
Article:
Abdessalem, A.B., Dervilis, N., Wagg, D. et al. (2019) Model selection and parameter estimation of dynamical systems using a novel variant of approximate Bayesian computation. Mechanical Systems and Signal Processing, 122. pp. 364-386. ISSN: 0888-3270

https://doi.org/10.1016/j.ymssp.2018.12.048

Mechanical Systems and Signal Processing 122 (2019) 364–386

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Model selection and parameter estimation of dynamical systems using a novel variant of approximate Bayesian computation

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

Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, United Kingdom

ARTICLE INFO

Article history:
Received 10 December 2017
Received in revised form 6 September 2018
Accepted 21 December 2018
Available online 29 December 2018

Keywords:
Dynamical systems
Approximate Bayesian computation
Model selection
Nested sampling
Wire rope isolator
Nonlinearity

ABSTRACT

Model selection is a challenging problem that is of importance in many branches of the sciences and engineering, particularly in structural dynamics. By definition, it is intended to select the most likely model among a set of competing models that best matches the dynamic behaviour of a real structure and better predicts the measured data. The Bayesian approach which is based essentially on the evaluation of a likelihood function is one of the most popular approach to deal with model selection and parameter estimation issues. However, in some circumstances, the likelihood function is either intractable or not available even in a closed form. To overcome this issue, the likelihood-free or approximate Bayesian computation (ABC) algorithm has been introduced in the literature, which relaxes the need for an explicit likelihood function to measure the level of agreement between model predictions and measurements. However, ABC algorithms suffer from a low acceptance rate of samples which is actually a common problem with the traditional Bayesian methods. To overcome this shortcoming and alleviate the computational burden, a new variant of the ABC algorithm based on an ellipsoidal *Nested Sampling* (NS) technique is introduced in this paper; it has been called ABC-NS. Through this paper, it will be shown how the new algorithm is a promising alternative to deal with parameter estimation and model selection issues. It promises drastic speedups and provides a good approximation of the posterior distributions. To demonstrate its robust computational efficiency, four illustrative examples are given. Firstly, the efficiency of the algorithm is demonstrated to deal with parameter estimation. Secondly, two examples based on simulated and real data are given to demonstrate the efficiency of the algorithm to deal with model selection in structural dynamics.

1. Introduction

Model selection and parameter estimation still remain challenging issues for dynamicists, particularly for systems with complex behaviours (such as the existence of bifurcations and/or chaos, nonlinearities, etc). In model selection, the common route is not to identify the true underlying model but rather to find a model which is useful. Box and Draper [1] made the famous statement "All models are wrong but some are useful". Typically the usefulness of a model is measured by its ability

* Corresponding author at: LARIS, EA, 7315, Angers University, 62 Avenue Notre Dame du Lac, 49000 Angers, France.
E-mail address: anis.ben-abdessalem@univ-angers.fr (A. Ben Abdessalem).

https://doi.org/10.1016/j.ymssp.2018.12.048
0888-3270/

to make predictions about unseen observations. In this sense, there is no true model (in the absence of simulated data); there is only a model from a set of competing models which performs better. This is the model that explains the data in an accurate and parsimonious way with minimal model complexity.

In structural dynamics, it is always the case that the number of possible models that could be used to explain a particular data set could be very large. Therefore, it is often desirable to compare models to see whether all components are necessary and then select the model with the highest evidence. This suggests the need of a sophisticated statistical tool which can be used to evaluate the performance of the competing models and then provide a formal rank. To solve this challenging problem, the Bayesian approach has been successfully applied in different domains (dynamics, genetics, biology, ecology, etc). Compared with other classical methods specifically the ones based on the evaluation of an information criterion, the Bayesian approach is more informative in the sense that one gets the full distributions of model parameters. For more details about the implementation of the Bayesian method for parameter estimation and model selection, the reader is referred to [2–9] and the references therein.

Before presenting the approximate Bayesian computation (ABC) or likelihood-free algorithms for model selection, the main methods and techniques which have been proposed in the literature to deal with model selection are discussed. The standard approach to compare between a set of competing models, is based on their ability to reproduce experimental data. Basically, the model comparison involves the definition of a suitable metric of fit, e.g., the sum of squared residuals or the euclidean distance for instance, and selecting a model that minimises this metric. The principal limitation of this approach is that it only identifies a single best model and yields point estimates of the underlying parameter values, without providing a meaningful description of the uncertainty among competing models and in their parameters. In addition, the estimated parameter values are usually local optima of the fitting metric. Bayesian model inference overcomes the limitation of a least-squares fitting approach by providing a rigorous method using available experimental data with prior knowledge, to yield a fuller description of model and parameter uncertainties. Classical information criterions have been widely used to deal with model selection, such as the Akaike information criterion (AIC) [10] and the Bayesian information criterion (BIC) [11]. Another still popular IC is the deviance information criterion (DIC) proposed by [12]. Several other (ICs) have been proposed in the literature and the question of which IC should be used in model selection is not a trivial task and is a still a matter of debate. To enforce parsimony (simpler models should be preferred as they generalise better), most of the mentioned ICs introduce an *ad hoc* penalty term while in ABC, the parsimony principle is naturally embedded as will be shown in the presented examples. The Reversible-Jump Markov-chain Monte Carlo (RJ-MCMC) algorithm is one of the methods which has been widely used in the literature; however, its major drawback is how to deal with multiple competing models with different dimensionalities. For a deep discussion and theory of RJ-MCMC, the reader is referred to the following paper [13]. In [14], a recently proposed MCMC type posterior probability sampling algorithm called TMCMC [15] was implemented. This algorithm estimates the model evidence by sampling the posterior probability distribution of the model by a sequence of non-normalised intermediate probability functions. It should be noted that an improved versions of the TMCMC algorithm have been proposed in the literature to correct the bias in the evidence observed in the original algorithm. For more details, the reader may refer to [16,17]. Skilling [18,19] proposed the nested sampling algorithm, which is able to estimate efficiently the evidence; it has been widely applied in several domains [20–24].

The application of the Bayesian approach requires the definition of a likelihood function to measure the level of agreement between the observed and simulated data. However, in some circumstances the likelihood function might not be available in a closed form. To overcome this issue, and make possible the inference of complex systems, the ABC algorithm has been introduced. ABC offers the possibility of using different features and metrics to measure the similarity between simulated and observed data to infer a given system. For this reason, ABC has attracted attention in a wide variety of applied disciplines (e.g., biology, psychology, genetics, machine learning to mention just a few [25–27]) and recently in structural dynamics for parameter estimation [28,29] and model selection [30,29,31,32]. The ABC algorithm has several advantages compared to the existing methods: (i) it is easy to understand and to implement, (ii) no burn-in period for most of the variants and no parameter distribution filtering is necessary as the a posteriori distributions are directly given at the last running step and (iii) offers the possibility to compare between a set of competing models simultaneously.

Several variants of the ABC algorithm have been proposed in the literature, including ABC based on Markov chain Monte Carlo sampling [33] and ABC based on sequential Monte Carlo (proposed by Sisson et al. [34]), which has proven to be more efficient than [33]. It should be noted that ABC was introduced initially to infer model parameters and then was extended to deal with model selection [35]. One common issue with the existing ABC algorithms which has been widely reported in the literature is the acceptance rate which decreases dramatically along the iterations (see, [36,37] for instance). Therefore, the computational requirements exponentially increase which is a major drawback.

In this paper, a new ABC algorithm based on the idea of an ellipsoidal nested sampling technique [21] will be proposed to overcome this issue, and it has been named ABC-NS. In the proposed algorithm, instead of removing one particle, as in the traditional nested sampling algorithm [18], a proportion of particles are removed in each iteration (called the population in the ABC jargon) based on assigned weights. In this study, different examples dealing with parameter estimation and model selection issues (using simulated and real data) have been proposed to demonstrate the computational efficiency. It has been shown that the new algorithm maintains a high acceptance rates over the populations, reduce the overall CPU time and provide a good precision on the posterior estimates. The gain in terms of computational efficiency offers the possibility to infer several competing models with a large number of parameters compared with other variants. To demonstrate the computational efficiency of the ABC-NS algorithm over the ABC-SMC, a numerical example is presented.

The rest of this paper is structured as follows. Section 2 starts out with the basics of the ABC algorithm and presents in detail the new variant of ABC algorithms. Section 3 presents two examples to demonstrate the efficiency of the novel algorithm to deal with parameter estimation. In Sections 4 and 5, two examples have been selected to demonstrate the efficiency of the algorithm in solving the model selection issue and form the core of the paper. The main conclusions are given in Section 6.

2. Approximate Bayesian computation

2.1. Basic theory

In the Bayesian method, the posterior probability density, $p(\boldsymbol{\theta}|u)$ given observed data u and a model $\mathcal{M}$, can be computed using Bayes' Theorem:

$$p(\boldsymbol{\theta}|u) = \frac{p(\boldsymbol{\theta})\mathcal{L}(u|\boldsymbol{\theta})}{\int_{\theta} p(\boldsymbol{\theta})\mathcal{L}(u|\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}} \propto p(\boldsymbol{\theta})\mathcal{L}(u|\theta) \tag{1}$$

where $p(\boldsymbol{\theta})$ is the prior probability of $\boldsymbol{\theta}$ (the vector of model parameters) and $\mathcal{L}(u|\boldsymbol{\theta})$ is the likelihood function. The denominator is a normalising constant.

However, as mentioned earlier, explicit forms for likelihood functions are rarely available. In addition, the evaluation of the integral in Eq. (1) is sometimes numerically difficult. The ABC methods remove the likelihood by evaluating the discrepancy between the observed data and the data generated by a simulation using a given model. It should be noted that the ABC offers the advantage to compare summary statistics instead of using the raw data. In the framework of ABC, an approximate form of the Bayes' Theorem is given by:

$$p(\boldsymbol{\theta}|u) \propto p(\Delta(u,u^*) < \varepsilon|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{2}$$

where $u^* \sim f(\cdot|\boldsymbol{\theta})$ is the simulated data, $\Delta(\cdot)$ is a discrepancy metric, $p(\Delta(u,u^*) < \varepsilon|\boldsymbol{\theta})$ is a distribution that measures how similar u and u^* and $\varepsilon > 0$ is a tolerance threshold (when ε tends towards 0, the approximated posterior distribution is a good approximation of the "true" posterior distribution).

The Bayesian method allows both levels of inference (i.e., parameters and models). For model comparison, the Bayesian method relies on the computation of posterior model probabilities $p(\mathcal{M}_j|u)$. Applying Bayes's rule to a set of models, the posterior probability of model $\mathcal{M}_j$ is given by:

$$p(\mathcal{M}_j|u) \propto p(\mathcal{M}_j)p(u|\mathcal{M}_j,\boldsymbol{\theta}_j) \propto p(\mathcal{M}_j)\int p(u|\mathcal{M}_j)p(\boldsymbol{\theta}_j|\mathcal{M}_j)\mathrm{d}\boldsymbol{\theta}_j \tag{3}$$

where $p(\mathcal{M}_j)$ is the prior model probability, $p(u|\mathcal{M}_j)$ is the evidence of model $\mathcal{M}_j, p(\boldsymbol{\theta}_j|\mathcal{M}_j)$ is the prior distributions of parameters and $\boldsymbol{\theta}_j$ is the vector of parameters in model $\mathcal{M}_j$.

The most simple implementation of the ABC algorithm is ABC rejection sampling denoted by ABC-RS and illustrated in Algorithm 1. While ABC-RS is simple to implement, it can be computationally prohibitive in practice. To overcome this problem, other variants have been proposed in the literature mentioned previously. In the next section, the new ABC algorithm is presented. It should be noted that in the ABC algorithms in general, the identification strategy starts at a coarse resolution (higher initial ε_1 value), which then is adaptively refined until a target tolerance threshold value defined by the user is reached, giving a gradually finer representation of the model parameters estimation or model probabilities.

