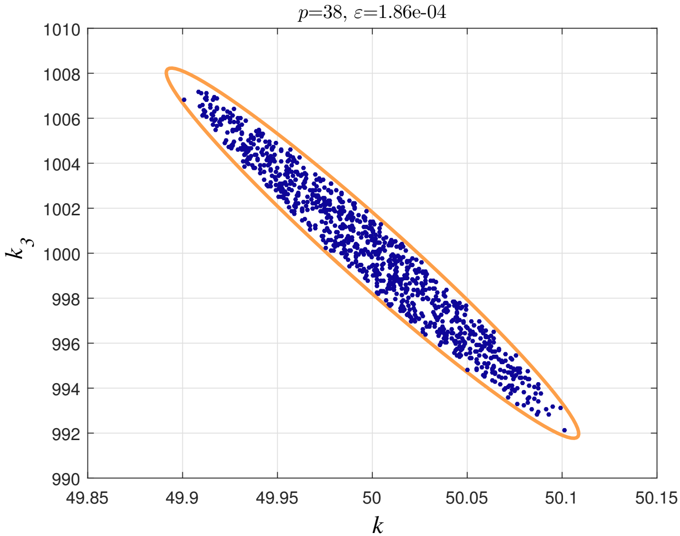
Example 2: ABC-NS for parameter estimation

- Illustration of the ABC-NS algorithm.



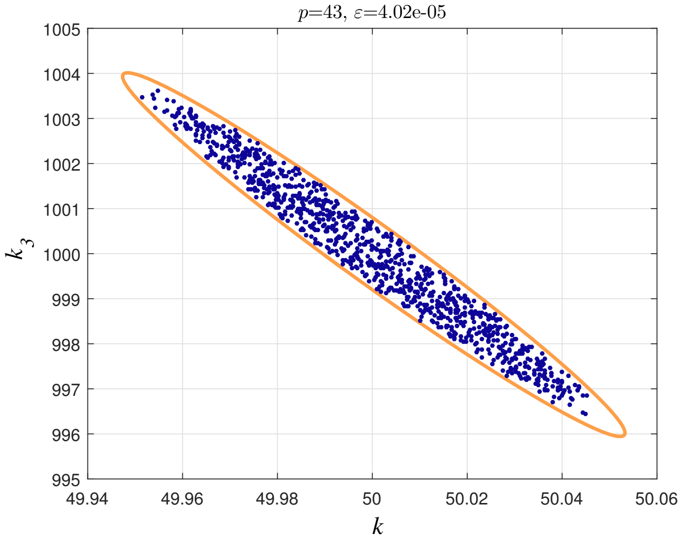
Example 2: ABC-NS for parameter estimation

- Illustration of the ABC-NS algorithm.



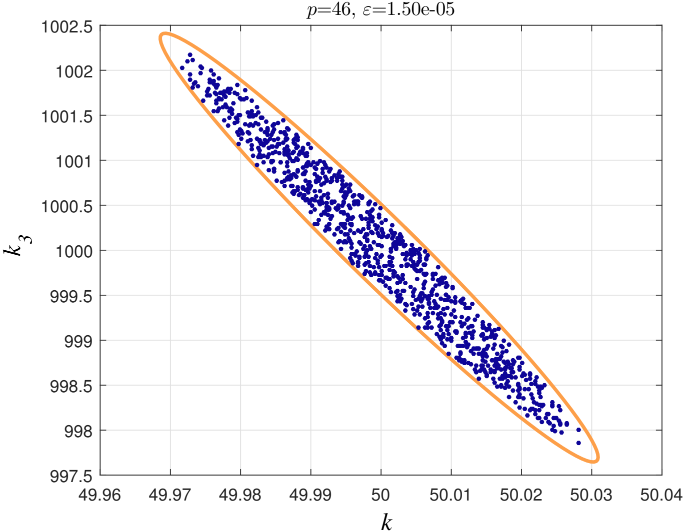
Example 2: ABC-NS for parameter estimation

- Illustration of the ABC-NS algorithm.



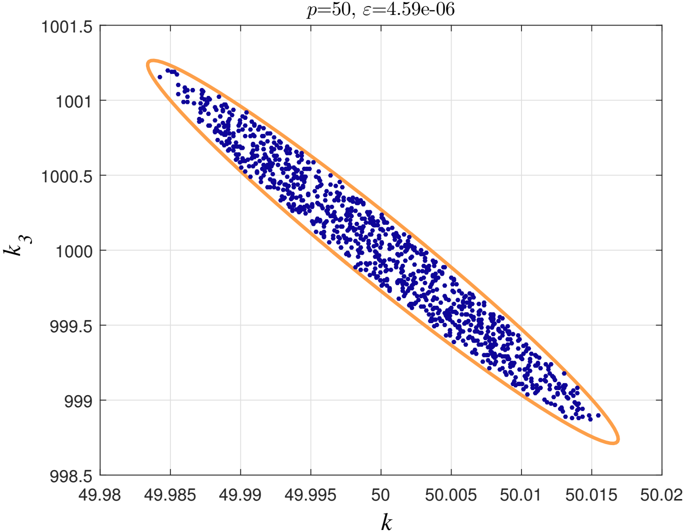
Example 2: ABC-NS for parameter estimation

- Illustration of the ABC-NS algorithm.



Example 2: ABC-NS for parameter estimation

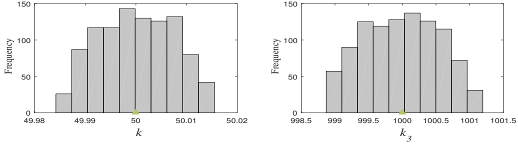
- Illustration of the ABC-NS algorithm.



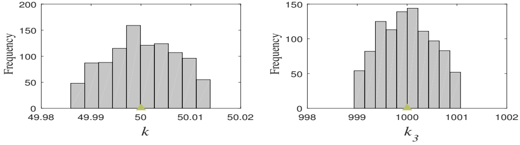
Example 2: ABC-NS for parameter estimation

→ Comparison between the posterior distributions obtained by using the:

- ABC-NS

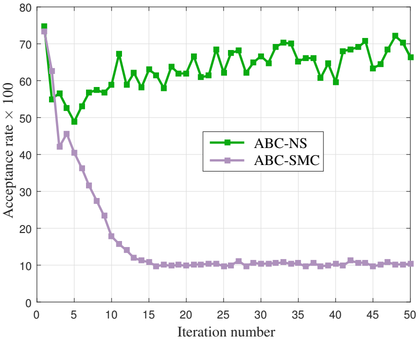


- ABC-SMC



Example 2: ABC-NS for parameter estimation

- Comparison between the acceptance rates.



