Reinforcement learning for model selection applied to a nonlinear dynamical system

Thiago G. Ritto[1,2], Sandor Beregi[2], David A.W. Barton[2]

[1]*Department of Mechanical Engineering, Universidade Federal do Rio de Janeiro*
Cidade Universitária - Ilha do Fundão, 21945-970, Rio de Janeiro, Brazil
tritto@mecanica.ufrj.br
[2]*Faculty of Engineering, University of Bristol Beacon House, Queens Road, BS8 1QU, Bristol, UK*
sandor.beregi@bristol.ac.uk, david.Barton@bristol.ac.uk

Abstract. In the context of digital twins, and the integration of physics-based models with machine learning tools, this paper proposes a new methodology for model selection and parameter identification, applied to nonlinear dynamic problems. Reinforcement learning is used for model selection through Thompson sampling, and parameter identification is performed using approximate Bayesian computation (ABC). These two methods are applied together in a one degree-of-freedom nonlinear dynamic model. Experimental data are used in the analysis, and two different nonlinear models are tested. The initial Beta distribution of each model is updated according to how successful the model is at representing the reference data (reinforcement learning strategy). At the same time, the prior Uniform distribution of the model parameters is also updated using a likelihood free strategy (ABC). In the end, the models' rewards and the posterior distribution of the parameters of each model are obtained. Several analyses are made and the potential of the proposed methodology is discussed.

Keywords: nonlinear dynamics, parameter identification, model selection, reinforcement learning, ABC

1 Introduction

A Digital twin (DT) is more than a computational model. It is a framework that fuses important elements aimed at supporting management decisions about a specific asset, such as sensing, data, computational model, and learning [1, 2]. In this context, choosing the most appropriate model (given the experimental data) and taking into account uncertainties are paramount. In addition, the combination of physics-based models with machine learning tools can leverage the DT capabilities [2, 3]. The present paper integrates reinforcement learning (RL) for model selection and approximate Bayesian computation (ABC) for parameter estimation in a nonlinear problem.

The problem considered is a forced nonlinear oscillator subjected to noise [4]. A Duffing-like model is considered and the control-based continuation strategy [5, 6] is used to measure experimentally both stable and unstable orbits.

RL is a machine learning technique where the learner must discover which actions yield the largest reward by trying them [7]. We are particularly interested in selecting the most appropriate nonlinear dynamical model given a set of experimental data, i.e. the one that better explains the data under analysis. We are inspired by the multi-armed bandit problem [7] where an agent selects one over n different options. After each choice it receives a reward depending on the selection made. It provides a trade-off between exploration (trying different arms) and exploitation (playing the arm with best results).

ABC has been used for model selection and parameter estimation [8]. This Bayesian strategy is very convenient because it is likelihood-free; hence, there is no need to construct a likelihood function. Here we use a similar framework to [8]. The key difference is that we apply RL for the model selection instead of considering a Uniform distribution to select the models.

ABC has been combined with RL in a control policy problem [9]. However, the application here and the way we employ the techniques are different. Furthermore, we focus our analysis in the selection of the best nonlinear dynamic model and, at the same time, update the probability density function of each model parameter given

some experimental data. We propose to use Thompson sampling [10] (Beta-Bernoulli bandit) to help in selecting the most appropriate model. Since the parameters of the model are not known, a prior Uniform distribution is considered and the ABC strategy [8] is used to construct the posterior distribution for the parameters. We believe that many applications might benefit from this new methodology since it provides a simple and efficient strategy for model selection and parameter estimation of DTs.

The next section depicts the proposed methodology (RL and ABC) for model selection and parameter calibration. Section 3 presents the nonlinear model analysed. The numerical results are shown in Section 4, and the concluding remarks are made in the last section.

2 Methodology for model selection and parameter calibration

This section depicts the methodology proposed in this paper. First, the reinforcement learning for model selection is discussed, then the ABC algorithm for parameter estimation is introduced. In the end of the section, the proposed algorithm is presented.