Algorithm 1 ABC-RS

Require u: observed data, $\mathcal{M}$: model, ε_1
1: **while** $i \leqslant N$ **do**
2: **repeat**
3: generate $\boldsymbol{\theta}^*$ from the prior distribution $p(\cdot)$
4: simulate u^* using the model $\mathcal{M}(\cdot)$
5: **until** $\Delta(u^*,u) < \varepsilon_1$
6: **set** $\boldsymbol{\Theta} = \boldsymbol{\theta}^*$
7: **end for**

2.2. ABC-NS implementation

In this section, a detailed description of the novel ABC algorithm is given. The ABC-NS algorithm broadly works following the same scheme as the ABC-SMC algorithm in [35]. The main novelties are in (i) the way of sampling, (ii) the weighting technique adopted from [25] and (iii) instead of dropping one particle per iteration, a proportion of particles is dropped based on the assigned weights, which speeds-up the algorithm without compromising the precision on the posterior esti-

mates. The iterative process is detailed in Algorithm 2. The algorithm starts by generating N particles from the prior satisfying the constraint $\Delta(u,u^*) < \varepsilon_1$ (here, u for observed data and u^* for simulated data). The accepted particles are then weighted (see, step 9) and the next tolerance threshold is defined based on the discrepancy values ranked in descending order (highest on top, see, step 11) as the $(\alpha_0 N)^{\text{th}}$ value where α_0 is the proportion of dropped particles defined by the user. Then, one assigns a weight of zero to the dropped particles. After that, the weights of the remaining particles are normalised (see, step 13). From the remaining particles, one selects $\beta_0 N$ particles based on the updated weight values, where β_0 is the proportion of particles, so-called "surviving" particles (see step 14). The surviving particles are then enclosed in an ellipsoid in which the mass center μ and the covariance matrix $\mathcal{C}$ are estimated based on the remaining particles; one denotes this ellipsoid by $\mathcal{E} = (\mu, \mathcal{C})$. The generated ellipsoid could be enlarged by a factor f_0 as mentioned in step 16 to ensure that the particles on the borders are inside. It should be noted that ellipsoidal sampling was firstly proposed in [21] to improve the efficiency of the nested sampling algorithm which has been widely used for Bayesian inference, mainly in cosmology [38]. Finally, the population is replenished by resampling $(1-\beta_0)N$ particles inside the enlarged ellipsoid (see step 20) following the scheme in [39] and a re-weighting step is carried out (see step 28). The procedure is repeated until a stopping criterion defined by the user is met. For the ABC-NS algorithm, the mean value and the covariance matrix of the "alive" particles are given by:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4}$$

$$\mathcal{C} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T \tag{5}$$

where n is the number of "alive" particles and x_i denotes the vector of those particles.

In the framework of this work and in the considered examples, the hyperparameters are selected as follows: the number of samples is set to 1000, α_0, β_0 and f_0 are set to 0.3, 0.6 and 1.1, respectively. It has been shown that the selected tuning parameters work quite well for the considered examples to maintain relatively high acceptance rates over the iterations. However, of course they can be optimised via common machine learning tools. In this work a brief discussion on the effects of the tuning parameters on the computational and statistical efficiencies is given.

Algorithm 2 ABC-NS SAMPLER

Require u: observed data, $\mathcal{M}(\cdot)$: model, ε_1, N, α_0, β_0, f_0
1: **set** $t = 1$
2: **for** $i = 1, \ldots, N$ **do**
3: **repeat**
4: Sample $\boldsymbol{\theta}^*$ from the prior distributions $p(\cdot)$
5: Simulate u^* using the model $\mathcal{M}(\cdot)$
6: **until** $\Delta(u,u^*) < \varepsilon_1$
7: **set** $\boldsymbol{\Theta}_i = \boldsymbol{\theta}^*$, $e_i = \Delta(u,u^*)$
8: **end for**
9: Associate a weight to each particle: $\omega_i \propto \frac{1}{\varepsilon_1}\left(1-\left(\frac{e_i}{\varepsilon_1}\right)^2\right)$
10: Sort e_i in descending order and store them in e^t.
11: Define the next tolerance threshold $\varepsilon_2 = e^t(\alpha_0 N)$
12: Drop particles with $\Delta(u,u^*) \geqslant \varepsilon_2$, $\omega_{j=1:\alpha_0 N} = 0$
13: Normalise the weights such that $\sum_{i=1}^{(1-\alpha_0)N}\omega_i = 1$
14: Select $\mathcal{A}^t = \beta_0 N$ particles from the remaining based on the weights
15: Define the ellipsoid by its centre of the mass and covariance matrix $\{\mu_t, \mathcal{C}_t\}$
16: Enlarge the ellipsoid by f_0 ▷For simplicity the same notation for the updated ellipsoid is kept
17: **for** $t = 2, \ldots, T$ **do**
18: **for** $j = 1, \ldots, (1-\beta_0)N$ **do**
19: **repeat**
20: Sample one particle $\boldsymbol{\theta}^*$ inside $\mathcal{E}_{t-1}$
21: Simulate u^* using the model $\mathcal{M}(\cdot)$
22: **until** $\Delta(u,u^*) < \varepsilon_t$
23: **set** $\boldsymbol{\Theta}_j = \boldsymbol{\theta}^*$, $e_j = \Delta(u,u^*)$
24: **end for**
25: Store the new particles in $\mathcal{S}_t$
26: Obtain the new particle set, $\mathcal{N}_{new} = [\mathcal{A}_{t-1}; \mathcal{S}_t]$ with their correspondent distance values e^t

(*continued on next page*)

27: Sort e^t and define $\varepsilon_{t+1} = e^t(\alpha_0 N)$
28: Associate a weight to each particle as in step (9)
29: Define the new set of selected particles $\mathcal{A}^t$ as in step (14)
30: Update the ellipsoid hyperparameters using $\mathcal{A}^t$, $\mathcal{E}_t = \{\mu_{t-1}, \mathcal{C}_{t-1}\}$ ▷The enlargement factor is kept constant
31: **end for**

3. ABC-NS for parameter estimation

3.1. Example 1: linear oscillator

A numerical example is presented here to demonstrate the computational and statistical efficiencies of the ABC-NS algorithm compared with the ABC-SMC. One considers the case of a SDOF oscillator in which the equation of motion is given by:

$$m\ddot{y} + c\dot{y} + ky = x(t) \tag{6}$$

where m is the mass (the mechanical system is assumed to have known mass, $m = 1$), c is the damping and k is the stiffness. $y, \dot{y}$ and $\ddot{y}$ are displacement, velocity and acceleration responses, respectively. The excitation $x(t)$ is a Gaussian sequence with mean zero and standard deviation 10.

The training data shown in Fig. 1 was synthetically generated by integrating numerically Eq. (6) using the fourth-fifth order Runge-Kutta method. The duration of measurements is $T = 5$ s with sampling frequency $f_0 = 100$ Hz, so that the number of data points is $n = 500$. Table 1 gives the true values used to generate the training data and their respective ranges.

Once the data has been generated, the ABC-NS algorithm illustrated in Algorithm 2 was applied to infer the model parameters. It is important to note that an appropriate final tolerance ε may be difficult to specify a priori. In this example, one considers that the convergence is met when the difference between two successive tolerance threshold values is less than 10^{-3}. Finally, the normalised mean square error (MSE) given by Eq. (7) is selected as a metric to measure the discrepancy between the observed and simulated data:

$$\Delta(z^*, z) = \frac{100}{n\sigma_{z^*}^2}\sum_{i=1}^{n}\left(z_i^* - z_i\right)^2 \tag{7}$$

where n is the size of the training data, $\sigma_{z^*}^2$ is the variance of the observed displacement; z^* and z are the observed and simulated displacements given by the model, respectively.

To be equivalent to ABC-NS, one considers for the implementation of the ABC-SMC algorithm, a percentage of alive particles $\alpha_0 = 0.6$. We set the sequence of tolerance levels obtained by ABC-NS for the ABC-SMC algorithm. A proposal PDF is assumed to be Gaussian to run the ABC-SMC algorithm (see [40], for more details).

To make a comparison between the ABC-NS and ABC-SMC algorithms, one focuses on the acceptance rate over the populations (measured by dividing the number of particles required to replenish a population by the total number of simulations) together with the quality of the posterior. From Fig. 2, one can see that the ABC-NS algorithm outperforms the ABC-SMC algorithm in terms of computational efficiency. The ABC-NS algorithm gives an acceptance rates for each simula-

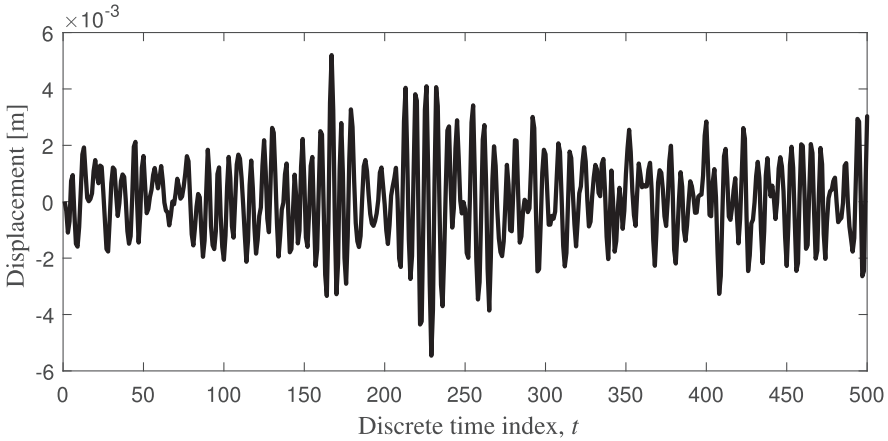

Fig. 1. Training data set.

Table 1
Parameter ranges of the linear model.

Parameter	True value	Lower bound	Upper bound
c	20	10	50
k	10^4	5000	15000

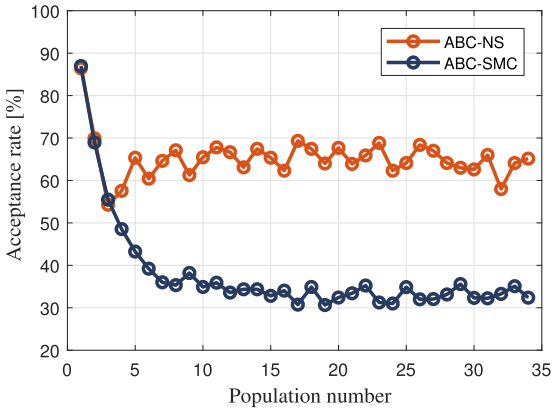

Fig. 2. Acceptance rates over populations: ABC-NS vs. ABC-SMC.

tion (population) in the range of 60–70%, while with the ABC-SMC, the acceptance rate decreases over the populations, it stabilises around 35% after few populations.

The ABC-NS and ABC-SMC results show that the mean estimates of the approximate posterior distribution is close to the true parameters. Fig. 3 displays a comparison between the posterior estimates based on 50 repeated runs of ABC-NS and ABC-SMC. One can see that the ABC-NS achieves the same or better results in terms of accuracy with less variation on the posterior estimates.

Next, one investigates the effects of the hyperparameters (f_0, α_0 and β_0) on the computational and statistical efficiencies. Fig. 4a shows the effect of f_0 on the acceptance rate, it should be noted that the pair (α_0, β_0) is set to (0.3, 0.6). One can see from Fig. 4a and b that by varying f_0 from 1 to 1.2 impacts slightly the acceptance rate over the populations. For all the values of f_0 and after few populations, the acceptance rates oscillates between 60 and 70 per cent. Overall and based on Fig. 5a and b, one can say that the effect on the posterior estimates is as well negligible.

Now, one examines the effects of the pair (α_0, β_0) on the computational and statistical efficiencies of the ABC-NS algorithm. Fig. 6 shows the acceptance rate over the populations for 5 pairs. To run simulations, the value of f_0 is set to 1.1. From the same figure, one can see that when the value of α_0 is low (i.e., the value of β_0 is high), the acceptance rate is high, however, several populations are required to ensure convergence. On the other side, a high value of α_0 is associated to a low acceptance rate which is expected as we need more particles to replenish the population. We believe that the best solution

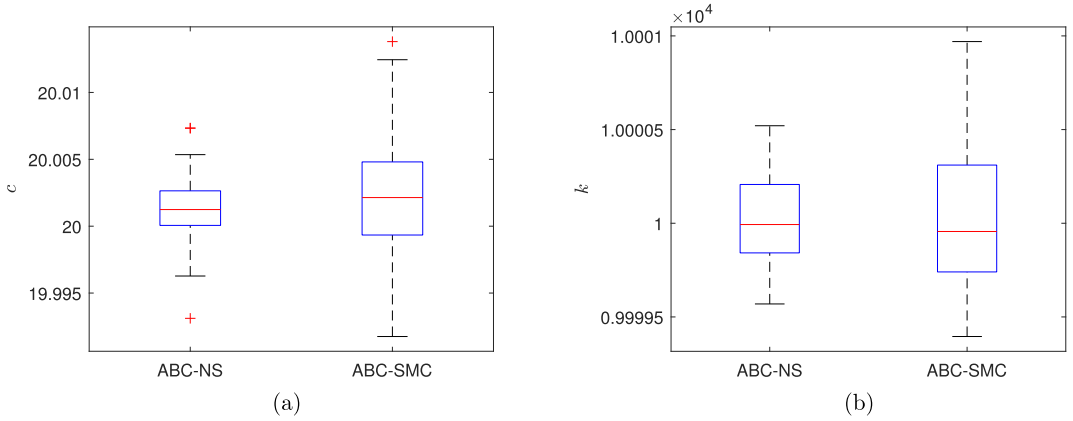

Fig. 3. Boxplots of estimated posterior mean values based on 50 repeated runs of ABC-NS and ABC-SMC.