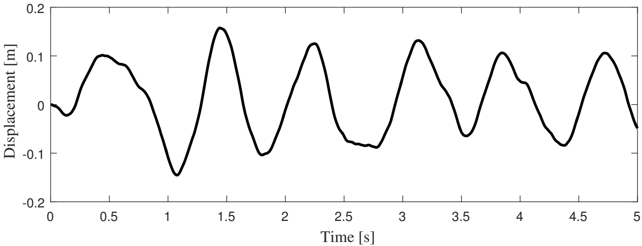
Example 3: ABC-NS for parameter estimation

1. The equation of motion is given by:

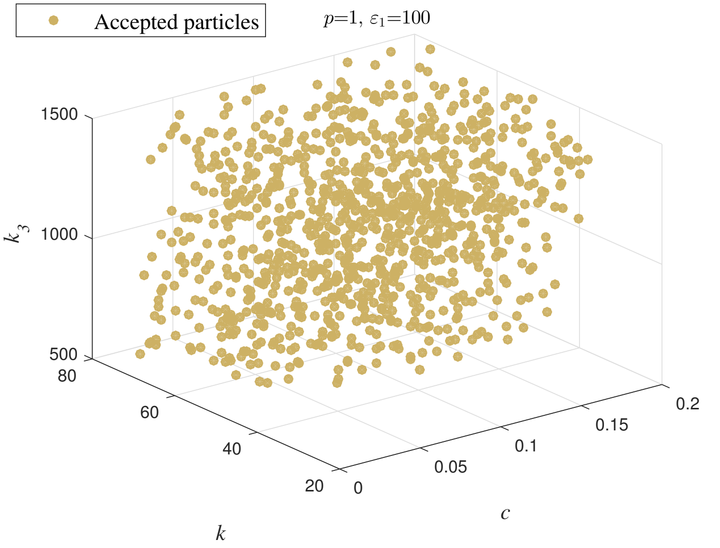
$$m\ddot{y} + \underline{c}\dot{y} + \underline{k}y + \underline{k_3}y^3 = f(t)$$

Objective: identify the unknown parameters $\Theta = (c, k, k_3)$ from the training data (free-of-noise).

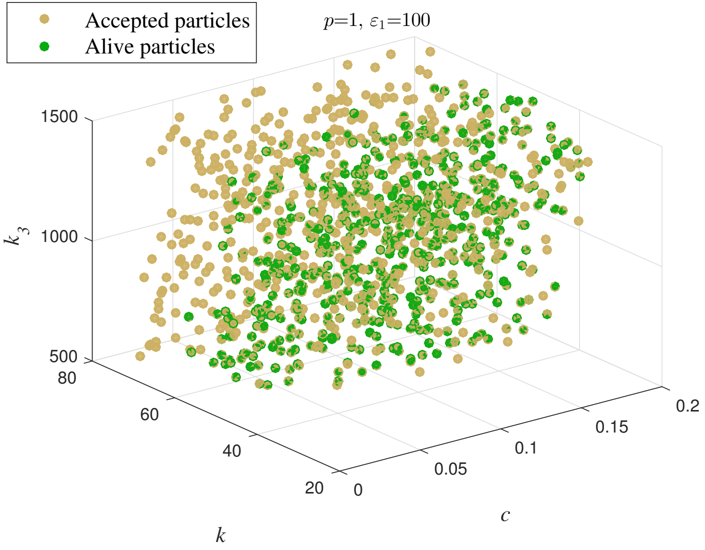
- $m = 1$.
- $c \sim \mathcal{U}(0.002, 0.2)$.
- $k \sim \mathcal{U}(20, 80)$.
- $k_3 \sim \mathcal{U}(500, 1500)$.



Example 3: ABC-NS for parameter estimation

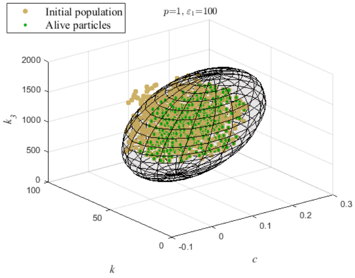


Example 3: ABC-NS for parameter estimation



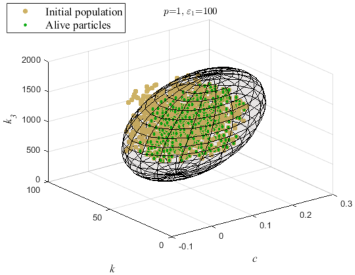
Example 3: ABC-NS for parameter estimation

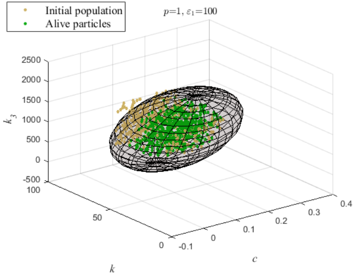
- Selection of the alive particles.



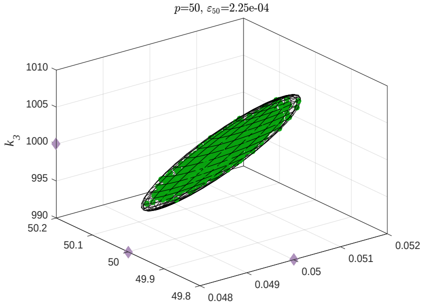
Example 3: ABC-NS for parameter estimation

- Selection of the alive particles.



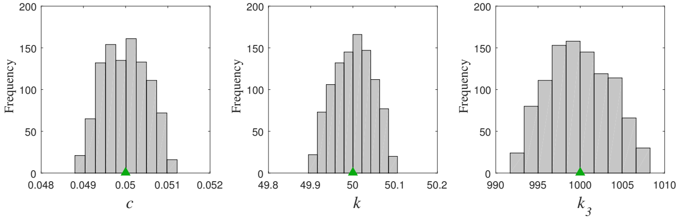


Example 3: ABC-NS for parameter estimation

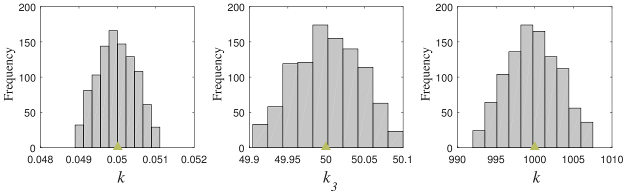


Example 3: ABC-NS for parameter estimation

- Histograms of the model parameters using ABC-NS.