2.1 Reinforcement learning – RL

A variation of the Thompson algorithm usually applied to multi-armed bandit problems is considered [10, 11]. We propose to use this strategy in the context of model selection [12]. Consider a Beta distribution for each model,

$$\pi(\theta, \mathcal{M}) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\,, \tag{1}$$

where α and β are the parameters of the Beta distribution with $\theta \in [0,1]$. The Bayesian framework is used to update the Beta probability density functions (PDFs) over Bernoulli trials. Since Beta conjugates with Bernoulli, for each new observation the parameters of the Beta distribution are updated according to

$$\alpha_{new} = \alpha_{old} + r \quad , \quad \beta_{new} = \beta_{old} + (1-r)\,, \tag{2}$$

where r is the reward which is equal to one if the model is rewarded, and equals to zero otherwise. First, a sample from each models' distribution is generated (θ_k) with $\alpha = \beta = 2$. The model corresponding to the highest value is chosen, then its parameters are updated according to the reward. The model is rewarded depending on how close its results are from the experimental data set. If it is close enough (see next section), then $r = 1$, otherwise $r = 0$.

In the beginning, the probability of choosing any of the models is the same. As the problem evolves, the models with more rewards are more likely to be selected. Note that adding one to α moves the Beta distribution to the right (closer to one), and adding one to β moves the Beta distribution to the left side.

2.2 Approximate Bayesian computation – ABC

The Bayesian approach has been widely used for statistic inverse problems [13]. In this approach, the parameters $\boldsymbol{\varphi}$ of model $\mathcal{M}$ are treated as random variables, whose probability density function (pdf) is updated by means of the Bayes formula

$$\pi(\boldsymbol{\varphi}\,|\mathbf{y}, \mathcal{M}) = \frac{\pi(\mathbf{y}\,|\boldsymbol{\varphi}, \mathcal{M})\,\pi(\boldsymbol{\varphi}, \mathcal{M})}{\pi(\mathbf{y}, \mathcal{M})}\,, \tag{3}$$

where $\boldsymbol{\varphi}$ is the vector composed of the parameters of model $\mathcal{M}$, and $\mathbf{y}$ represents the data used in the learning process. The terms shown in the equation are: the posterior PDF of $\boldsymbol{\varphi}$, which is $\pi(\boldsymbol{\varphi}\,|\mathbf{y}, \mathcal{M})$; the prior PDF $\pi(\boldsymbol{\varphi}, \mathcal{M})$; the likelihood function $\pi(\mathbf{y}\,|\boldsymbol{\varphi}, \mathcal{M})$; and the normalisation constant $\pi(\mathbf{y}, \mathcal{M})$. Usually an additive noise is considered

$$\mathbf{y} = \mathbf{y}_m(\boldsymbol{\varphi}, \mathcal{M}) + \mathbf{e}\,, \tag{4}$$

in which $\mathbf{y}_m(\cdot)$ represents the model prediction, and $\mathbf{e}$ is, for instance, a Gaussian added noise. From the additive noise model, we can obtain the likelihood function

$$\pi(\mathbf{y}\,|\,\boldsymbol{\varphi}, \mathcal{M}) = \pi_{\mathbf{e}}(\mathbf{y} - \mathbf{y}_m(\boldsymbol{\varphi}, \mathcal{M}))\,. \tag{5}$$

Unfortunately, we do not know the structure of the error and, consequently, the likelihood function. To circumvent this problem we apply the approximate Bayesian computation (ABC) [8], where instead of assuming a

likelihood function we make a direct comparison of the model prediction (with parameter φ^*) and the experiment, for instance

$$\rho(\mathbf{y}, \mathbf{y}_m(\boldsymbol{\varphi}^*, \mathcal{M})) = \frac{||\mathbf{y} - \mathbf{y}_m(\boldsymbol{\varphi}^*, \mathcal{M})||^2}{||\mathbf{y}||^2} \,. \tag{6}$$

To obtain the updated posterior PDF of parameters, a simple rejection method is considered [8].

2.3 Proposed algorithm

1 Sample from the Beta distributions (with parameters α and β) and choose the candidate model $\mathcal{M}^*$ with the greatest θ;
2 Sample a candidate set of parameters $\boldsymbol{\varphi}^*$ from $\pi(\boldsymbol{\varphi}|\mathcal{M}^*)$;
3 Compute the prediction $\mathbf{y}_m(\boldsymbol{\varphi}^*)$;
4 Evaluate the results using the metric $\rho(\mathbf{y}, \mathbf{y}_m(\boldsymbol{\varphi}^*))$. Accept $\mathcal{M}^*$ and $\boldsymbol{\varphi}^*$ if the error is lower than a threshold ϵ, and set the reward $r = 1$. Otherwise, reject $\mathcal{M}^*$ and $\boldsymbol{\varphi}^*$ and set $r = 0$.
5 Update the parameters of the Beta distribution of the model $\mathcal{M}^*$ according to the reward ($\alpha^* = \alpha^* + r$ and $\beta^* = \beta^* + (1 - r)$)
6 Go back to 1.