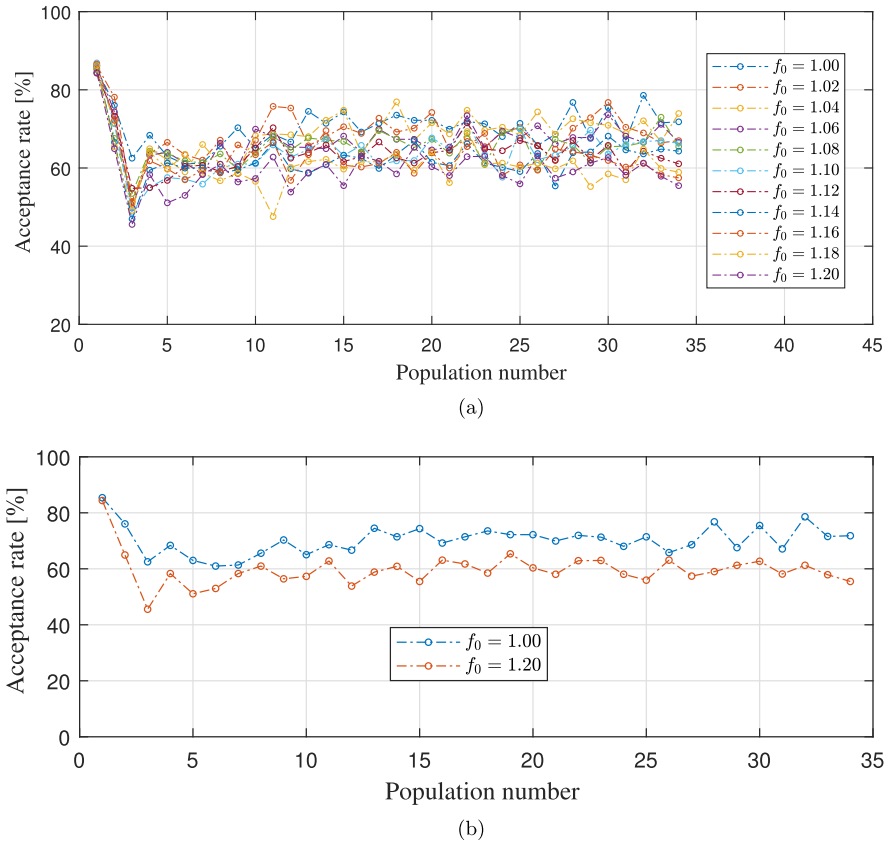

Fig. 4. (a) The effect of f_0 on the acceptance rate, (b) as for (a) only for the low and high values of f_0.

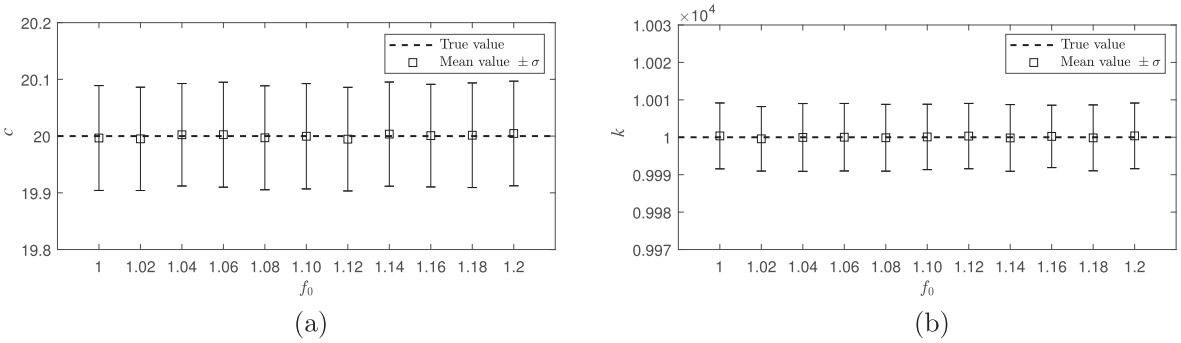

Fig. 5. The effect of f_0 on the posterior estimates (a) c and (b) k.

is a trade-off between the rejected and the remained particles at each population. In our case studies, the values of α_0 and β_0 are set to 0.3 and 0.6, respectively.

Fig. 7a and b show the evolution of the mean values ($\pm\sigma$) of c and k for the 5 different pairs. One can see that for the different pairs, the posterior estimates are quite similar. The mean squared error (MSE) (computed using the following formula: $(\frac{1}{n}\sum_{i=1}^{n}(\theta_i - \theta_{true}))$ for the inferred parameters are shown on the same figures. One can see that the evolution of the MSE values shows the same tendency for both parameters. The MSE is high when the value of α_0 is low which means that from population to population a few new particles are injected. The MSE is at lowest value when α_0 is high which means a more "good" particles have been added to the population. However, as one can see from the previous figure, the acceptance rate is low when the value of α_0 is high.

In this example, a brief discussion on the effects of each hyperparameter on the statistical and computational efficiencies is given. One can say that a good choice of the hyperparameters is the one who guarantees the best trade-off between the

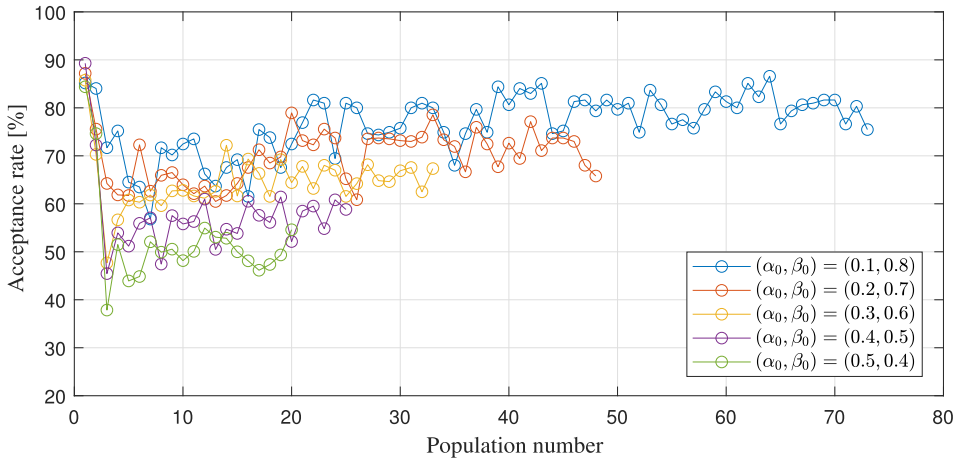

Fig. 6. Acceptance rates over populations: the effect of the pair (α_0, β_0).

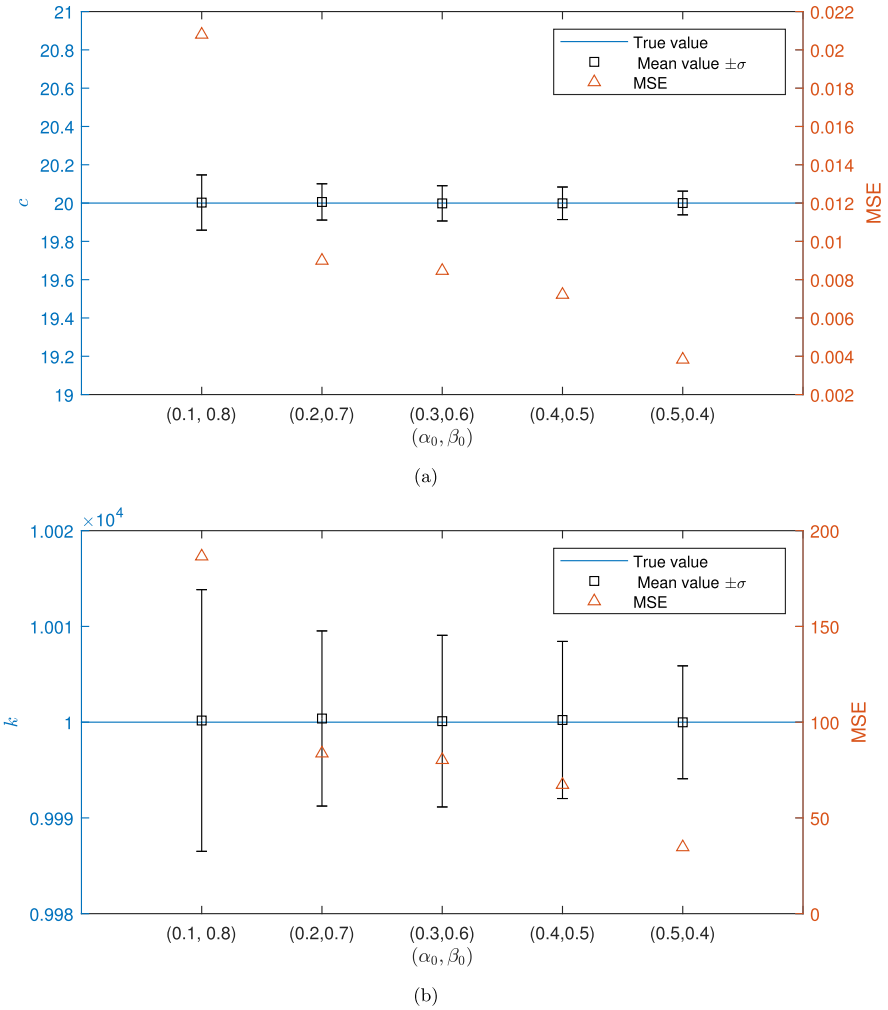

Fig. 7. The effect of the pair (α_0, β_0) on the posterior estimates, (a) c, (b) k.

computational and the statistical efficiencies. Of course, a finer study is required to provide a general guideline to the user to define those hyperparameters.

3.2. Example 2: parameter estimation of the g-and-k distribution

To demonstrate the efficiency of the ABC-NS for parameter estimation, the case study of the *g*-and-*k* distribution is presented. Such a distribution has also been used for testing several ABC approaches as in [41–43]. This distribution can model many types of behaviour through just a small number of parameters, and offers an ideal alternative to complex convolutions of "regular" distributions. The *g*-and-*k* distribution is defined by its cumulative distribution function and no explicit likelihood function is available. The *g*-and-*k* distribution, has four parameters describing location, scale, skewness and kurtosis and is thus able to model many asymmetric distributions. The quantile function of the *g*-and-*k* distribution (inverse distribution function) is given by:

$$F^{-1}(X;A,B,c,g,k) = A + B\left[1 + c\frac{1-\exp(-g\cdot r(X))}{1+\exp(-g\cdot r(X))}\right](1+r^2(X))^k r(X) \tag{8}$$

where $r(X)$ is the $r(X)$th standard normal quantile, A and B are location and scale parameters and g and k are related to skewness and kurtosis.

$\theta = (A,B,g,k)$ is the vector of distribution parameters; given that c is kept fixed to $c = 0.8$ following Rayner and MacGillivray [44] (the parameter c measures the overall asymmetry). It is noted that the normal distribution is a special case of the *g*-and-*k* distribution, with $g = 0$ and $k = 0$. Parameter restrictions are $B > 0$ and $k > -0.5$. An evaluation of Eq. (8) returns a draw (X-th quantile) from the *g*-and-*k* distribution.

We follow the simulation set up for data $y_1,\ldots,y_n$ of size $n = 10^4$ generated as in [41] with $\theta = (3,1,2,0.5)$. Fig. 8a and b show the histogram and the empirical distribution function of the data, revealing a peaked distribution with a long right-hand tail. Actually, it is extremely simple to obtain accurate inference for all parameters by reducing the dimensionality of the problem as in [43] using a smaller set of summaries. It is assumed here that $S(y) = (P_{20},P_{40},P_{60},P_{80},\text{skew}(y))$ as in [43], that is the 20–40-60-80th percentiles of the data and the sample skewness. A uniform prior $U(0,10)$ on each parameter is considered. The objective now is to estimate parameters with summaries $S(y) = (y_{(1)},\ldots,y_{(n)})$ (the sequence of ordered data) using the metric given by Eq. (9) to measure the degree of similarity:

$$J(y,z) = \left(\sum_{i=1}^{n}[S_i(z) - S_i(y)]^2\right)^{(1/2)} \tag{9}$$

with S_i the i-th element of S and $z = (z_1,\ldots,z_n)$ a vector of samples from the *g*-and-*k* distribution.

3.3. Results

To retrieve the unknown parameters, one uses the ABC-NS algorithm as illustrated in Algorithm 2. The hyperparameters used to run the ABC-NS are defined as follows: $\alpha_0 = 0.3, \beta_0 = 0.6, f_0 = 1.1, N = 1000$ and the convergence criterion is set when the difference between two consecutive tolerance threshold values is less than 10^{-5}. The tolerance level ε has been chosen in a recursive way: first a very large value (equal to 100) has been selected and then the next value is defined based on the distances obtained in the previous population as illustrated in Algorithm 2. Fig. 9 shows the marginal posterior distributions at the last population when the tolerance value is equal to $\varepsilon = 0.0074$.