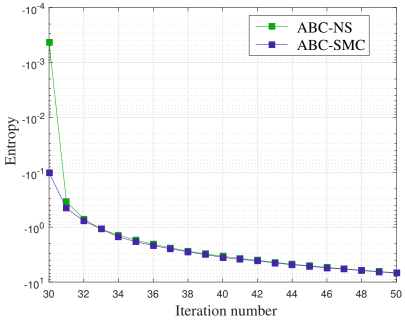


- Histograms of the model parameters using ABC-SMC.



Example 3: ABC-NS for parameter estimation

- Comparison between the posterior distributions using the differential entropy.



Example 3: ABC-NS for parameter estimation

- Comparison between the acceptance rates.

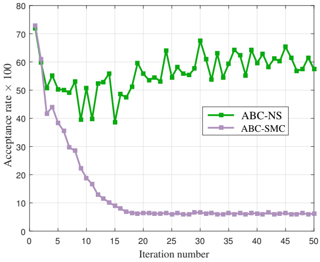


Figure: 3 unknown parameters

Example 3: ABC-NS for parameter estimation

- Comparison between the acceptance rates.

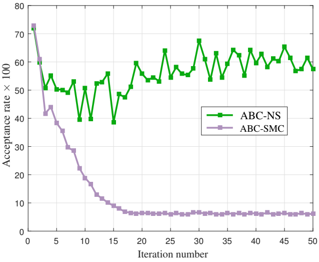


Figure: 3 unknown parameters

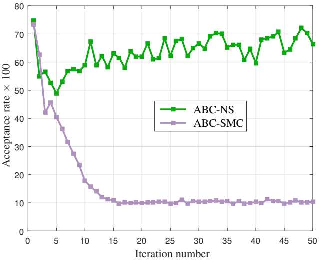


Figure: 2 unknown parameters

Example 4: ABC-NS for Model selection

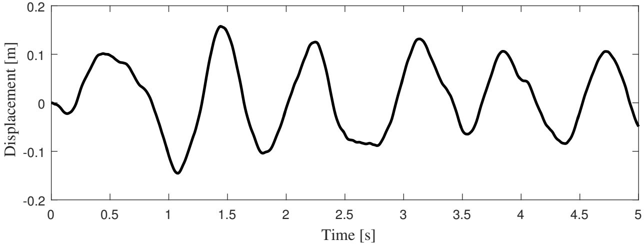
- Application of the **ABC-NS** for model selection.

→ Two competing models: each model is characterised by a set of parameters θ.

$$\begin{aligned}\mathcal{M}_1 &: \quad m\ddot{z} + c\dot{z} + kz = f(t),\\ \mathcal{M}_2 &: \quad m\ddot{z} + c\dot{z} + kz + k_3 z^3 = f(t).\end{aligned}$$

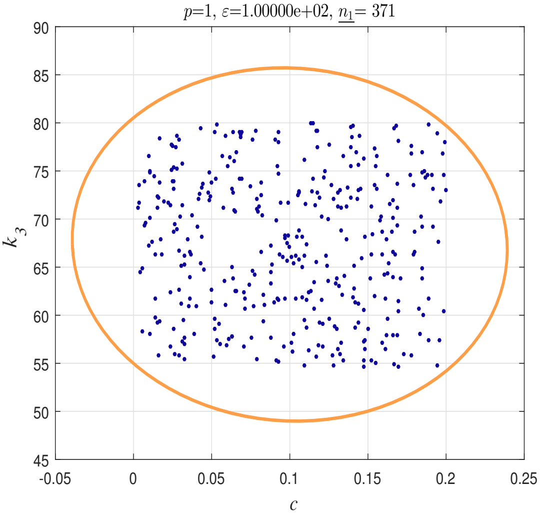
→ Objective: Identify the most likely model.

- The training data was synthetically generated from the cubic model (free-of-noise).



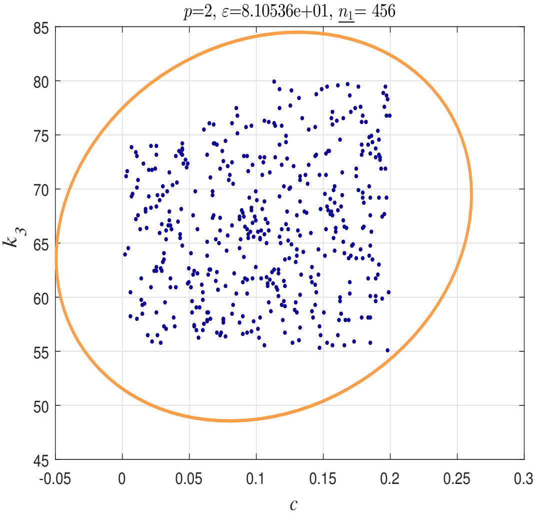
Example 4: ABC-NS for model selection

- Evolution of the number of particles for the linear model.



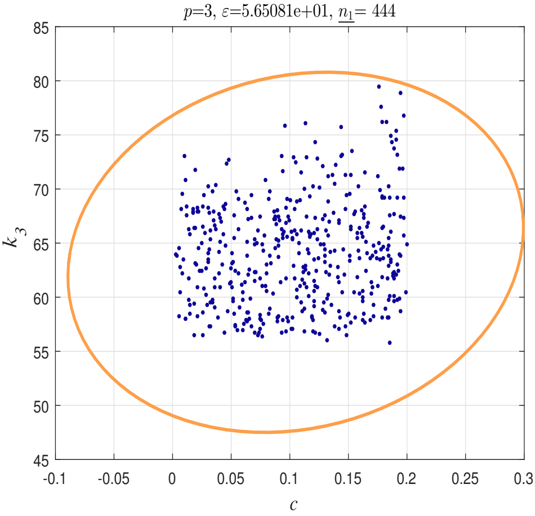
Example 4: ABC-NS for model selection

- Evolution of the number of particles for the linear model.