For the RL, initially we set $\alpha = \beta = 2$, which yields a symmetric Beta distribution with mean equals to $1/2$. For ABC, we consider independent random Uniform random variables.

3 Nonlinear dynamical system

The model and experimental data considered here were taken from [4]. It consists of a Duffing-like oscillator with the equation of motion

$$\ddot{x}(t) + b\dot{x}(t) + \omega_n^2 x(t) + \mu x^3(t) + \nu x^5(t) + \rho x^7(t) = A\cos(\omega t)\,, \tag{7}$$

where t is the time, ω_n is the natural frequency of the system, b is the damping parameter, μ, ν and ρ are the constants related to the nonlinear terms, A is the force amplitude, and ω is the forcing frequency.

The second model considered in this work is a simpler model, where $\rho = 0$:

$$\ddot{x}(t) + b\dot{x}(t) + \omega_n^2 x(t) + \mu x^3(t) + \nu x^5(t) = A\cos(\omega t)\,, \tag{8}$$

We want to know which one of these two models (Eqs. 7 or 8) is the best to represent a specific data set, according to the proposed strategy. Even though the first model is more elaborated and might produce a lower error, the strategy we employed takes into account other ingredients such as parameter uncertainties. The model classification depends, for instance, on the prior distribution of the parameters and on the sensitivity of the response with respect to them [14].

If we consider the non-dimensional time $\tau = \omega_n t$, then $d/dt = \omega_n d/d\tau$, and Eq. 7 becomes

$$x''(\tau) + \hat{b}x'(\tau) + x(\tau) + \hat{\mu}x^3(\tau) + \hat{\nu}x^5(\tau) + \hat{\rho}x^7(\tau) = \delta_{st}\cos(\zeta\tau)\,, \tag{9}$$

where $'$ is the derivative with respect to τ, $\zeta = \omega/\omega_n$, $\hat{b} = b/\omega_n$, $\hat{\mu} = \mu/\omega_n^2$, $\hat{\nu} = \nu/\omega_n^2$, $\hat{\rho} = \rho/\omega_n^2$, and $\delta_{st} = A/\omega_n^2$. The steady-state periodic solutions of the oscillator can be obtained [4] by the method of multiple scales [15]. This method provides an analytical expression for the forcing amplitude as a function of the amplitude of the fundamental harmonic component of the steady-state solution X, thus providing an implicit representation of the solution. For model 1, this expression of the forcing amplitude reads

$$\mathbf{y}_{m1} = \frac{|y_{Am1} \times y_{Bm1}|}{\delta_{st}}\,, \tag{10}$$

in which

$$
\begin{aligned}
y_{Am1} &= \left((35/64)X^7\hat{\rho} + (5/8)X^5\hat{\nu} + (3/4)X^3\hat{\mu} - X(\zeta^2-1)\right) , \\
y_{Bm1} &= \left(\frac{(\zeta^2 - 1 - (35/64)X^6\hat{\rho} - (5/8)X^4\hat{\nu} - (3/4)X^2\hat{\mu})^2 + \hat{b}^2\zeta^2}{(\zeta^2 - 1 - (35/64)X^6\hat{\rho} - (5/8)X^4\hat{\nu} - (3/4)X^2\hat{\mu})^2} \right)^{1/2} ,
\end{aligned}
\tag{11}
$$

where the following property was used: $\cos(\text{atan}(\psi)) = 1/\sqrt{1+\psi^2}$. For model 2, the analytical expression for the forcing amplitude as a function of the amplitude of the fundamental harmonic component of the steady-state solution X is obtained considering $\rho = 0$,

$$
\mathbf{y}_{m2} = \frac{|y_{Am2} \times y_{Bm2}|}{\delta_{st}} , \tag{12}
$$

in which

$$
\begin{aligned}
y_{Am2} &= \left((5/8)X^5\hat{\nu} + (3/4)X^3\hat{\mu} - X(\zeta^2-1)\right) , \\
y_{Bm2} &= \left(\frac{(\zeta^2 - 1 - (5/8)X^4\hat{\nu} - (3/4)X^2\hat{\mu})^2 + \hat{b}^2\zeta^2}{(\zeta^2 - 1 - (5/8)X^4\hat{\nu} - (3/4)X^2\hat{\mu})^2} \right)^{1/2} .
\end{aligned}
\tag{13}
$$