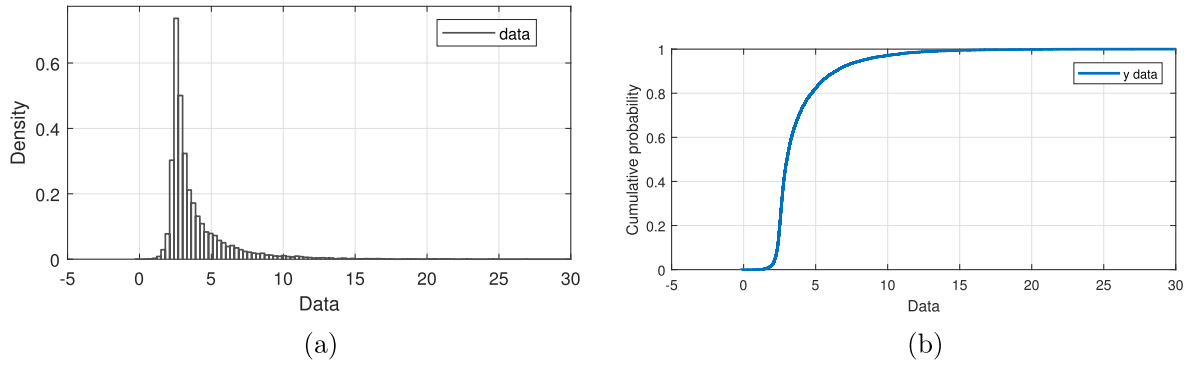

Fig. 8. (a) Histogram of simulated data set from *g*-and-*k* distribution, (b) empirical cumulative distribution function.

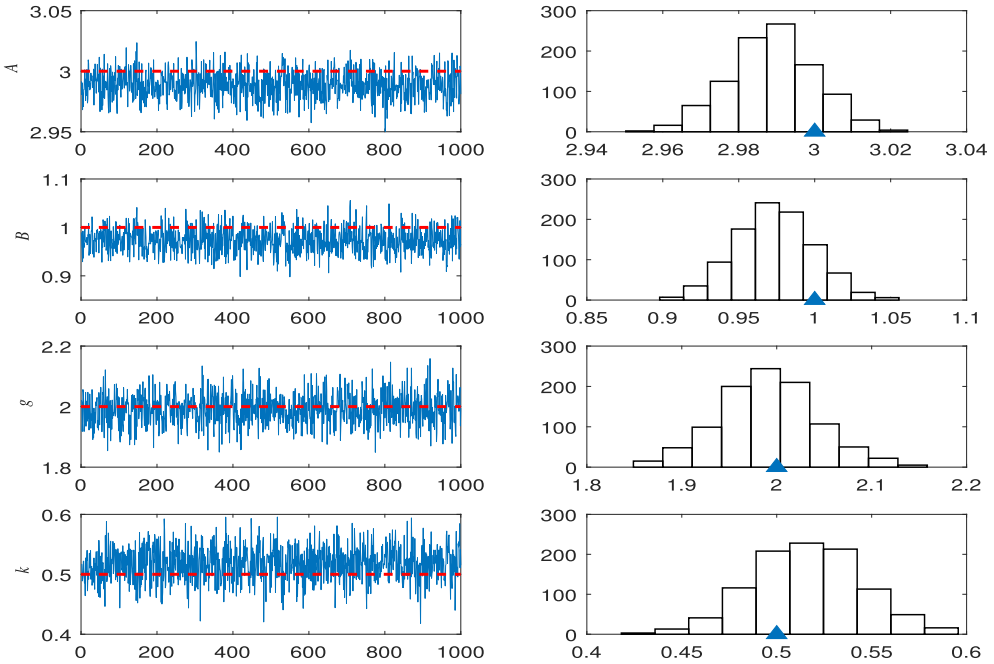

Fig. 9. (left) *g*-and-*k* distribution: trace plots for the ABC-NS run at the last population, (right) marginal posterior distributions of the *g*-and-*k* distribution parameters, (the blue triangles show the true values).

From Table 5, one can see that ABC-NS provides a good inference of the distribution parameters. The 95% credible intervals for the posterior distribution of all parameters contain the true values. This example shows that by selecting a suitable set of summary statistics, one can provide a good estimates of the model parameters. From the obtained results, one can see that *g* is identifiable with high uncertainty compared with the other parameters of the distribution, mainly for the first two tolerance values. It should be noted that it is possible to further reduce the uncertainty in the distribution parameters by decreasing the final tolerance threshold, of course by introducing computational cost.

The second example illustrates the statistical efficiency of the ABC-NS in dealing with parameter estimation from data. One can see how, by selecting an appropriate summary statistics, one can efficiently make Bayesian inference circumventing the issue of intractable likelihood functions. In the rest of this paper, the efficiency of the ABC-NS to deal with *model selection* is investigated through two examples using simulated and real data. (Table 2).

Table 2
Posterior estimates for the *g*-and-*k* distribution parameters at different target tolerance values.

Tolerance value	Parameter	Posterior estimates		
		Mean	Std dev	[2.5%, 97.5%] percentiles
$\varepsilon = 0.1345$	A	2.9615	0.0573	[2.9313, 3.0587]
	B	0.9839	0.0458	[0.9499, 1.0712]
	g	2.5434	1.5117	[1.9146, 9.3087]
	k	0.4964	0.0402	[0.4689, 0.5740]
$\varepsilon = 0.0444$	A	2.9842	0.0201	[2.9702, 3.0230]
	B	0.9873	0.0332	[0.9653, 1.0528]
	g	2.0383	0.1317	[1.9406, 2.2894]
	k	0.4970	0.0374	[0.4715, 0.5688]
$\varepsilon = 0.0169$	A	2.9869	0.0129	[2.9777, 3.0113]
	B	0.9794	0.0283	[0.9614, 1.0335]
	g	2.0064	0.0648	[1.9604, 2.1281]
	k	0.5085	0.0330	[0.04858, 0.5754]
$\varepsilon = 0.0074$	A	2.9888	0.0115	[2.9815, 3.0114]
	B	0.9738	0.0257	[0.9568, 1.0237]
	g	1.9906	0.0515	[1.9569, 2.0980]
	k	0.5163	0.0290	[0.04965, 0.5716]

4. ABC-NS for model selection using Duffing-type oscillators

4.1. Example 1: cubic and cubic-quintic models

The performance of the ABC-NS algorithm is now investigated for model selection by considering two candidate models: the cubic and cubic-quintic Duffing oscillators denoted by $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. The equation of motion associated to each model is given by:

$$\mathcal{M}_1 : \ddot{z} + c\dot{z} + kz + k_3 z^3 = f(t) \tag{10}$$

$$\mathcal{M}_2 : \ddot{z} + c\dot{z} + kz + k_3 z^3 + k_5 z^5 = f(t) \tag{11}$$

where c is the damping, k is the linear stiffness, k_3 and k_5 are the non-linear stiffness coefficients. $z, \dot{z}$ and $\ddot{z}$ are displacement, velocity and acceleration responses, respectively. The excitation $f(t)$ is a Gaussian sequence with mean zero and standard deviation of 10.

To make Bayesian inference, a noisy training data set was generated from the cubic-quintic model and shown in Fig. 10 (the first 500 data points). It has been corrupted with Gaussian noise of standard deviation 1% RMS. The fourth and fifth-order Runge-Kutta algorithm is chosen to integrate the equations of motion. To evaluate the model predictability, a set of testing data has been generated shown in Fig. 10 (the second 500 points). Table 3 summarises the prior lower and upper bounds associated to each unknown parameter of the competing models.

For the ABC-NS implementation, the same scheme shown in Algorithm 2 is followed by considering the candidate models as additional parameters. One sets the prior probabilities of each model to be equal, i.e., $p(\mathcal{M}_1) = p(\mathcal{M}_2) = \frac{1}{2}$. In the ABC-NS for model selection, one treats the pair $(\mathcal{M}_k, \theta^{(k)})$ with $\mathcal{M}_k$ as a candidate model and $\theta^{(k)}$ its vector of unknown parameters. For a given $(\mathcal{M}_k, \theta^{(k)})$, the pair is accepted or rejected based on a discrepancy value. At the end of the algorithm, the model probability for $\mathcal{M}_k$ is approximated using Eq. (12).

$$p(\mathcal{M}_k|u^*) \approx \frac{\text{Accepted particles for } \mathcal{M}_k}{\text{Total number of particles } N} \tag{12}$$

The convergence criterion used here is when the difference between two successive tolerance values is less than 10^{-7} ($\varepsilon(j) - \varepsilon(j+1) < 10^{-7}$, $j$isthepopulationnumber). For the rest, the same hyperparameters defined earlier have been used. It should be noted that the number of the dropped and remaining particles are rounded to the nearest integer in each step of the algorithm such that the sum of the dropped and new particles is equal to N. Finally, the normalised mean square error (MSE) given by Eq. (7) and used in the first example is selected as a metric to measure the discrepancy between the observed and simulated data.

4.2. Results and discussion

Fig. 11 shows the model posterior probabilities over a selected number of populations. One can see how the ABC-NS algorithm oscillates between the competing models and finishes by converging to the correct model when the tolerance values become very small. From the same figure, one can see that the algorithm clearly tries first to favour the cubic model, this can be seen from populations 1 to 37. Then, when the cubic model is no longer able to match the data very well, the ABC-NS algorithm jumps to the complex model to accommodate the nonlinearity coming from the quintic term. The probability of selecting this "true" model approaches when the tolerance value is very small and close to zero. This demonstrates that the parsimony principle [45] is well embedded in the ABC-NS algorithm by favouring first simpler models. As mentioned earlier, in the classical methods based on the evaluation of an information criterion, overly-complex models are penalised through an *ad hoc* penalty term while the ABC algorithm naturally favours simpler models as shown here, which is a major advantage circumventing the issue of which information criterion is more suitable to compare models.

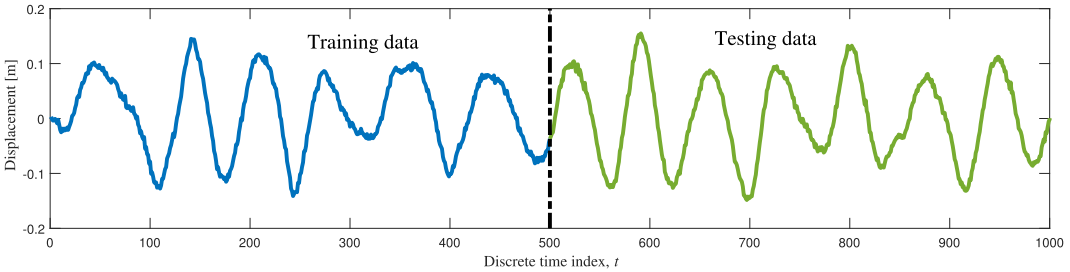

Fig. 10. Training and testing data sets.

Table 3
Parameter ranges of the competing models.

Parameter	True value	Lower bound	Upper bound
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

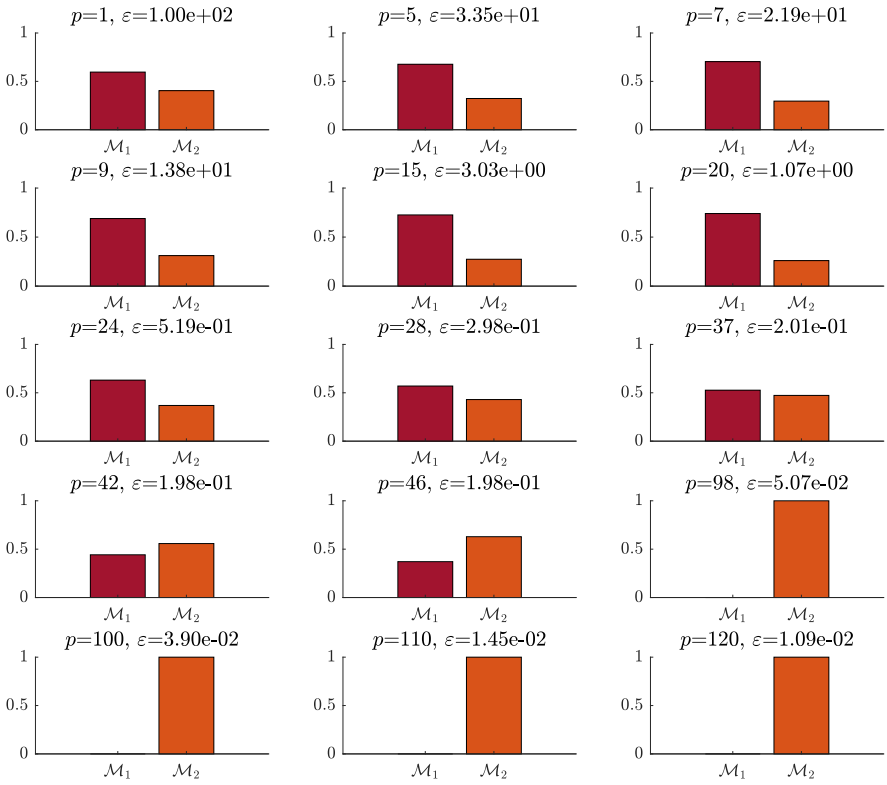

Fig. 11. Model posterior probabilities considering the cubic and cubic-quintic models.