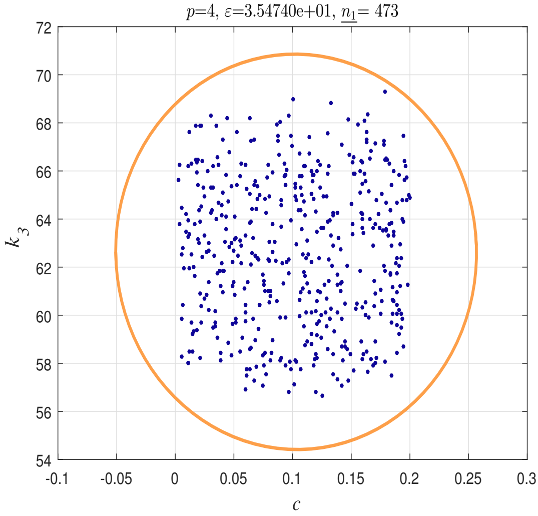
Example 4: ABC-NS for model selection

- Evolution of the number of particles for the linear model.



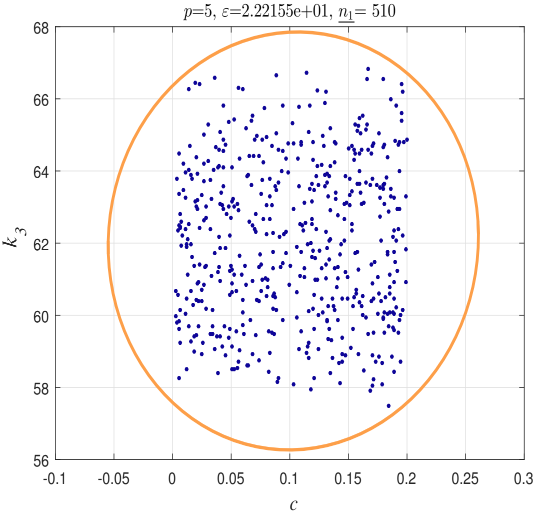
Example 4: ABC-NS for model selection

- Evolution of the number of particles for the linear model.



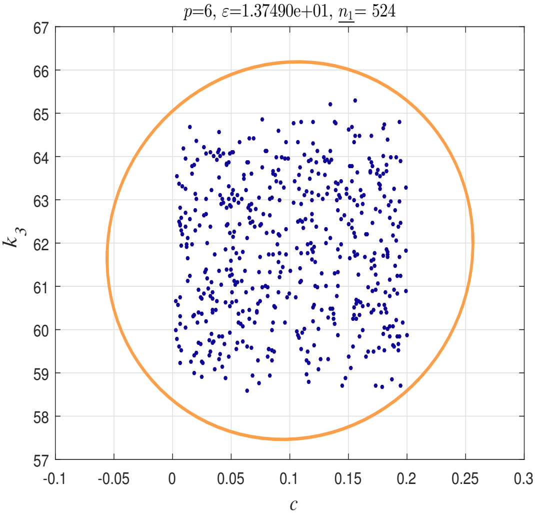
Example 4: ABC-NS for model selection

- Evolution of the number of particles for the linear model.



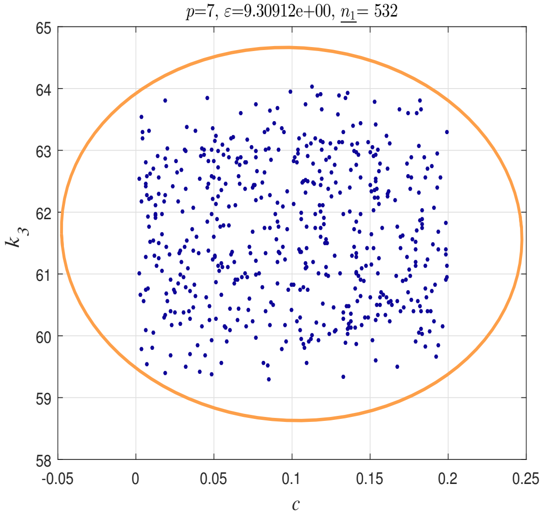
Example 4: ABC-NS for model selection

- Evolution of the number of particles for the linear model.



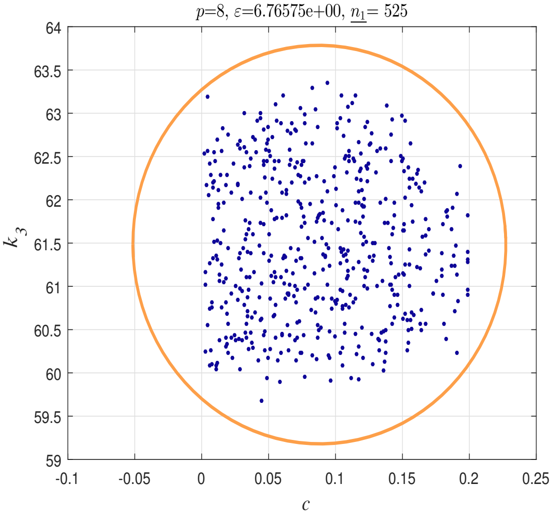
Example 4: ABC-NS for model selection

- Evolution of the number of particles for the linear model.



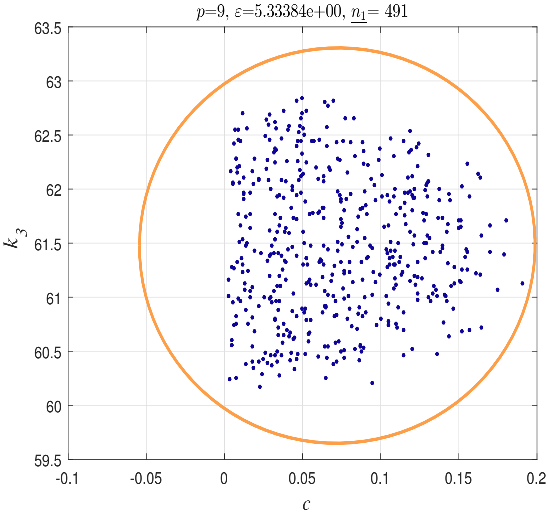
Example 4: ABC-NS for model selection

- Evolution of the number of particles for the linear model.



