This is a repository copy of *Computational system identification of continuous-time nonlinear systems using approximate Bayesian computation*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/93277/

Version: Accepted Version

Article:
Krishnanathan, K., Anderson, S.R., Billings, S.A. et al. (1 more author) (2015) Computational system identification of continuous-time nonlinear systems using approximate Bayesian computation. International Journal of Systems Science. ISSN 0020-7721

https://doi.org/10.1080/00207721.2015.1090643

Reuse
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

To appear in the *International Journal of Systems Science*
Vol. 00, No. 00, Month 20XX, 1–14

Computational system identification of continuous-time nonlinear systems using approximate Bayesian computation

Kirubhakaran Krishnanathan[a] and Sean R. Anderson[a*] and Stephen A. Billings[a] and Visakan Kadirkamanathan[a]

[a]*Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, S1 3JD, UK*

(*Received 00 Month 20XX; final version received 00 Month 20XX*)

In this paper we derive a system identification framework for continuous-time nonlinear systems, for the first time using a simulation-focused computational Bayesian approach. Simulation approaches to nonlinear system identification have been shown to outperform regression methods under certain conditions, such as non-persistently exciting inputs and fast-sampling. We use the approximate Bayesian computation (ABC) algorithm to perform simulation-based inference of model parameters. The framework has the following main advantages: (i) parameter distributions are intrinsically generated, giving the user a clear description of uncertainty, (ii) the simulation approach avoids the difficult problem of estimating signal derivatives as is common with other continuous-time methods, and (iii) as noted above, the simulation approach improves identification under conditions of non-persistently exciting inputs and fast-sampling. Term selection is performed by judging parameter significance using parameter distributions that are intrinsically generated as part of the ABC procedure. The results from a numerical example demonstrate that the method performs well in noisy scenarios, especially in comparison to competing techniques that rely on signal derivative estimation.

Keywords: models; NARMAX; continuous-time systems; system identification and signal processing; Bayesian estimation; computational system identification; nonlinear; approximate Bayesian computation

1. Introduction

In the field of nonlinear system identification (SID), there are many techniques available for obtaining discrete-time models, where the model structure is described by a difference equation based on sampled data (Baldacchino, Anderson, & Kadirkamanathan, 2012; Chen, Billings, & Luo, 1989; Kukreja, Galiana, & Kearney, 2004; K. Li, Peng, & Bai, 2006; Piroddi & Spinelli, 2003). There are, however, far fewer techniques available for the identification of nonlinear continuous-time (CT) models, where the model structure is described by a differential equation (Billings, 2013). There are a number of reasons why CT models are attractive for SID: (i) they are easier to interpret and can to some extent facilitate physical understanding, (ii) they tend to be compact, (iii) they permit identification for irregularly sampled data and (iv) they exhibit more stability and less ill-conditioning (Garnier & Wang, 2008).

There are two main types of approach to continuous-time SID (CT-SID) from sampled data, which are the direct and indirect methods (Billings, 2013; Rao & Unbehauen, 2006; Unbehauen & Rao, 1990), i.e. direct identification of the model in CT typically using estimates of signal derivatives, or indirect identification by first identifying a discrete-time model and then secondly mapping it to the CT domain (L. Li & Billings, 2001). The direct approach is more efficient because

*Corresponding author. Email: s.anderson@sheffield.ac.uk

it requires only a single step of identification, without the complication of mapping the nonlinear model between discrete- and continuous-time domains. However, direct identification does usually incur the problem of signal derivative estimation.