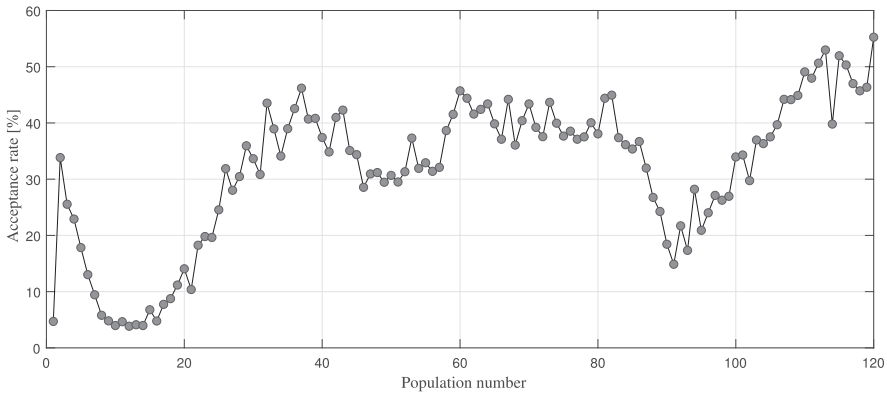

Fig. 12. Acceptance rate over the populations.

Fig. 12 shows the acceptance rate over the populations, one can see how the ABC-NS algorithm maintains a relatively high acceptance rate over the populations. At early populations, the acceptance rate decreases because the input space to be explored is large, then steadily rises as the volume of the search space shrinks down. From population 30 to population

81, it stabilises around an average value higher than 35 per cent and then decreases as a finer representation of the data is required at that stage until the elimination of the cubic model (population 91). Then, from population 91, the acceptance rate rises again as the search domain for the selected model is well defined.

Fig. 13 shows the histograms of the unknown parameters associated to the cubic-quintic model obtained at the last population. One can see how the histograms are well peaked around the true values. Table 4 summarises the statistics of the posterior estimates associated with the selected model. The estimated model parameters are then used to make future predictions and evaluate the model predictability. From Fig. 14, it can be seen that the training and testing data sets are well predicted. The 99% confidence interval is found pointwise by generating randomly a large number of samples, simulating the model responses and then the 99% confidence interval is found pointwise. The normalised MSE estimated on the first 500 data points is equal to 0.2580 and is 0.4698 on the testing data. Based on [2], the discrepancy measure formulated in Eq. (7) has shown that a value less than 5 generally indicates a good fit while a value less than 1 means an excellent fit. It should be noted that the obtained posterior estimates could be easily refined and therefore reduce the uncertainty on the parameters by further decreasing the final tolerance threshold value. The main advantage of the ABC-NS algorithm in comparison with its predecessors is that this can be done at low computational cost as the region from where one can sample a "good" candidates is well delimited by the ellipsoids.

Finally, in order to check the repeatability of the model posterior probabilities, the ABC-NS algorithm is run 20 times. From Fig. 15, one can see how the ABC-NS produces repeatable results with small variations. Clearly the algorithm tries first to favour the simpler model (here the cubic model) (see the model posterior probabilities at populations 1, 11, 31, 51, 76 and 91) and when a higher predictive performance is required then the algorithm switches to the more complex model to justify the increase in complexity. In short, a simpler (but equally accurate) explanation for data always has the greater evidence.

5. Characterisation of the dynamics of a wire rope isolators using ABC-NS

5.1. Experimental set-up

The last system consists of characterising the dynamics of a wire rope isolator (WRI) used for vibration isolation. WRIs have found a vast number of application in medical equipment, mechanical machinery, and military hardware due to their superior performance for the isolation of impact and vibration. However, the dynamical properties of mechanical isolators are typically non-linear and these characteristics are seldom well defined, which may cause problems for design calculation and computer simulations. The system considered in this paper has been proposed within the framework of the European COST Action F3 working group in "Identification of non-linear systems" [46]. The aim of this benchmark was to identify the dynamic properties of resilient mounts used for vibration isolation in industrial applications using different methods.

Fig. 16a shows the experimental set-up of the WRI mounted between a load mass m_2 and a base mass m_{1b} while Fig. 16b is a schematic illustration. The excitation produced by an electro-dynamic shaker corresponds to a white noise sequence,

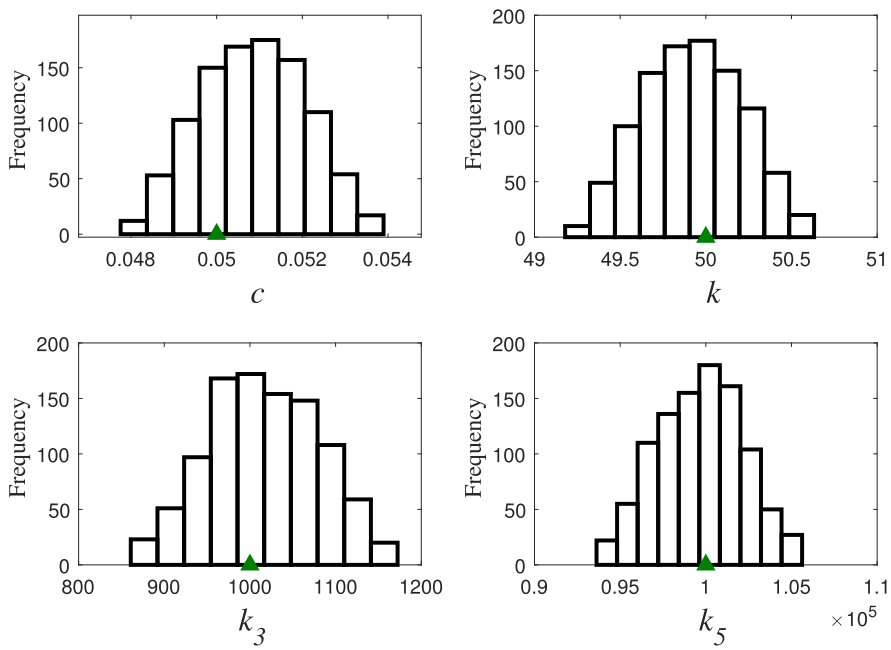

Fig. 13. Histograms for the cubic-quintic model parameters.

Table 4
Posterior estimates for the cubic-quintic model parameters.

Parameter	True value	Posterior estimates		
		Mean, μ	Std. Dev, σ	[5th, 95th] percentiles
c	0.05	0.05086	1.2167×10^{-3}	$[1.87 \times 10^{-2}, 5.2827 \times 10^{-2}]$
k	50	49.9211	2.8995×10^{-1}	[49.4382, 50.3844]
k_3	10^3	1016.5110	63.8396	[914.045, 1123.6337]
k_5	10^5	9.9670×10^4	2.5015×10^3	$[9.5183 \times 10^4, 1.036 \times 10^5]$

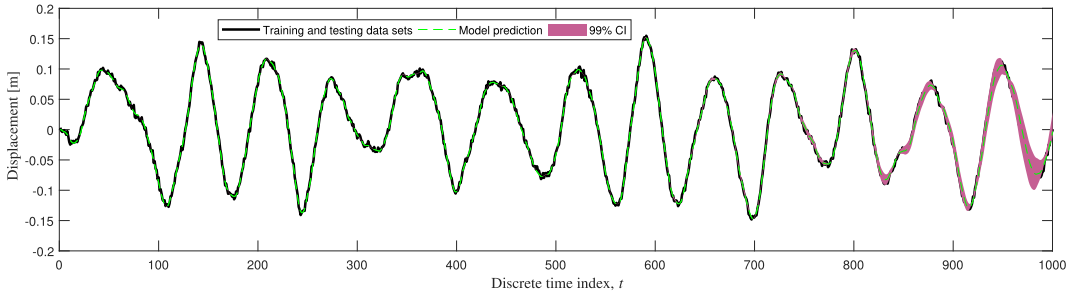

Fig. 14. Model prediction using the cubic-quintic model on training and testing data sets.

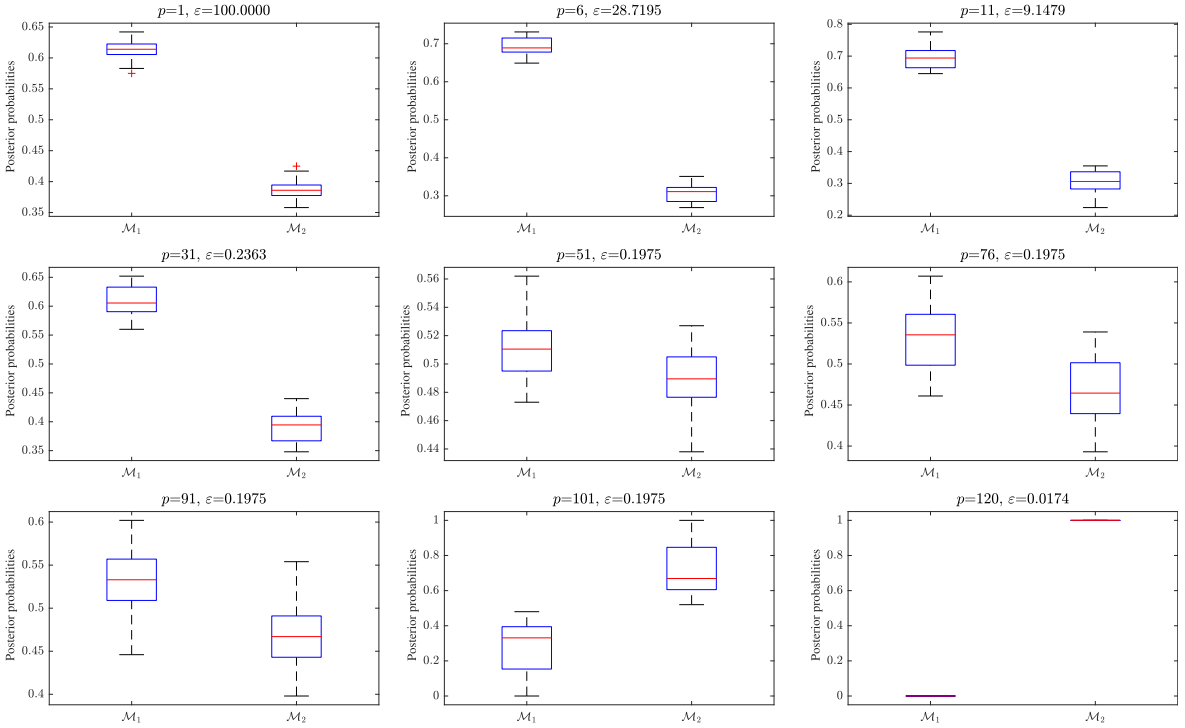

Fig. 15. Boxplots of the model posterior probabilities over some selected populations.

low-pass filtered at 400 Hz. The motion and forces experienced by the isolators are measured; in particular, the acceleration responses $\ddot{x}_2$ and $\ddot{x}_{1b}$ of the load mass and bottom plate, the applied f and the relative displacement x_{12} between the top and bottom plates. Five excitations ranging from 0.5 up to 8 V have been considered. For more details concerning the experimental set-up and the methods presented for the identification of the system, the reader is referred to the following references [49,50,47,48]. Table 5 illustrates the characteristics of the testing system and the WRI properties.

Wire rope isolators have different response characteristics depending on the selected properties mentioned in Table 5. To determine their dynamic characteristics, a series of dynamic tests were conducted by imposing a random excitation at dif-

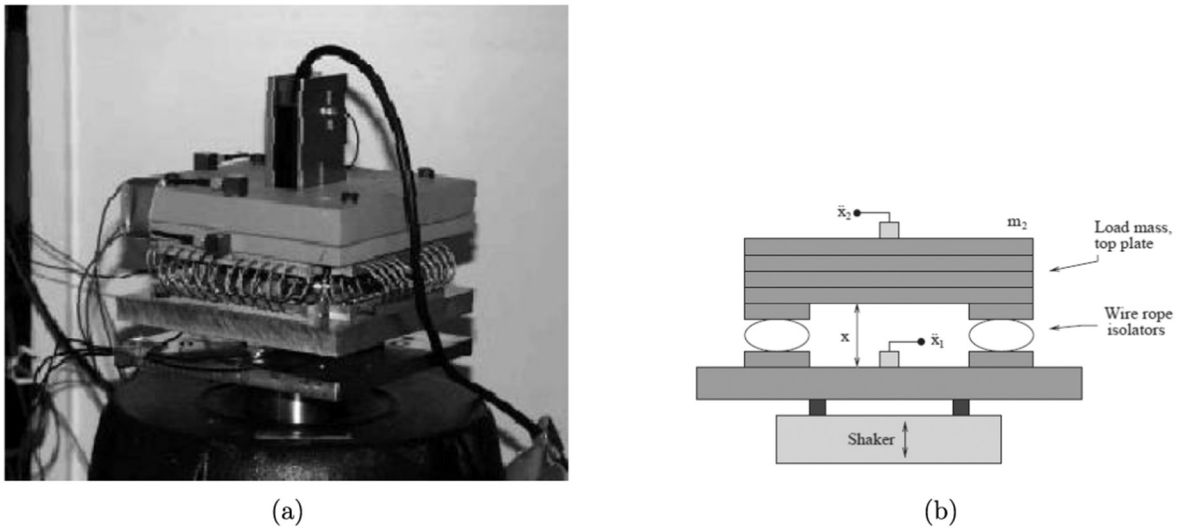

Fig. 16. (a) Experimental set-up: schematic configuration of the experiment, (b) illustration of the dynamical system under consideration [50].

Table 5
System characteristics and geometrical properties of helical wire rope isolators.