4 Numerical results

The simulated results must be compared with the available experiments. To this end, we take the experimental data (red circles) shown in Fig. 1 (left), and consider a function with respect to X (the vibration amplitude). This allows us to use least squares regression to fit a polynomial to the experimental data, which will be used to compute the error between simulation and experiments. To respect the physics of the system, X must go to zero as the amplitude of excitation A goes to zero, we solve

$$
[X]\mathbf{p} = \mathbf{y}_{points} , \tag{14}
$$

where $\mathbf{p}$ is the vector composed of the polynomial coefficients to be found, $\mathbf{y}_{points} = [A_1 \, ... \, A_n]^T$ is the vector with the n experimental points. To fit a seventh-order polynomial that passes through the origin, the Vandermonde matrix $[X]$ is constructed suppressing the first column with one in each entry. Applying the least squares approximation, $\mathbf{p} = ([X]^T[X])^{-1}[X]^T\mathbf{y}_{points}$, the function obtained is given by:

$$
\mathbf{y}_{exp} = 0.53X - 4.72X^2 + 16.11X^3 - 26.49X^4 + 22.18X^5 - 10.02X^6 + 3.11X^7 . \tag{15}
$$

Figure 1 (right) shows the results of models 1 and 2, using the identified parameters, together with the experimental data. Both models are able to reproduce the main characteristic of the system, although the shape they produce are a little different. Model 1 presents a lower error (0.17%) compared with model 2 (0.80%).

Figure 2 shows the prior and posterior PDF of the parameters for the two models analysed. The initial Uniform distribution is updated after the calibration process. The cubic parameter μ is the one that gained the most information in model 1, in the sense that its final distribution is dissimilar to the Uniform one. For model 2, the parameter ν also gained considerable information, and it can be seen that it is less likely to return lower values, in the range analysed. The set of parameters that yields the lower error, comparing with the experiment, is $\mu_1 = 0.423$, $\nu_1 = -0.105$, $\rho_1 = 0.011$ and $b_1 = 0.002$ for model 1 and $\mu_2 = 0.423$, $\nu_2 = -0.093$ and $b_2 = 0.008$ for model 2.

Figure 3 shows that there is a negative correlation between parameters μ and ν for model 1 (-0.65) and model 2 (-0.56). As μ increases, ν decreases. Note also that the values of μ are a little greater for model 2, which compensates the fact that it considers $\rho = 0$.

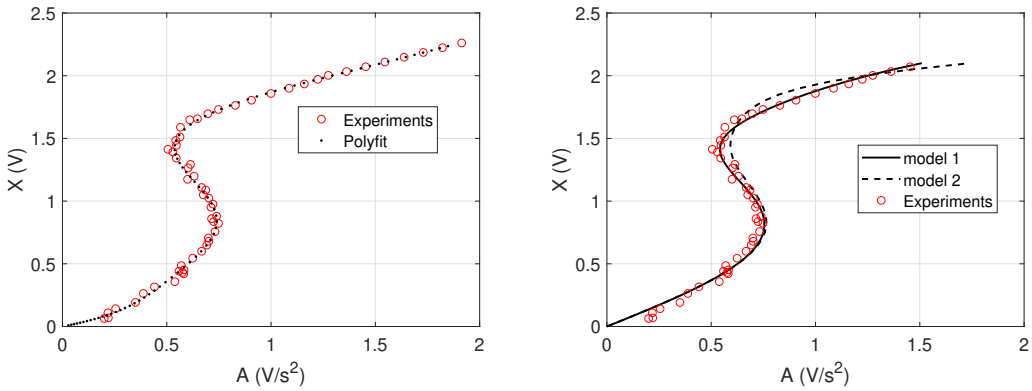

Figure 1. Left: experimental data (red circles) and *polyfit* function obtained via least squares. Right: comparison of the simulations with the experimental data. The black continuous and dashed lines are the results of models 1 and 2 obtained with the identified parameters.