Current approaches for direct identification of CT nonlinear systems make use of a number of approaches for signal derivative estimation: delayed state-variable filters (Tsang & Billings, 1994), Kalman smoothing (Coca & Billings, 1999), the delta-operator (Anderson & Kadirkamanathan, 2007) and differencing methods that account for the nonlinear zero-dynamics (Yuz & Goodwin, 2005). An advantage of the simulation approach for CT-SID is that it avoids the need to estimate derivatives from sampled data, and for discrete-time models simulation-based SID has been promoted as improving both term selection and parameter estimation under conditions of non-persistently exciting inputs and fast sampling (Billings, 2013; Piroddi & Spinelli, 2003). A simulation-based method for nonlinear CT-SID has not yet been developed. This highlights a gap, which this paper aims to fill, by developing a novel simulation-based approach to nonlinear CT-SID, addressing the problem of oversampling and overfitting that can lead to poor prediction.

In this investigation, we develop the identification algorithm in a Bayesian framework, which intrinsically generates information on uncertainty in parameters (Peterka, 1981). Characterisation of uncertainty is important in control engineering (Gevers, 2005), but also in other areas where SID is now commonly applied such as the life sciences (Anderson, Lepora, Porrill, & Dean, 2010; Krishnanathan, Anderson, Billings, & Kadirkamanathan, 2012; Kukreja, Galiana, & Kearney, 2003). In SID, computational Bayesian (or probabilistic) methods are gaining popularity due to advances in processing power (Baldacchino, Anderson, & Kadirkamanathan, 2013; Falsone, Piroddi, & Prandini, 2015; Ninness & Henriksen, 2010). The computational estimation framework we develop here for CT-SID is based on approximate Bayesian computation (ABC), which is a rejection sampling algorithm (Beaumont, Zhang, & Balding, 2002; Tavare, Balding, Griffiths, & Donnelly, 1997). In practice, we use a computationally more efficient approach to ABC, known as ABC-sequential Monte Carlo (ABC-SMC), which reduces computation time (Sisson, Fan, & Tanaka, 2007, 2009; Toni, Welch, Strelkowa, Ipsen, & Stumpf, 2009).

In Bayesian estimation for dynamic systems, the ABC methods have become popular due to the fact that they can be used on a range of problems without modification, for instance (i) systems described by either ordinary, delay, deterministic or stochastic differential equations, (ii) systems that have different distributions of noise (e.g. Gaussian, Laplacian, uniform) and (iii) for model selection (Toni et al., 2009). Currently, ABC methods have only been applied to parameter estimation and model selection problems for known model structures. The full SID problem includes the important case where the model structure is unknown *a priori*. Hence, a further key novel contribution of this work is the extension of ABC methods to the case where the nonlinear model structure is unknown. This is a non-trivial extension because structure detection in nonlinear SID is an extremely challenging problem (Sjoberg et al., 1995).

To solve the structure detection problem, the inspiration for our basic approach comes from Kukreja et al. (2004), where structure detection for a discrete-time nonlinear model was solved by one-step-ahead regression with bootstrapping: in that work, parameter distribution ranges generated by bootstrapping for a superset of potential model terms were checked to see if they included zero - these terms were pruned from the superset resulting in a parsimonious model structure. In the case of our proposed algorithm for CT-SID, parameter distributions intrinsically generated by ABC-SMC for a superset of model terms are similarly checked to see if they include zero - these terms are judged to be unnecessary and are pruned from the model. We extend this basic selection algorithm here by identifying model terms that improve simulation performance, enhancing the robustness of the overall identification scheme. The main result is a new algorithm for nonlinear CT-SID, in a computational Bayesian framework.

The paper is structured as follows. In section 2 we define the nonlinear model representation and describe the ABC-SMC parameter estimation algorithm. In section 3 we develop the identification framework. Numerical examples of this new identification framework are given in section 4, along

with a comparison to one of few established nonlinear CT-SID methods (Coca & Billings, 1999). Finally, the contributions of the paper are summarised in section 5.