System characteristics	
Load mass, m_2	2.2 kg
Bottom plate, m_{1b}	1.1 kg
Shaker base	3.7 kg
Wire diameter	2 mm
Length	110 mm
Number of loops	10
Loop diameter	30 mm

ferent excitation levels. Random excitation consists of a white noise, low-pass filtered (LPF) (18 db/oct) at 400 Hz. Different levels of excitation were produced and the time signals were recorded with a sampling frequency of4096 Hz. As an example, the displacement and inertial force versus relative displacement for three levels of excitation are shown in Fig. 17 a–f. Recorded inertial force-relative displacement loops showed hysteretic behaviour for all five amplitudes (only three of them are shown for brevity). A hysteretic behaviour can be seen which means that a hysteretic model could potentially describe the data reasonably well.

5.2. Selection of the candidate models

For the competing models, one selects the linear model given by the following expression as a competing model:

$$m\ddot{y} + c\dot{y} + Ay = x(t) \tag{13}$$

where (m, c, A) are the parameters to be identified and $x(t)$ is the applied excitation.

It should be noted that the linear model is selected on purpose to analyse the behaviour of the algorithm and to investigate if this model could be used to describe the dynamics of the WRI, mainly at low excitation levels. To describe the hysteretic behaviour of the WRI, a well developed mathematical model of hysteresis would be useful to make predictions and avoids the time-consuming experimental work. In this study, the Bouc-Wen model is used to model the hysteretic behaviour. Many researchers have used the Bouc-Wen model to perform mathematical modelling of the hysteresis system in several areas including hysteretic isolators [51,52]. The Bouc-Wen model of hysteresis is given by:

$$m\ddot{y} + g(y, \dot{y}) + z(y, \dot{y}) = x(t) \tag{14}$$

where m is the mass of the system, $g(y, \dot{y}) = c\dot{y} + ky$ is the polynomial part of the restoring force, $z(y, \dot{y})$ is the hysteretic part and $x(t)$ is the excitation force.

The hysteretic component is defined by Wen [53] via an additional equation of motion:

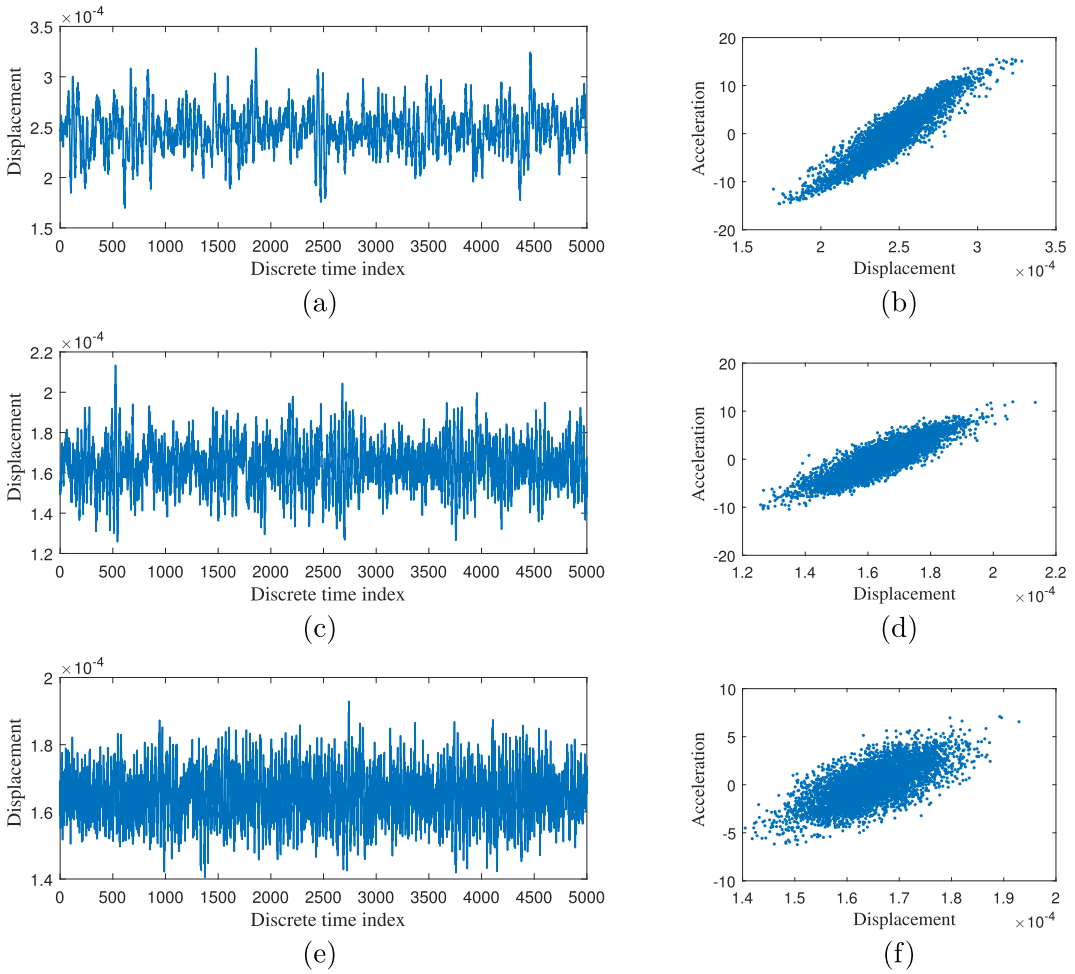

Fig. 17. Displacement and inertial force versus relative displacement under excitation levels of (a, b) 8 V, (c, d) 4 V and (e, f) 2 V.

$$\dot{z} = \begin{cases} -\alpha|\dot{y}|z^n - \beta\dot{y}|z^n| + A\dot{y}, & \text{for } n \text{ odd} \\ -\alpha|\dot{y}|z^{n-1}|z| - \beta\dot{y}|z^n| + A\dot{y}, & \text{for } n \text{ even} \end{cases} \tag{15}$$

where (c, α, β, A) are the parameters to be identified, n is a discrete parameter.

The parameters (A, α, β) govern the shape and smoothness of the hysteresis loop. The equations offer a simplification from the point of view of parameter estimation in that the stiffness term in Eq. (14) can be combined with the term A in the state equation for z. The coupled Eqs. (14) and (15) were integrated forward in time using fourth-order Runge-Kutta integration.

In total, five competing models will be considered to perform model selection using ABC-NS: the linear model given by Eq. (13) and hysteretic models (Eqs. (14) and (15)) by varying n from 1 to 4. The priors assigned to the model parameters are given in Table 6. All other settings for the ABC-NS algorithm were specified as before.

In the first part of this example, one aims to select the most likely model among the competing ones for each excitation amplitude. The data set contains 1000 samples representing a short recording period of the acceleration of the top plate. The

Table 6
Priors on model parameters.

Parameter	Lower bound	Upper bound
c	-10^3	10^3
k	-10^5	10^6
α	-10^4	10^5
β	-10^4	10^5
A	0	10^6

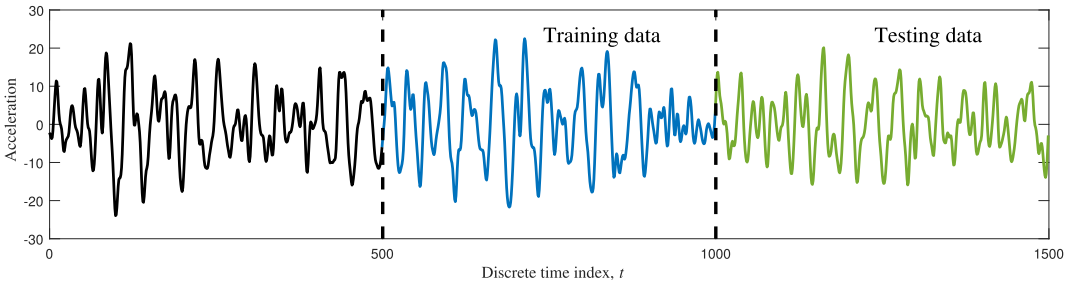

Fig. 18. Training and testing data sets using an excitation of amplitude 8 V.

data set is split into training data (from 501 to 1000) and testing data (from 1001 to 1500) as shown in Fig. 18. It should be noted that the transient part (from 0 to 500) has been ignored in computing the MSE, to reduce the effect of initial conditions.

5.3. Results and analyses

Following the same scheme as before, the ABC-NS for model selection is implemented by assuming the same prior probability to the competing models. Fig. 19 shows the model posterior probabilities over some selected populations. One can see that the most likely model is $\mathcal{M}_2$. As before, one can see that the parsimony principle is deeply embedded in the ABC-NS algorithm. It tries firstly to select the simpler models and then when those models are not able to describe the dynamics due the complexity of the data and the presence of nonlinearity, then the algorithm jumps to a more sophisticated model. Table 7 summarises the statistics of model parameters for the selected excitation levels from where one can see that the

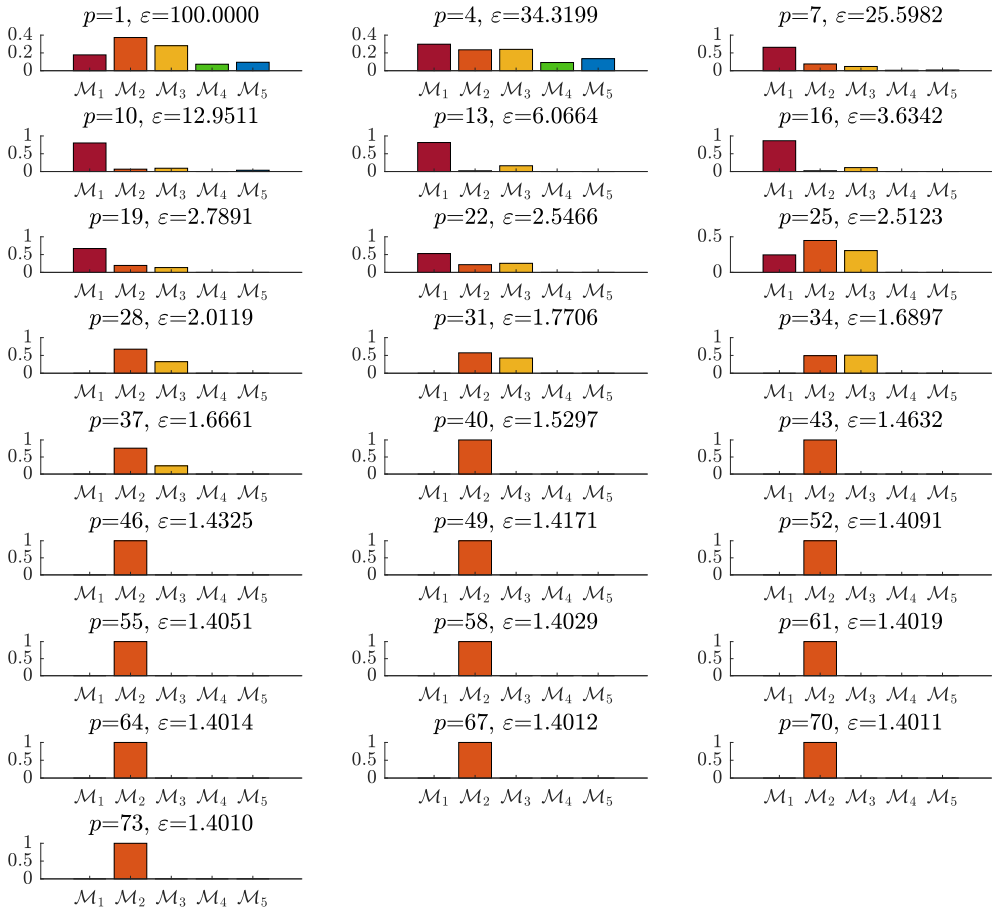

Fig. 19. Model posterior probabilities under excitation amplitude 8 V.

Table 7
Posterior estimates using an excitation amplitude of 8 V.

Excitation level	Moments	Model parameters			
		c	α	β	A
8 V	μ	78.7315	3.9135×10^3	1.5644×10^4	3.3328×10^5
	σ	0.2414	29.2111	97.9579	4.6459×10^2

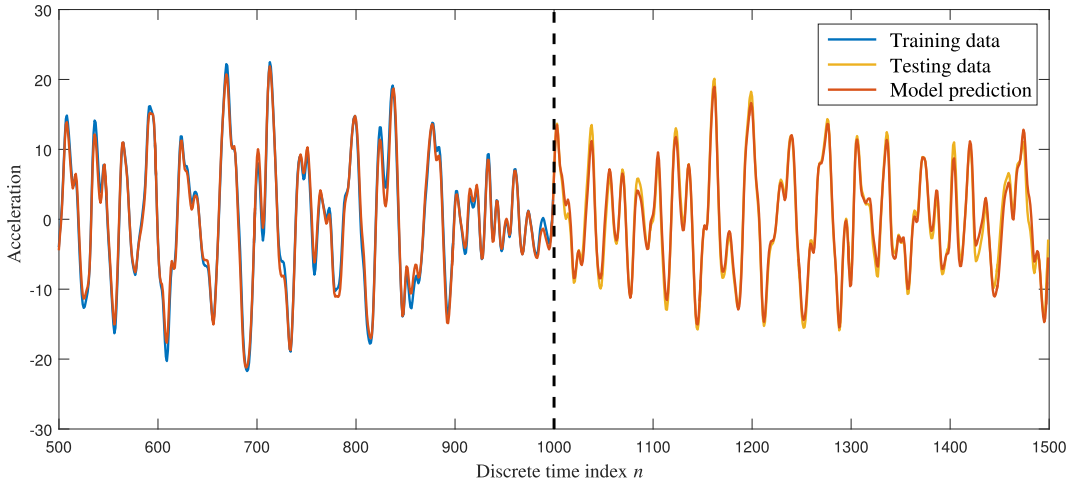

Fig. 20. Model prediction under excitation amplitude 8 V.

parameters have been estimated with a reasonable amount of uncertainties. Fig. 20 shows the model prediction on the training and testing data sets from where one can see a good agreement.