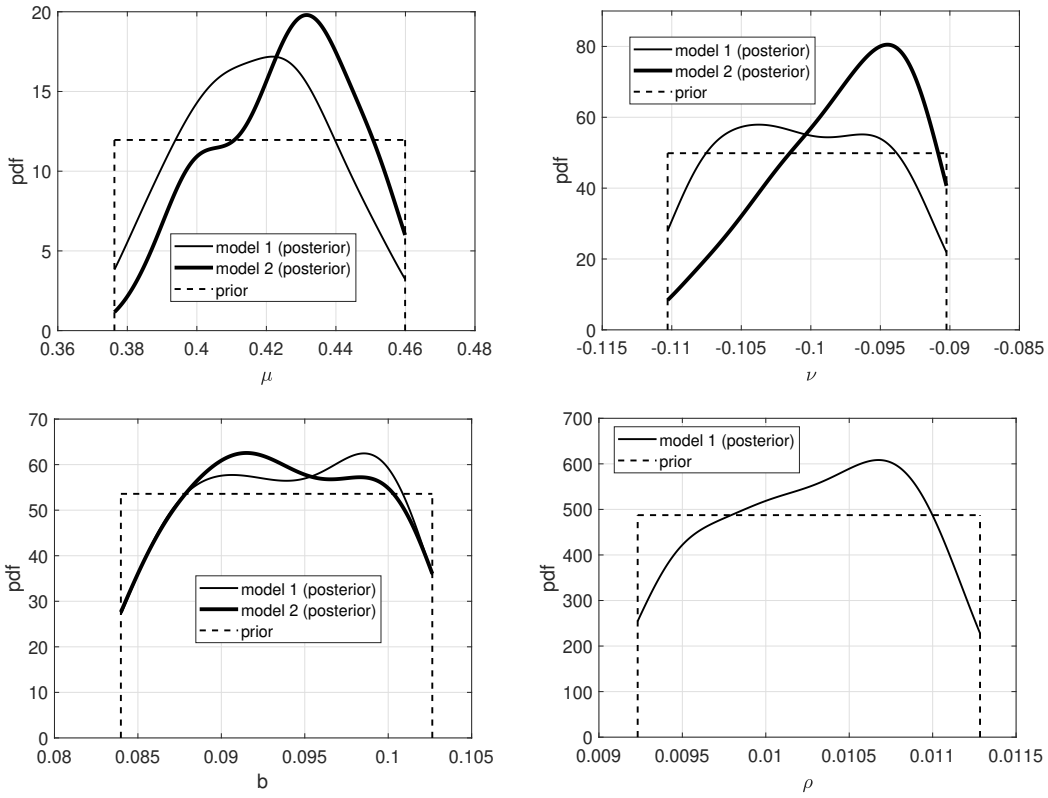

Figure 2. Prior and posterior probability density functions (pdf) of the parameters of models 1 and 2. The parameters μ, ν and b appear in both models ($\mu x^3 + \nu x^5 + bx'$), but only model 1 has ρ (related to ρx^7).

Figure 4 shows the 95% probabilistic envelope considering the calibrated models (posterior distributions), together with the experiments. The stochastic model encompasses the available experiments. However, the probabilistic envelopes of the two models are quite different for high amplitudes of excitation. For both models, the probabilistic envelope is very thin if the amplitude of the excitation is small. Up to $V = 0.5V/s^2$, the non-linearity is not activated; remember that the uncertain parameters are related to the nonlinear part of the equation (cubic, 5th-order, and 7th-order terms). As the amplitude of excitation increases the envelopes get wider.

Figure 5 shows the reward attributed to each model and their prior/posterior Beta PDFs. After each simulation, a reward (0 or 1) is given for the model that was chosen. It can be seen that model 2 is the first one selected, and it is positively rewarded, but, as the simulations evolve, model 1 outperforms model 2, considering: (i) the available