2. Model definition and parameter estimation

2.1. *Continuous-time nonlinear model representation*

The output $y(t) \in \mathbb{R}$ of a continuous-time nonlinear output-error process can be represented as

$$z^{(n)}(t) = F(\mathbf{x}(t)), \tag{1}$$
$$y(t) = z(t) + e(t), \tag{2}$$

where $z(t) \in \mathbb{R}$ is the unknown noise-free system output, $u(t) \in \mathbb{R}$ is the known system input, $z^{(n)}(t) \in \mathbb{R}$ indicates the n^{th} derivative of $z(t)$, and the measurement noise $e(t)$ can follow a range of distributions, e.g. Gaussian, uniform or Laplacian. The function $F(.)$ describes the dynamics of the nonlinear CT process and $\mathbf{x}(t) \in \mathbb{R}^{2n}$ is the vector of input-output derivatives,

$$\mathbf{x}(t) = \left(z(t), \ldots, z^{(n-1)}(t), u(t), \ldots, u^{(n-1)}(t)\right). \tag{3}$$

The nonlinear function $F(\mathbf{x}(t))$ can be decomposed and represented by a linear sum of basis functions $\phi_j(\mathbf{x}(\mathbf{t}))$ which can have varying forms including wavelet, polynomial or radial functions,

$$F(\mathbf{x}(t)) = \sum_{j=1}^{N_\theta} \theta_j \phi_j(\mathbf{x}(t)), \tag{4}$$

where N_θ is the number of model terms and $\theta_j \in \mathbb{R}$ is the parameter associated with basis function $\phi_j(.)$.

2.2. *Parameter estimation by approximate Bayesian computation: ABC*

The Bayesian inference problem considered here is the estimation of the conditional distribution of model parameters, $\boldsymbol{\theta}$, given the observed output data samples, $\mathbf{y}$,

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{p(\mathbf{y})}, \tag{5}$$

where $p(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood function, $\pi(\boldsymbol{\theta})$ is the prior distribution, $p(\mathbf{y})$ is the marginal likelihood and

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{N_\theta})^\top, \tag{6}$$
$$\mathbf{y} = \left(y(t_0), y(t_1), \ldots, y(t_{N_y-1})\right)^\top, \tag{7}$$

where $y(t_k)$ is a sampled observation at sample time t_k, where $t_k = kT$ for $k = 0, \ldots, N_y - 1$, T is the sample-time and N_y is the number of data samples.

The likelihood function $p(\mathbf{y}|\boldsymbol{\theta})$ can be difficult to compute, complicating the estimation procedure. However, it is often a simple task to simulate a model of the process described by $p(\mathbf{y}|\boldsymbol{\theta})$. The ABC estimation algorithm exploits this situation by sampling directly from the posterior distribution to

directly obtain an estimate of the conditional distribution $p(\boldsymbol{\theta}|\mathbf{y})$ (Beaumont et al., 2002; Tavare et al., 1997).

The usual ABC procedure is to define some prior distribution for the parameters, $\pi(\boldsymbol{\theta})$, often a uniform distribution in the absence of more specific information, then simulate the model with a sample drawn from the prior. The simulation output is then compared to the observed data using a distance measure - for dynamic systems this is usually some norm of the simulation error. Parameters that generate simulations that are beyond some distance threshold from the observations are rejected. The parameters that are within the distance threshold are accepted and stored. Accepted parameters form a numerical approximation $p(\hat{\boldsymbol{\theta}}|\mathbf{y})$ of the true posterior $p(\boldsymbol{\theta}|\mathbf{y})$. The basic ABC algorithm is:

(1) Draw $\boldsymbol{\theta}^* \sim \pi(\boldsymbol{\theta})$
(2) Simulate $\mathbf{y}^* \sim p(\mathbf{y}|\boldsymbol{\theta}^*)$
(3) Reject $\boldsymbol{\theta}^*$ if $d\left(S(\mathbf{y}^*), S(\mathbf{y})\right) > \epsilon$, else accept $\boldsymbol{\theta}^*$ into $p(\hat{\boldsymbol{\theta}}|\mathbf{y})$

where $S(.)$ describes summary statistics (that may be vector-valued), $d(.)$ is a distance measure between simulated and observed values, and ϵ is a threshold value. For estimation of parameters in dynamic systems the model predictions can be compared directly to observations from the system, without the need for summary statistics (Toni et al., 2009).