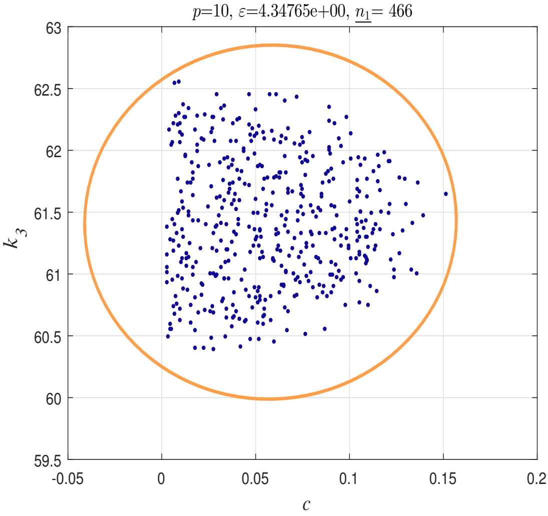
Example 4: ABC-NS for model selection

- Evolution of the number of particles for the linear model.



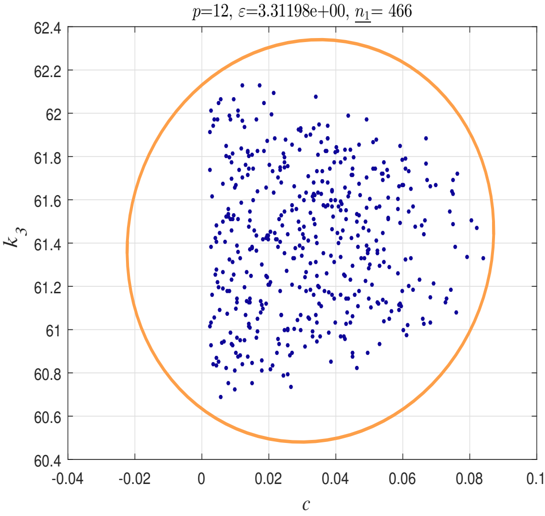
Example 4: ABC-NS for model selection

- Evolution of the number of particles for the linear model.



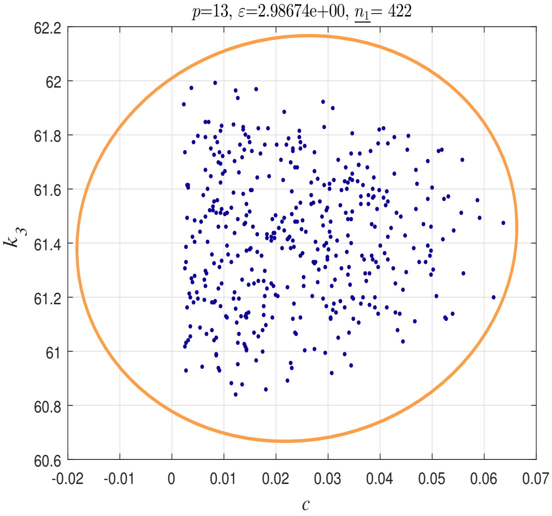
Example 4: ABC-NS for model selection

- Evolution of the number of particles for the linear model.



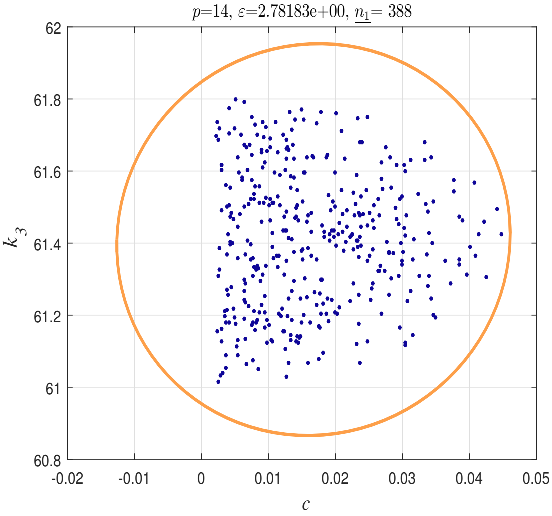
Example 4: ABC-NS for model selection

- Evolution of the number of particles for the linear model.



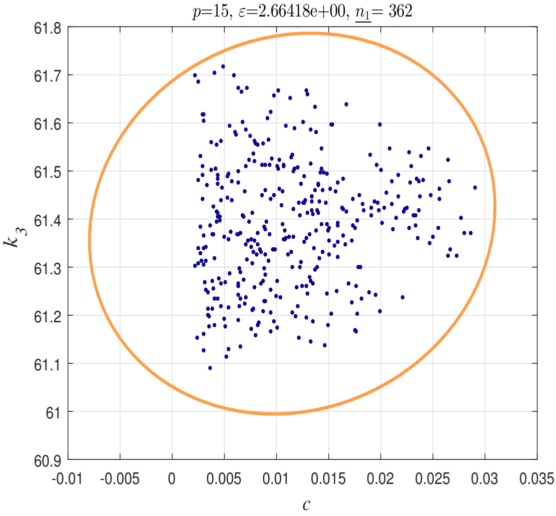
Example 4: ABC-NS for model selection

- Evolution of the number of particles for the linear model.



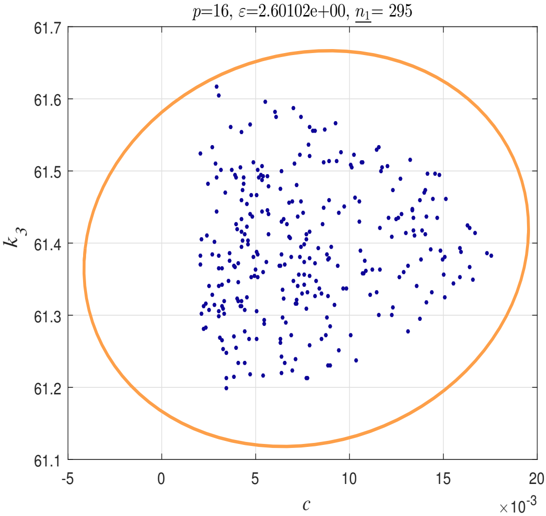
Example 4: ABC-NS for model selection

- Evolution of the number of particles for the linear model.