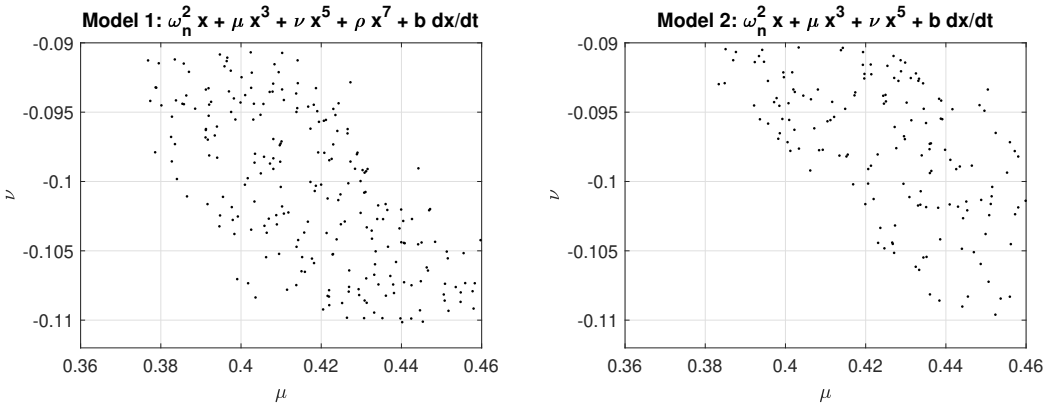

Figure 3. Correlation between parameters μ and ν: model 1 (left) and model 2 (right).

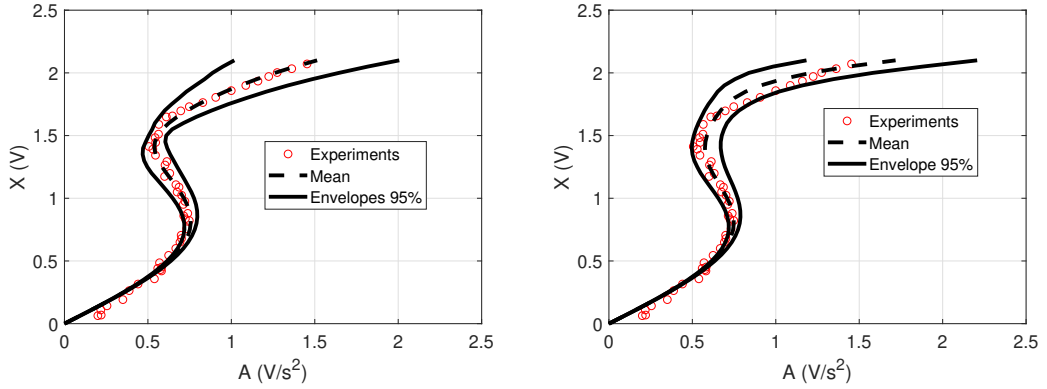

Figure 4. Experiments (red circles) together with the probability envelope of model 1 (left) and model 2 (right). The black continuous line represents the percentiles of 2.5 and 97.5% and the dashed line is the mean.

experiments, (ii) the interval of the parameters, and (iii) the value of the threshold ϵ. Figure 5 (right) shows the initial Beta PDF with $\alpha = 2$ and $\beta = 2$. The updated PDFs are thinner with mean greater than 50%. For model 1 the parameter values reach $\alpha_1 = 237$ and $\beta_1 = 177$ (model 1) and $\alpha_2 = 158$ and $\beta_2 = 136$ (model 2). The conclusion is that both models are very good to explain the experiment analysed.

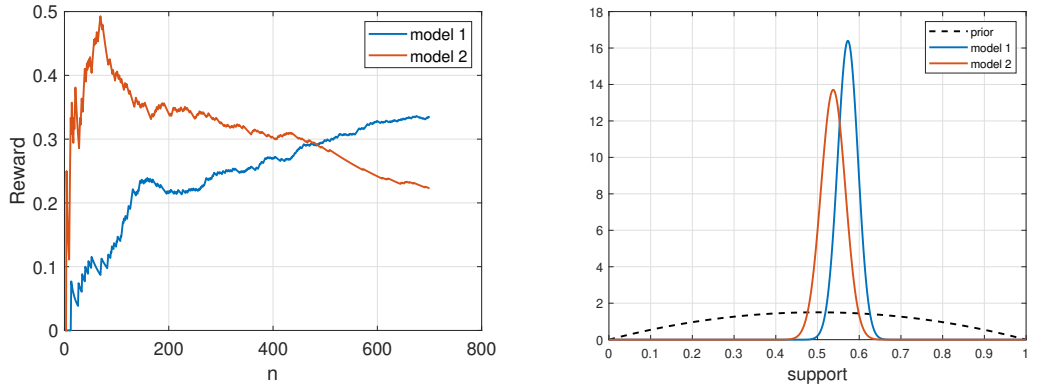

Figure 5. Left: reward attributed to each model along the 700 simulations. Right: prior and posterior Beta distribution for each model.