Different choices of distance measure $d(.)$ can be used depending on the noise distribution assumed, such as the L_2 norm for normally distributed noise, the L_1 norm for Laplacian noise or the L_∞ norm for uniform noise.

2.3. *Computationally efficient parameter estimation using ABC-SMC*

The algorithm used here for estimating the model parameters was a more computationally efficient extension of the basic ABC algorithm, known as ABC-SMC. A shortcoming of the basic ABC algorithm described above is the low acceptance rate when the prior distribution is very different to the true posterior. A low acceptance rate would require many simulations to adequately represent the posterior. To increase the computational efficiency of ABC, therefore, more sophisticated approaches have been developed (Beaumont, 2010). One such method is the ABC sequential Monte Carlo (ABC-SMC) algorithm (Sisson et al., 2007, 2009; Toni et al., 2009), which has proved effective in dynamic systems modelling (Holmes et al., 2012; Liepe et al., 2012).

The main idea of the ABC-SMC algorithm is to iterate population estimates generated by ABC, gradually decreasing the error tolerance ϵ_k at each iteration k. The posterior distribution at iteration k becomes the sampled prior distribution at $k+1$. Hence, the ABC-SMC algorithm reaches the target posterior in a sequential manner.

The error threshold sequence is chosen so that it decreases at each iteration, hence $\epsilon_1 > \ldots > \epsilon_K$, where K is the number of iterations. The first and final thresholds can be tuned by performing the basic ABC estimation algorithm for L samples and setting $\epsilon_1 = 2d_{min}$ and $\epsilon_K = 1.2d_{min}$, where d_{min} denotes the minimum of the vector of all L distance measures. The ABC-SMC algorithm is described for the nonlinear CT model parameter estimation in Algorithm 1.

The model simulation step was performed by deterministic simulation of the model defined in (1) and (2), using a fourth order Runge-Kutta method. The distance measure of simulations from observations was obtained from the sum-of-squared errors,

$$d = \sum_{j=1}^{N_y} (y(t_j) - y^*(t_j))^2. \tag{8}$$

The L_2 norm used here for $d(.)$ is suited to normally distributed noise but for other types of noise it would be possible to use an alternative, for example an L_1 norm for Laplacian noise or an L_∞

Algorithm 1 Parameter Estimation by ABC-SMC

Require: no. of iterations K, no. of param. samples L,
prior $\pi(\boldsymbol{\theta})$ and error sequence $\epsilon_1 > \ldots > \epsilon_K$
for $k = 1$
  for $j = 1 : L$
    draw $\boldsymbol{\theta}_j^* \sim \pi(\boldsymbol{\theta})$ and simulate $\mathbf{y}_j^* \sim p(\mathbf{y}|\boldsymbol{\theta}_j^*)$
    until $d(S(\mathbf{y}_j^*), S(\mathbf{y})) \leq \epsilon_1$
  end for
  set each weight $w_j^1 = \frac{1}{L}$
end for
for $k = 2 : K$
  for $j = 1 : L$
    sample $\boldsymbol{\theta}_j^*$ from $\boldsymbol{\theta}^{k-1}$ with probabilities w^{k-1}
      perturb $\boldsymbol{\theta}_j^*$ to obtain $\boldsymbol{\theta}_j^{**} \sim \mathcal{K}(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$
      simulate $\mathbf{y}_j^* \sim p(\mathbf{y}|\boldsymbol{\theta}_j^{**})$ until $d(S(\mathbf{y}_j^*), S(\mathbf{y})) \leq \epsilon_k$
  end for
  set each $\boldsymbol{\theta}_j^k = \boldsymbol{\theta}_j^{**}$
  set each $w_j^k = \frac{\pi(\boldsymbol{\theta}_j^k)}{\sum_{i=1}^{L} w_i^{k-1} \mathcal{K}(\boldsymbol{\theta}_i^{k-1}|\boldsymbol{\theta}_j^k)}$, and normalise
end for
Add all final parameter sample estimates $\boldsymbol{\theta}_j^K$, $j = 1, \ldots, L$, to the distribution $p(\hat{\boldsymbol{\theta}}|\mathbf{y})$.
end Algorithm 1