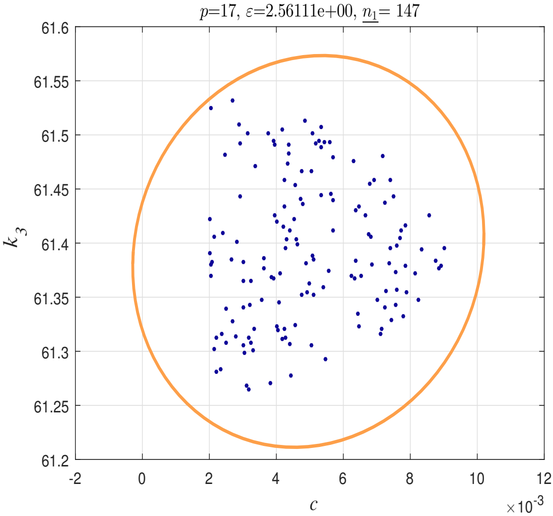
Example 4: ABC-NS for model selection

- Evolution of the number of particles for the linear model.



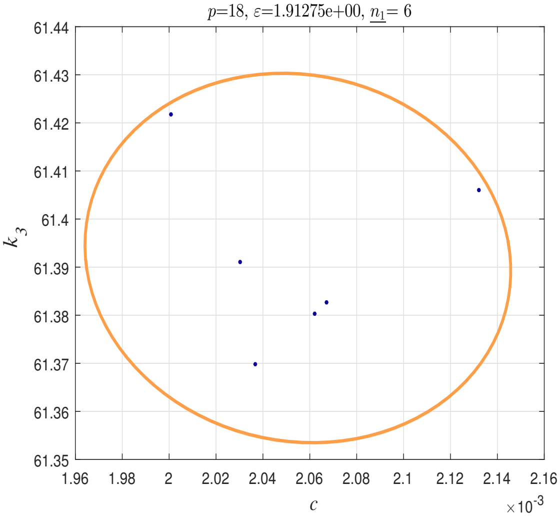
Example 4: ABC-NS for model selection

- Evolution of the number of particles for the linear model.



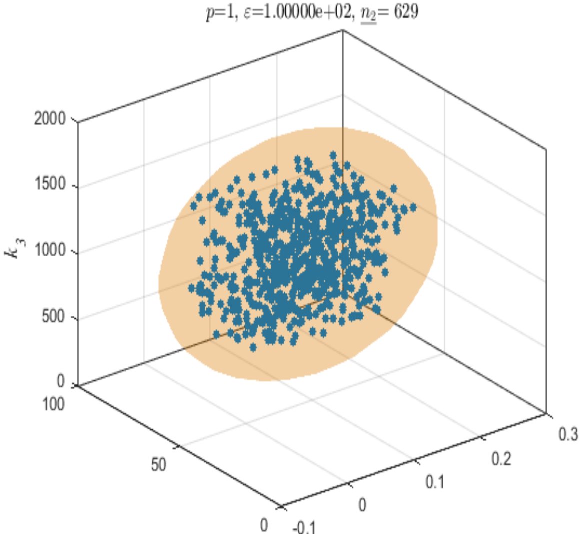
Example 4: ABC-NS for model selection

- Evolution of the number of particles for the linear model.



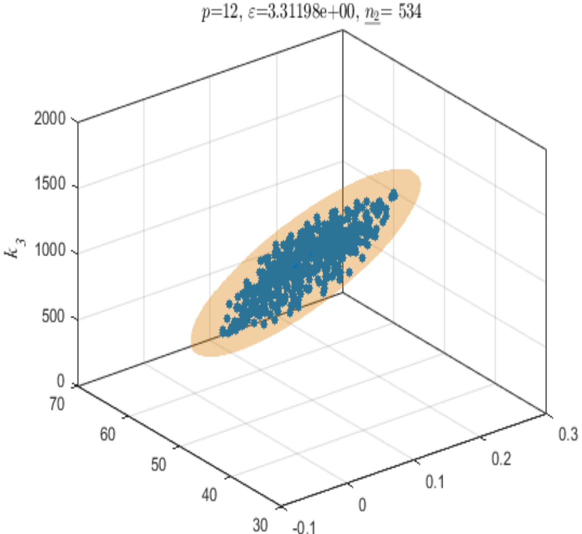
Model selection using ABC-NS

- Evolution of the number of particles for the cubic model.



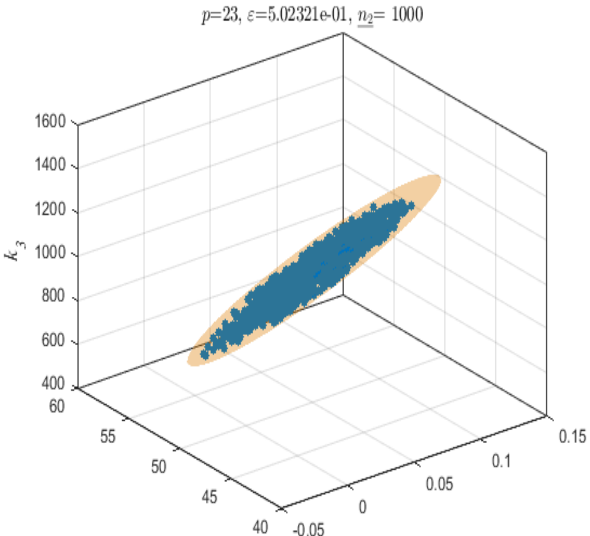
Model selection using ABC-NS

- Evolution of the number of particles for the cubic model.



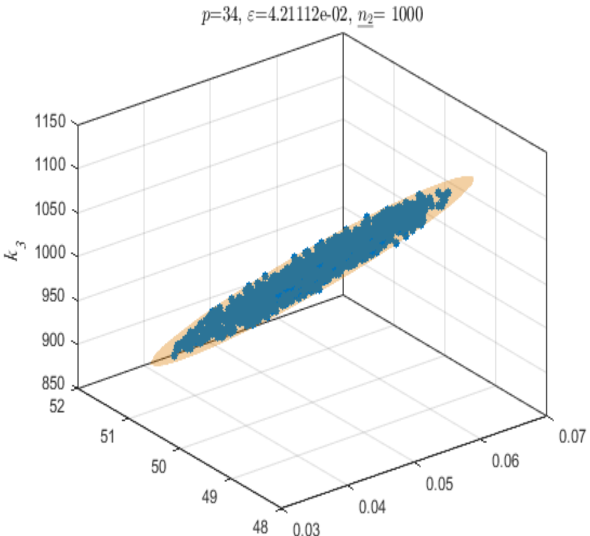
Model selection using ABC-NS

- Evolution of the number of particles for the cubic model.