Fig. 21 shows how the ABC-NS maintains a high acceptance rate over the populations required for convergence. This reduces considerably the computational time to converge to the most likely model.

The same procedure is now applied for the rest of the data sets for all the excitation levels to determine the most likely models. The Bayesian inference using ABC-NS with the same hyperparameters as earlier is performed. Table 8 shows the most likely model for each data set and the MSE estimated on the training and testing data sets. Table 9 shows the statistics of the model parameters from which, one can see as before, a reduced amount of uncertainty on the parameters. Model prediction for excitation 4 V is shown in Fig. 22, where one can see a good correlation between real and simulated data on training and testing data sets. A good agreement is shown as well for the other excitation levels not shown here for brevity.

The obtained results show that the Bouc-Wen model describes reasonably well the dynamics of the WRI. One point should be noted that the selection of the appropriate model varies with the excitation level which is not desirable mainly

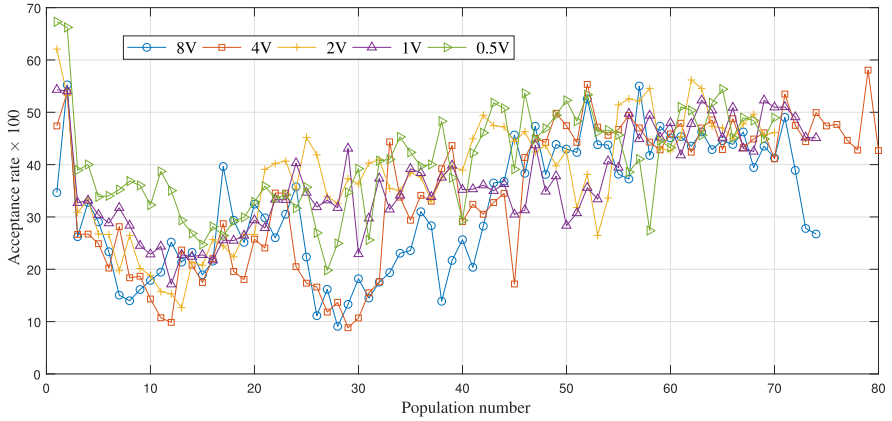

Fig. 21. Acceptance rate through the populations.

Table 8
Normalised MSE evaluated on training and testing data using mean posterior estimates.

Excitation amplitude (V)	Selected model	Training data	Testing data
8	$\mathcal{M}_2$	1.4010	2.4538
4	$\mathcal{M}_2$	0.7787	0.6338
2	$\mathcal{M}_3$	0.1783	0.2782
1	$\mathcal{M}_5$	0.1848	0.3368
0.5	$\mathcal{M}_4$	0.5550	0.9999

Table 9
Posterior estimates using the other excitation amplitudes.

Excitation level (V)	Moments	Model parameters			
		c	α	β	A
4	μ	70.9807	1.7234×10^3	3.3253×10^4	4.5797×10^5
	σ	0.1400	12.9543	58.2083	2.4260×10^2
2	μ	41.2794	1.1003×10^3	2.9893×10^4	4.7203×10^5
	σ	0.2280	90.9088	1.7992×10^2	3.1932×10^2
1	μ	29.6899	4.1047×10^3	2.6767×10^4	4.7829×10^5
	σ	0.1537	88.9378	1.8534×10^2	2.0254×10^2
0.5	μ	−3.0524	3.4739×10^4	-1.7453×10^4	4.5444×10^5
	σ	0.1422	1.8659×10^2	3.0277×10^2	1.5292×10^2

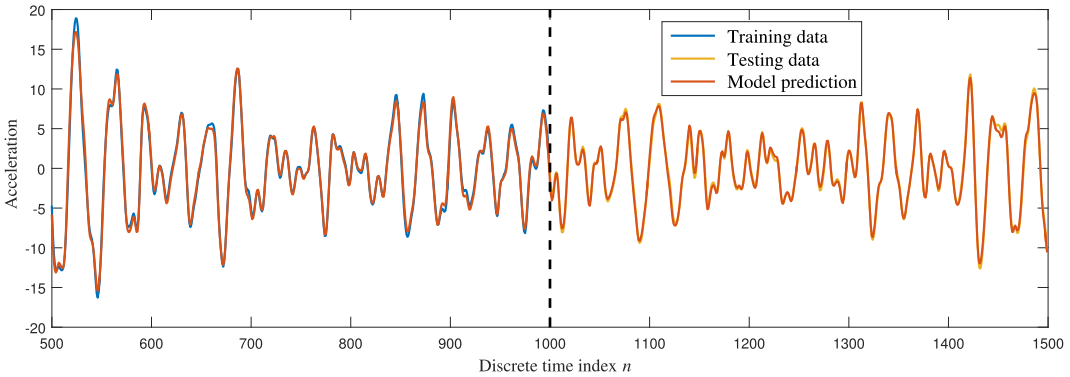

Fig. 22. Model prediction under an excitation amplitude of 4 V.

when the practitioner needs one model to describe the dynamics of the WRI independently of the excitation level. This point is addressed in the next subsection where the objective is to select one model which performs better considering all the testing data sets and all the excitation levels simultaneously. To answer this question, one uses a *confusion matrix* introduced and widely used in machine learning.

Before closing this section, one important point should be highlighted at low excitation levels (0.5 V, 1 V and 2 V). It has been noticed that the linear model is favoured at the beginning of the algorithm and has been eliminated at an advanced stage, as can be seen from Fig. 23 (at population 26 with $\varepsilon = 0.812$). This leads one to think that the system behaviour might be predominantly linear, though the algorithm converges to one of the hysteretic models when the performance requirement in terms of prediction is high. To confirm this result, one uses the linear model to check its ability to make predictions. Figs. 24 and 25 show that using the linear model at two different tolerance threshold values, an acceptable agreement is shown between the predicted and simulated data (in Fig. 25, the 99% CI is not shown as it is indistinguishable from the model prediction).

5.4. Confusion matrix

Although model selection can be straightforwadly performed for each excitation amplitude alone. The practitioner may prefer one single model which can be used to describe the dynamics of the isolator regardless of the excitation level. In such

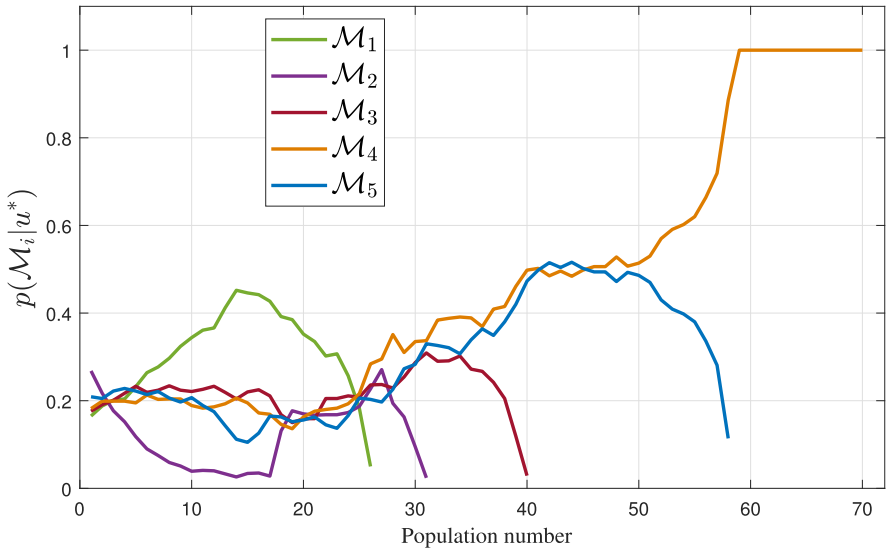

Fig. 23. Evolution of the model posterior probabilities over the populations using an excitation amplitude of 0.5 V.

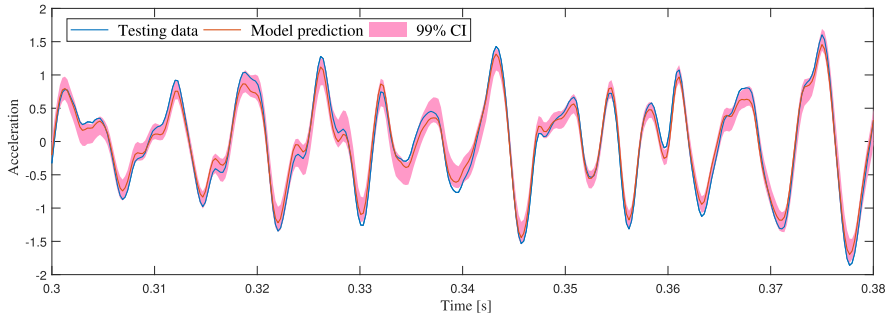

Fig. 24. Model prediction using the linear model under an excitation amplitude of 0.5 V ($\varepsilon = 5.44$).

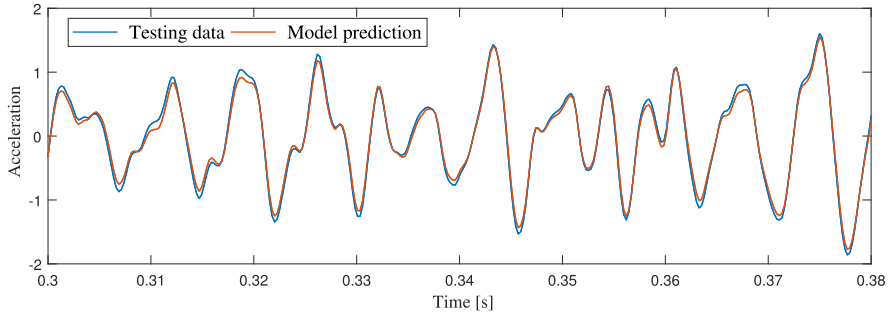

Fig. 25. Model prediction using linear model under excitation amplitude of 0.5 V ($\varepsilon = 0.812$).

cases, a problem may arise of deciding which of the competing models can provide acceptable predictions considering all the testing data sets. To answer this challenging question, one introduces the *confusion matrix*, a concept widely adopted in machine learning as mentioned earlier. The objective behind its use is to train the model considering one amplitude and then quantify the performance of the model prediction based on the same metric (normalised MSE) using the rest of the testing data sets as illustrated in Fig. 26. The confusion matrix is given below for a better understanding of its use in this context. This allows one to give a clear idea of how each model performs considering all the testing data sets. Then, the selection of the best model is done straightforwadly based on a comparison of the MSE values considering all the testing data sets.

		Testing data		
		A_1	$\cdots$ A_j $\cdots$	A_n
Training data	A_1	MSE_{11}	MSE_{1j}	MSE_{1n}
	A_i	MSE_{i1}	MSE_{ij}	MSE_{in}
	A_n	MSE_{n1}	MSE_{nj}	MSE_{nn}

Fig. 26. The confusion matrix concept.

Fig. 27 gives the confusion matrices associated to the competing models. One can clearly see that $\mathcal{M}_3^{n=2}$ is the best model considering all the testing data sets based on the MSE values. This means that this model generalises better under different excitation levels and should be selected if the practitioner prefers one model rather than a model for each excitation amplitude. The confusion matrix associated to $\mathcal{M}_5$ is not shown here as it shows some numerical instability and therefore it has been eliminated from the competition.

Based on these results, one may conclude that the Bouc-Wen model is a good way for describing the hysteretic behaviour of the WRI. It has been shown that this model is efficient and has a good performance in terms of prediction under different excitation amplitudes. Overall, the simulated data were in good agreement with experimental results.