Note: parameter samples are denoted as $\boldsymbol{\theta}^*$, and $\boldsymbol{\theta}^{**}$ after perturbation. $\mathcal{K}$ is a parameter perturbation kernel (uniform random walk).

norm for uniform noise.

The ABC-SMC algorithm (Algorithm 1) has quadratic complexity in the number of samples, L, i.e. the algorithm complexity is $\mathcal{O}(L^2)$ (Toni et al., 2009). In practice, the most time-consuming steps in the ABC algorithm are the model simulations, which are typically performed many thousands of times. The model simulations are inherently parallelisable, due to their independence. Therefore, we exploited the ready availability of multi-core desktop machines to decrease the computation time. Similar performance enhancements could be obtained with graphics processing units (GPUs) (Henriksen, Wills, Schon, & Ninness, 2012; Lee, Yau, Giles, Doucet, & Holmes, 2010), which would require more specialist implementations.

2.4. *Defining the parameter prior for ABC-SMC*

For the nonlinear CT model, the prior distribution of the parameters, $\pi(\boldsymbol{\theta})$, was defined here as a uniform distribution. In the absence of specific information on the prior, it was scaled using the least-squares parameter estimate obtained from the CT-SID method of Coca and Billings (1999). Hence, the prior distribution was defined as,

$$\pi(\boldsymbol{\theta}) \sim U(-2\boldsymbol{\gamma}, 2\boldsymbol{\gamma}), \tag{9}$$

with range parameter $\boldsymbol{\gamma}$,

$$\boldsymbol{\gamma} = (\Psi^\top \Psi)^{-1} \Psi^\top \hat{\mathbf{y}} \tag{10}$$

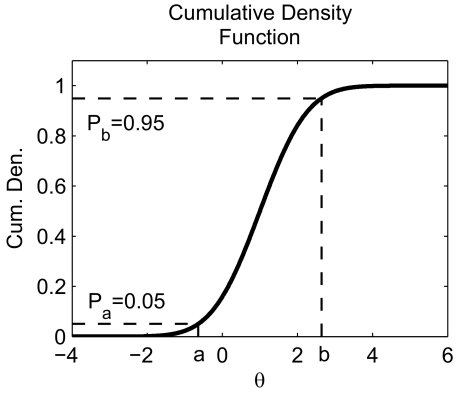

Figure 1. Term selection via the cumulative density function. The cumulative density function for a parameter θ is constructed by the ABC-SMC estimation algorithm. The model term is rejected if zero lies between the limits corresponding to the 5% and 95% probability levels, i.e. $a \leq 0 \leq b$.

where

$$\Psi = \left(\boldsymbol{\psi}(t_1)^\top, \ldots, \boldsymbol{\psi}(t_{N_y})^\top\right)^\top \tag{11}$$

$$\boldsymbol{\psi}(t) = \left(\hat{y}(t), \ldots, \hat{y}^{(n-1)}(t), u(t), \ldots, u^{(n-1)}(t)\right) \tag{12}$$

$$\hat{\mathbf{y}} = \left(\hat{y}^{(n)}(t_1), \ldots, \hat{y}^{(n)}(t_{N_y})\right)^\top \tag{13}$$

where the derivative estimates $\hat{y}^{(j)}(t)$, $j = 0, \ldots, n$, were obtained from a Kalman smoothing algorithm, described in Coca and Billings (1999).