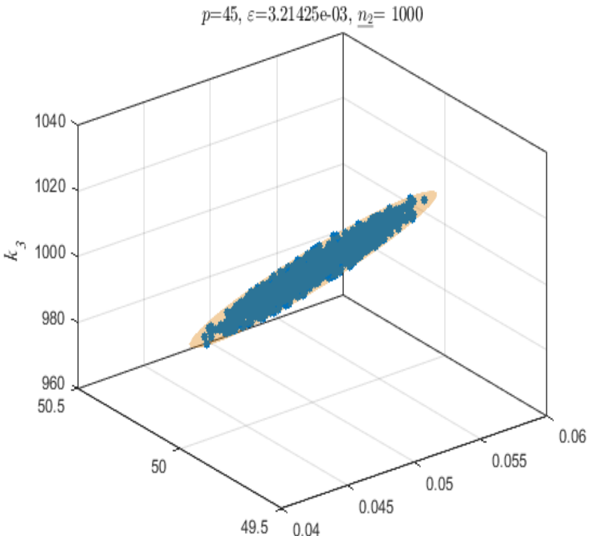
Model selection using ABC-NS

- Evolution of the number of particles for the cubic model.



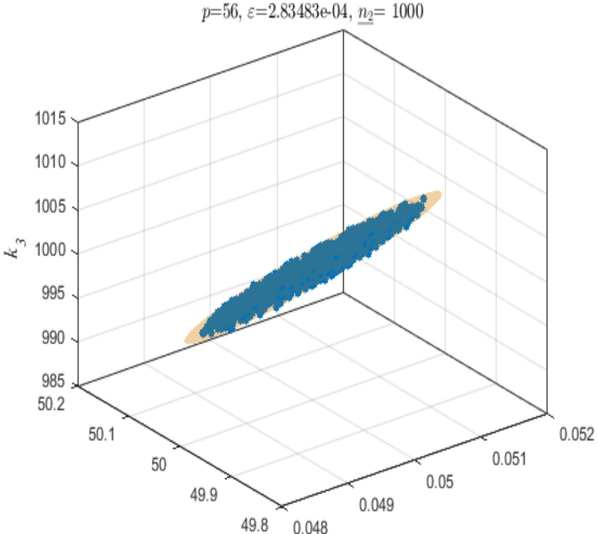
Model selection using ABC-NS

- Evolution of the number of particles for the cubic model.



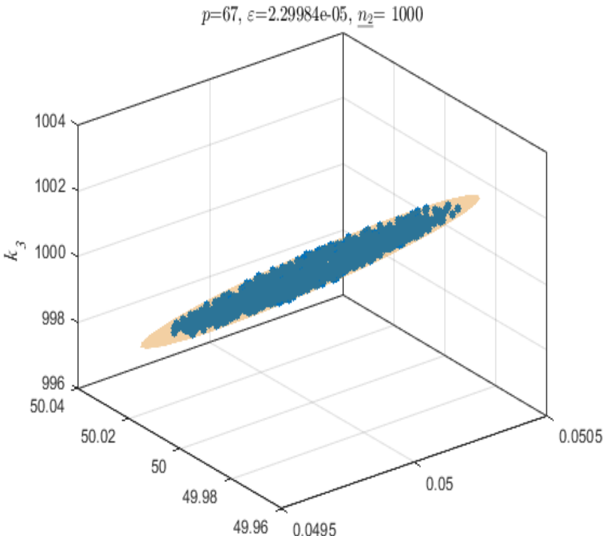
Model selection using ABC-NS

- Evolution of the number of particles for the cubic model.



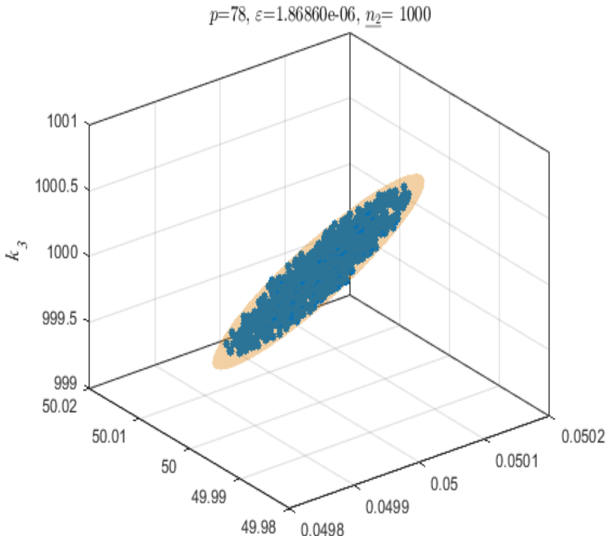
Model selection using ABC-NS

- Evolution of the number of particles for the cubic model.



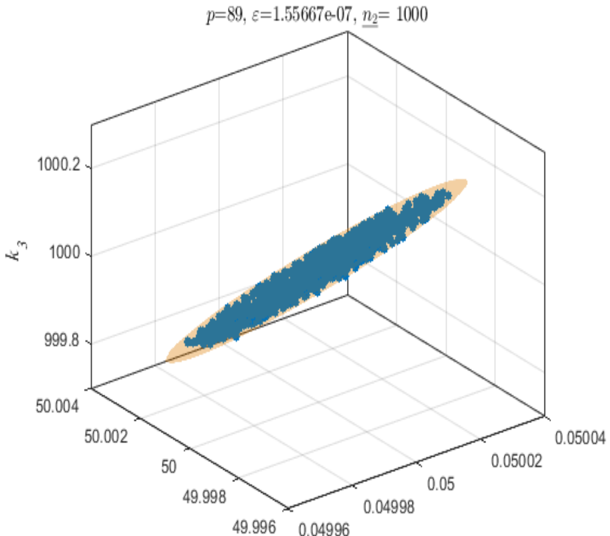
Model selection using ABC-NS

- Evolution of the number of particles for the cubic model.