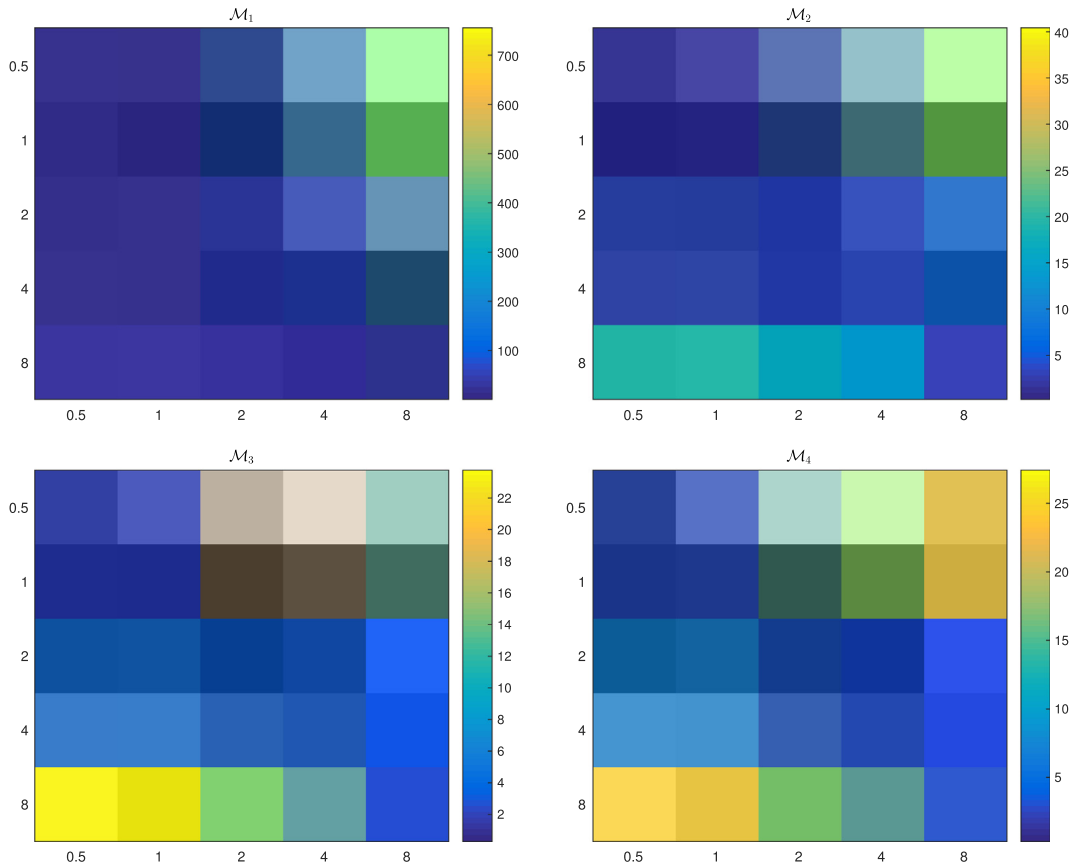

Fig. 27. Confusion matrices considering all the testing data sets (colorbars display the normalised MSE values). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

6. Conclusions

A new approximate Bayesian computation algorithm based on an ellipsoidal nested sampling method named ABC-NS has been proposed in this paper for parameter estimation and model selection. It has been shown through four examples using simulated and real data how the ABC-NS outperforms the popular ABC-SMC and overcomes the low efficiency observed after few iterations by employing a more sophisticated technique to generate new samples. ABC-NS maintains a high acceptance rate over the populations, which speeds up considerably the algorithm without compromising the precision of the posterior estimates. As a result, significant savings in computational effort can be achieved, which is desirable particularly to enable larger models to be analysed, the use of more computationally intensive forward simulation models and the inclusion of additional uncertain parameters. Moreover, as it has been shown, the parsimony principle is naturally embedded in the ABC algorithm. The ABC tells one which models are supported by the data in a straightforward way without any additional cost. In conclusion, likelihood-free or approximate Bayesian computation algorithms represent a simple and efficient way to handle highly complicated problems. It is extremely useful mainly when the likelihood function is intractable or cannot be approached in a closed form offering the possibility to make Bayesian inference by using different kinds of features and metrics representative of the data.

Acknowledgement

The support of the UK Engineering and Physical Sciences Research Council (EPSRC) through grant reference No. EP/K003836/1 is greatly acknowledged.

References

[1] G.E. Box, N.R. Draper, Empirical Model-Building and Response Surfaces, Wiley, New-Work, 1987.
[2] K. Worden, J.J. Hensman, Parameter estimation and model selection for a class of hysteretic systems using Bayesian inference, Mech. Syst. Signal Process. 32 (2012) 153–169.
[3] Ph. Bisaillon, R. Sandhu, M. Khalil, C. Pettit, D. Poirel, A. Sarkar, Bayesian parameter estimation and model selection for strongly nonlinear dynamical systems, Nonlinear Dyn. 82 (2015) 1061–1080.
[4] J.L. Beck, K.V. Yuen, Model selection using response measurements: Bayesian probabilistic approach, J. Eng. Mech. 130 (2) (2004) 192–203.
[5] R. Sandhu, M. Khalil, A. Sarkara, D. Poirel, Bayesian model selection for nonlinear aeroelastic systems using wind-tunnel data, Comput. Methods Appl. Mech. Eng. 282 (2014) 161–183.
[6] J.L. Beck, S.K. Au, Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation, J. Eng. Mech.- ASCE 128 (2002) 380–391.
[7] J.Y. Ching, Y.C. Chen, Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging, J. Eng. Mech. ASCE 133 (2007) 816–832.
[8] C. Soize, Bayesian posteriors of uncertainty quantification in computational structural dynamics for low-and medium-frequency ranges, Comput. Struct. 126 (2013) 41–55.
[9] K.V. Yuen, Bayesian Methods for Structural Dynamics and Civil Engineering, John Wiley & Sons, 2010.
[10] H. Akaike, Information theory and an extension of the maximum likelihood principle, Breakthroughs in Statistics, vol. I, Springer, 1992, pp. 610–624.
[11] G. Schwarz et al, Estimating the dimension of a model, Ann. Stat. 6 (2) (1978) 461–464.
[12] D.J. Spiegelhalter, N.G. Best, B.P. Carlin, A. van der Linde, Bayesian measures of model complexity and fit, J. R. Stat. Soc. Ser. B (Methodol.) 64 (4) (2002) 583–639.
[13] P.J. Green, Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, Biometrika 82 (4) (1995) 711–732.
[14] M. Muto, L.J. Beck, Bayesian updating and model class selection for hysteretic structural models using stochastic simulation, J. Vib. Control 14 (2008) 7–34.
[15] J.Y. Ching, Y.C. Chen, Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging, J. Eng. Mech.- ASCE 133 (2007) 816–832.
[16] W. Betz, I. Papaioannou, D. Straub, Transitional Markov Chain Monte Carlo: observations and Improvements, ASCE J. Eng. Mech. 142 (5) (2016).
[17] S. Wu, P. Angelikopoulos, C. Papadimitriou, Petros Koumoutsakos, Bayesian annealed sequential importance sampling: an unbiased version of transitional Markov Chain Monte Carlo, ASCE-ASME J. Risk Uncertainty Eng. Syst. Part B 4 (1) (2018), 011008-1.
[18] J. Skilling, Nested sampling for general Bayesian computation, Bayesian Anal. 1 (4) (2006) 833–860, https://doi.org/10.1214/06-BA127.
[19] J. Skilling, Nested sampling, in: R. Fischer, R. Preuss, U. Toussaint (Eds.), Bayesian Inference and Maximum Entropy Methods in Science and Engineering, AIP Conference Proceedings, 735, 2004, pp. 395–405.
[20] F. Feroz, M.P. Hobson, Multimodal nested sampling: an efficient and robust alternative to MCMC methods for astronomical data analysis, Monthly Notices R. Astron. Soc. 384 (2007) 449–463.
[21] P. Mukherjee, D. Parkinson, A.R. Liddle, A nested sampling algorithm for cosmological model selection, Astrophys. J. 638 (2006) L51–L54.
[22] D. Parkinson, P. Mukherjee, P. Liddle, A Bayesian model selection analysis of WMAP3, Phys. Rev. D 73 (2006) 123523.
[23] A. Ben Abdessalem, Model selection, updating and prediction of fatigue crack propagation using nested sampling algorithm, in: Conference: 23 Congrès Français de Mécanique, 2017, Lille, France.
[24] A. Ben Abdessalem, F. Jenson, P. Calmon, Quantifying uncertainty in parameter estimates of ultrasonic inspection system using Bayesian computational framework, Mech. Syst. Signal Process. 109 (2018) 89–110.
[25] M.A. Beaumont, W. Zhang, D.J. Balding, Approximate Bayesian computation in population genetics, Genetics 162 (4) (2002) 2025–2035.
[26] B.M. Turner, T. Van Zandt, A tutorial on approximate Bayesian computation, J. Math. Psychol. 56 (2012) 69–85.
[27] A. Ben Abdessalem, N. Dervilis, D. Wagg, K. Worden, Automatic kernel selection for Gaussian processes regression with approximate Bayesian computation and sequential Monte Carlo, Front. Built Environ. 3:52 (2017), https://doi.org/10.3389/fbuil.2017.00052.
[28] M. Chiachio, J.L. Beck, J. Chiachio, G. Rus, Approximate Bayesian computation by subset simulation, SIAM J. Scientific Comput. 36 (3) (2014), A1339-A1338.
[29] A. Ben Abdessalem, N. Dervilis, D. Wagg, K. Worden, Identification of nonlinear dynamical systems using approximate Bayesian computation based on a sequential Monte Carlo sampler, in: International Conference on Noise and Vibration Engineering, September 19-21, 2016, Leuven (Belgium).
[30] M.K. Vakilzadeh, J.L. Beck, T. Abrahamsson, Using approximate Bayesian computation by subset simulation for efficient posterior assessment of dynamic state-space model classes, SIAM J. Scientific Comput. 40 (1) (2018) B168–B195.

[31] A. Ben Abdessalem, N. Dervilis, D. Wagg, K. Worden, Model selection and parameter estimation in structural dynamics using approximate Bayesian computation, Mech. Syst. Signal Process. 99 (2018) 306–325.
[32] A. Ben Abdessalem, N. Dervilis, D. Wagg, K. Worden, ABC-NS: a new computational inference method applied to parameter estimation and model selection in structural dynamics, in: Conference: 23 Congrès Français de Mécanique, 2017, Lille, France.
[33] P. Marjoram, J. Molitor, V. Plagnol, S. Tavare, Markov chain Monte Carlo without likelihoods, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 15324–15328.
[34] S. Sisson, Y. Fan, M. Tanaka, Sequential Monte Carlo without likelihoods, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 1760–1765.
[35] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, M.P.H. Stumpf, Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems, J. R. Soc. Interface 6 (2009) 187–202.
[36] F.V. Bonassi, Approximate Bayesian Computation for Complex Dynamic Systems (Ph.D. thesis), Duke University, 2013.
[37] F.V. Bonassi, M. West, Sequential Monte Carlo with adaptive weights for approximate bayesian computation, Bayesian Anal. 10 (1) (2015) 171–187.
[38] F. Feroz, M.P. Hobson, M. Bridges, MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics, Monthly Notice R. Astron. Soc. 398 (4) (2009) 1601–1614.
[39] J.R. Shaw, M. Bidges, M.P. Hobson, Efficient Bayesian inference for multimodal problems in cosmology, Monthly Notice R. Astron. Soc. 000 (2006) 1–7.
[40] E. Jennings, M. Madigan, astroABC: an approximate bayesian computation sequential Monte Carlo sampler for cosmological parameter estimation, Astron. Comput. 19 (2017) 16–22.
[41] D. Allingham, R.A.R. King, K.L. Mengersen, Bayesian estimation of quantile distributions, Stat. Comput. 19 (2009) 189–201.
[42] Christopher C. Drovandi, Anthony N. Pettitt, Likelihood-free Bayesian estimation of multivariate quantile distributions, Comput. Stat. Data Anal. 55 (2011) 2541–2556.
[43] U. Picchini, R. Anderson, Approximate maximum likelihood estimation using data-cloning ABC, Comput. Stat. Data Anal. 105 (2017) 166–183.
[44] G. Rayner, H. Macgillivray, Weighted quantile-based estimation for a class of transformation distributions, Comput. Stat. Data Anal. 39 (4) (2002) 401–433.
[45] H. Walach, Ockham's razor, in: Wiley Interdisciplinary Reviews Computational Statistics, vol. 2, Sage Publications, 2007.
[46] C.O.S.T. Technical Report, Action F3, VTT Technical, Research Centre of Finland (1999).
[47] M. Juntunen, J. Linjama, Presentation of the VTT benchmark, Mech. Syst. Signal Process. 17 (1) (2003) 179–182.
[48] G. Kerschen, K. Worden, Alexander F. Vakakisc, J.C. Golinval, Past present and future of nonlinear system identification in structural dynamics, Mech. Syst. Signal Process. 20 (2006) 505–592.
[49] G. Kerschen, On the Model Validation in Non-linear Structural Dynamics (Thèse de doctorat), Université de Liège, 2002.
[50] M. Peifer, J. Timmer, H.U. Voss, Non-parametric identification of non-linear oscillating systems, J. Sound Vib. 267 (2003) 1157–1167.
[51] G.F. Demetriades, M.C. Constantinou, A.M. Reinhorn, Study of wire rope systems for seismic protection of equipment in buildings, Eng. Struct. 15 (5) (1993) 321–334.
[52] M.C. Constantinou, I.G. Tadjbakhsh, Hysteretic dampers in base isolation: random approach, J. Struct. Eng. 111 (4) (1985).
[53] Y. Wen, Method for random vibration of hysteretic systems, ASCE J. Eng. Mech. Division 102 (2) (1976) 249–263.