5 Concluding remarks

This paper proposes a methodology to simultaneously calibrate the parameters and select models to match experimental data using approximate Bayesian computation (ABC) to update the prior distribution over the model parameters, and reinforcement learning (RL) to select models. It was applied to a nonlinear Duffing-like system. A model that considers the cubic, 5th and 7th-order terms is compared with another that disregards the 7th-order term.

The results show that the strategy seems to work properly. It was successful in selecting the best model and updating the PDFs of the parameters. The parameters related to the cubic and 5th-order terms are the ones that gained more information in the process, while the PDFs of the damping and the 7th-order term parameter remained close to the Uniform distribution. In addition, the cubic and the 5th-order term parameter presented some negative correlation. The next steps of this investigation are to consider (i) time dependent parameters and (ii) a more efficient ABC strategy (e.g. MCMC).

Acknowledgments. The first author would like to acknowledge that this investigation was financed in part by the Brazilian agencies: *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES) - Finance code 001 - Grant PROEX 803/2018 and CAPES-PRINT - Grant 88887.569759/2020-00, and *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPQ) - Grant 400933/2016-0. The other two authors would like to acknowledge the support of the Engineering and Physical Sciences Research Council (EPSRC) via grant number EP/R006768/1.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] D. Wagg, K. Worden, R. Barthorpe, and P. Gardner. Digital twins: State-of-the-art and future directions for modelling and simulation in engineering dynamics applications. ASME, vol. 6, n. 3, pp. 030901, 2020.
[2] T. G. Ritto and F. A. Rochinha. Digital twin, physics-based model, and machine learning applied to damage detection in structures. Mechanical Systems and Signal Processing, vol. 155, pp. 107614, 2021.
[3] K. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar. Integrating physics-based modeling with machine learning: A survey. ArXiv, vol. 2003.04919v3, pp. 1–11, 2020.
[4] S. Beregi, D. A. W. Barton, D. Rezgui, and S. A. Neild. Robustness of nonlinear parameter identification in the presence of process noise using control-based continuation. Nonlinear Dynamics, vol. 104, pp. 885–900, 2021.
[5] J. Sieber, A. Gonzalez-Buelga, S. Neild, D. Wagg, and B. Krauskopf. Experimental continuation of periodic orbits through a fold. Physical Review Letters, vol. 100, n. 24, pp. 244101, 2008.
[6] D. Barton and S. Burrow. Numerical continuation in a physical experiment: investigation of a nonlinear energy harvester. Journal of Computational and Nonlinear Dynamics, vol. 6, n. 1, pp. 011010, 2011.
[7] R. S. Sutton and A. G. Barto. Reinforcement Learning: An Introduction. MIT Press, 2nd edition, 2018.
[8] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. Stumpf. Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of the Royal Society Interface, vol. 6, n. 31, pp. 187–202, 2009.
[9] C. Dimitrakakis and N. Tziortziotis. Abc reinforcement learning. In 30th International Conference on Machine Learning, ICML, pp. 1721–1729, 2013.
[10] D. Russo, B. V. Roy, A. Kazerouni, I. Osband, and Z. Wen. Tutorial on thompson sampling. Foundations and Trends in Machine Learning, vol. 11, n. 1, pp. 1–96, 2018.
[11] O.-C. Granmo. Solving two-armed bernoulli bandit problems using a bayesian learning automaton. International Journal of Intelligent Computing and Cybernetics, vol. 2, n. 3, pp. 207–234, 2010.
[12] J. Beck and K.-V. Yuen. Model selection using response measurements: Bayesian probabilistic approach. Journal of Engineering Mechanics, vol. 130, n. 2, pp. 192–203, 2004.
[13] J. Kaipio and E. Somersalo. Statistical and Computational Inverse Problems. Springer, 2004.
[14] T. Ritto and L. Nunes. Bayesian model selection of hyperelastic models for simple and pure shear at large deformations. Computers and Structures, vol. 156, pp. 101–109, 2015.
[15] A. H. Nayfeh. Introduction to Perturbation Techniques. Wiley, New York, 1981.