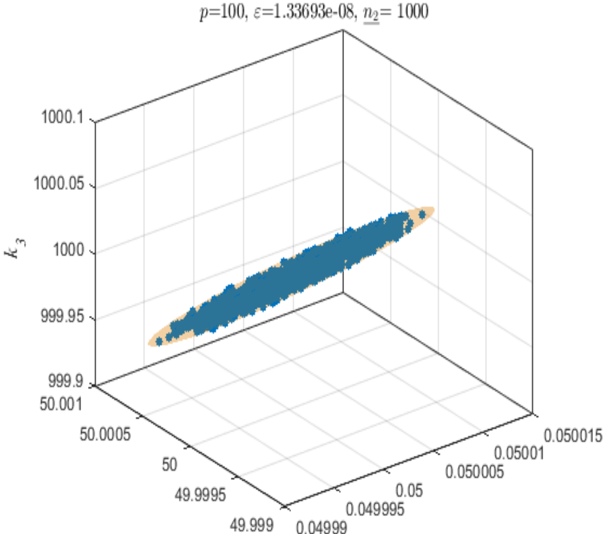
Model selection using ABC-NS

- Evolution of the number of particles for the cubic model.



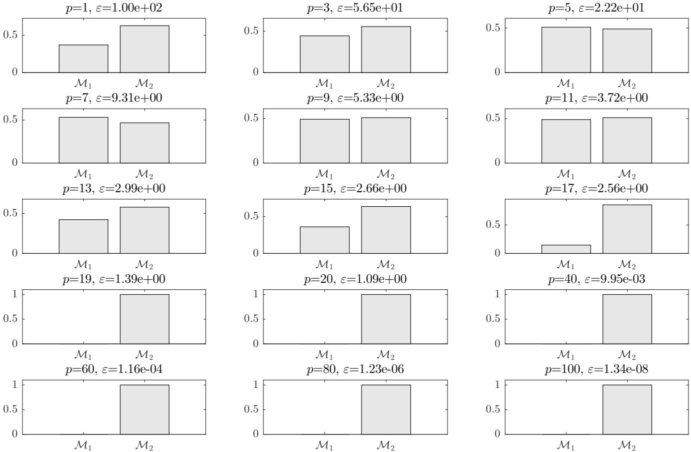
Model selection using ABC-NS

- Evolution of the number of particles for the cubic model.



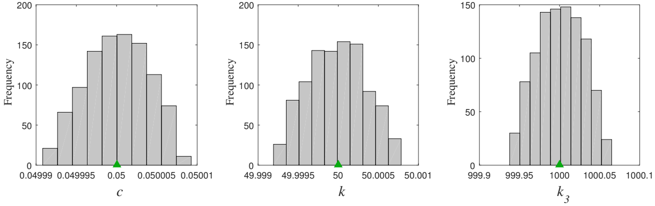
Example 4: ABC-NS for Model selection

- After few iterations, the linear model is eliminated and the algorithm converges to the right model: model posterior probabilities over few populations:

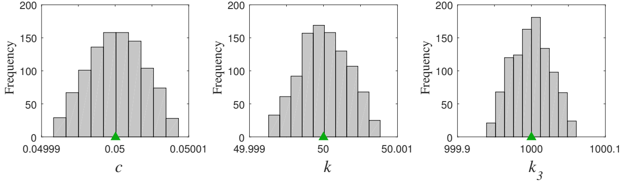


Example 4: ABC-NS for Model selection

- Histograms of the model parameters using ABC-NS.

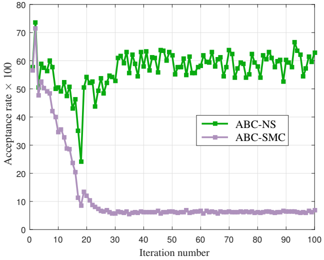


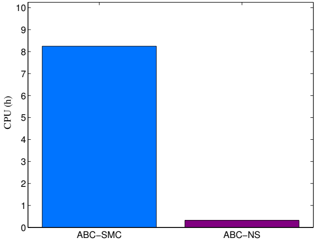
- Histograms of the model parameters using ABC-SMC.



Example 4: ABC-NS for Model selection

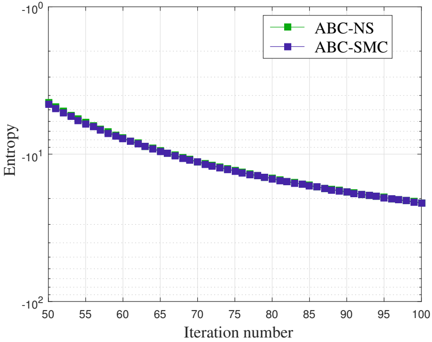
- Comparison between the acceptance rates and CPU time.





Example 4: ABC-NS for Model selection

- Comparison between the posterior distributions using the differential entropy.



Outline

1. Aims
2. The general Bayesian approach
3. ABC-SMC for model selection
4. ABC-NS as a new alternative
5. Conclusions

Conclusions

1. The parsimony principle is well embedded in the ABC algorithm.
2. ABC-SMC and ABC-NS provide good estimates.
3. The ABC-NS outperforms the ABC-SMC in terms of computational efficiency.
4. The ABC-NS can be applied simultaneously for parameter estimation and model selection.
5. ABC-NS requires less parameters to be tuned compared with the ABC-SMC.

- To run the ABC-NS algorithm, Matlab files `calc_ellipsoid.m` and `draw_from_ellipsoid.m` from the MultiNest package[Ref. 4] have been used.

Future work

- Compare the ABC-NS with other variants of ABC algorithms.
- Introduce the use of the ABC-NS for real applications.
- Extend the use of the ABC to infer systems with complex behaviours and systems with larger datasets by extracting useful features.

- More details and examples about the implementation of the ABC-NS algorithm and hyperparameters selection will be investigated and presented in a forthcoming journal paper.

References

T. Toni, D. Welch, N. Strelkowa, A. Ipsen and M.P. Stumpf.
Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems.
Journal of the Royal Society Interface, 6, 187–202.

J. Skilling.
Nested sampling for general Bayesian computation.
Bayesian Analysis, 1(4):833–860, 2006.

P. Mukherjee, D. Parkinson and A.R. Liddle.
A Nested Sampling Algorithm for Cosmological Model Selection.
The Astrophysical Journal Letters, 638(2):L51, 2006.

F. Feroz, M.P. Hobson and M. Bridges.
MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics.
Mon. Not. R. Astron. Soc., 398:1601–1614, 2009.

Acknowledgement

The authors gratefully acknowledge the support which has been provided through the EPSRC Programme Grant "Engineering Nonlinearity" (EP/K003836/1).

To cite this version:

Anis Ben Abdessalem, Keith Worden, Nikolaos Dervilis, David Wagg. Recent advances in approximate Bayesian computation methodology: application in structural dynamics. ENL Workshop, January 2017, Bristol, United Kingdom.