3. Nonlinear continuous-time model identification framework

In this section we develop the identification framework for the nonlinear CT model. First, a simple one-stage approach to structure detection is derived, based on a parameter significance test. The significance test makes use of the parameter distributions intrinsically generated by the ABC-SMC algorithm. This works effectively for terms with sensitive parameters. For terms with less sensitive parameters, we then derive a two-stage algorithm that follows the significance test with an evaluation of simulation performance.

3.1. *One-Stage Model Structure Detection*

The ABC-SMC algorithm naturally generates parameter distributions as part of the estimation procedure. Here, we exploit this feature by developing a structure detection algorithm that makes direct use of these distributions. Similarly to the approach of Kukreja et al. (2004), we use a significance test to prune 'false' parameters from a superset of model terms, where significance is determined from the parameter distributions using a quantile test.

The quantile test selects parameter estimates that cannot be distinguished from zero: the test finds the intervals from the cumulative distribution of the parameters. The algorithm proceeds as follows. We first define an initial superset of candidate model terms, $\mathcal{M}_0$, composed of basis functions, $\phi_j(\mathbf{x}(t))$, which appear in the model decomposition described in (4),

$$\mathcal{M}_0 = \{\phi_1(\mathbf{x}(t)), \ldots, \phi_{N_0}(\mathbf{x}(t))\}, \tag{14}$$

Algorithm 2 One-Stage Model Structure Detection

Require: derivative order n and polynomial order q
Define: superset of model terms $\mathcal{M}_0$, where $N_0 = |\mathcal{M}_0|$
Implement: Algorithm 1 for $\mathcal{M}_0$ (estimate parameters)
Initialise: $\mathcal{M}_1 = \mathcal{M}_0$
for $j = 1 : N_0$ (quantile test for each parameter θ_j)
  if $a_j \leq 0 \leq b_j$ (from (16) and (17))
    discard model term ϕ_j from $\mathcal{M}_1$
  else
    retain model term ϕ_j in $\mathcal{M}_1$
  end if
end for
Implement: Algorithm 1 for $\mathcal{M}_1$ (re-estimate parameters)
end Algorithm 2

of cardinality

$$N_0 = |\mathcal{M}_0|. \tag{15}$$

In the case of polynomial basis functions, the initial superset of candidate terms $\mathcal{M}_0$ would typically be constructed using all polynomial transformations of input-output signal derivatives up to order $n-1$ for an n^{th}-order dynamic model, with polynomial degree q (cubic nonlinearities are often assumed in practice, i.e. $q = 3$).

We then estimate all parameters of the model terms in the set $\mathcal{M}_0$ using Algorithm 1 (ABC-SMC). We select the terms of model $\mathcal{M}_1$ by forming quantile intervals a_j and b_j for each estimated parameter $\hat{\theta}_j$ and pruning the term if zero lies in the interval, i.e. if $a_j \leq 0 \leq b_j$ (see Figure 1), where a_j and b_j are derived from the relationships

$$P_a^{(j)} = \int_{-\infty}^{a_j} p(\hat{\theta}_j|\mathbf{y}) = 0.05, \quad j = 1, \ldots, N_0 \tag{16}$$

$$P_b^{(j)} = \int_{-\infty}^{b_j} p(\hat{\theta}_j|\mathbf{y}) = 0.95, \quad j = 1, \ldots, N_0 \tag{17}$$

where $P_a^{(j)}$ and $P_b^{(j)}$ are the 5% and 95% probability levels respectively for the j^{th} model parameter.

The quantile test is used here unlike the percentile test used by Kukreja et al. (2004) because the posterior distributions obtained in the ABC-SMC framework can be skewed. The one-stage structure detection algorithm is fully described in Algorithm 2.

3.2. *Two-Stage Model Structure Detection*

In this section we describe an enhanced two-stage structure detection algorithm. To motivate this enhancement, we first note that an advantage of the one-stage algorithm based on the quantile test is that it is computationally efficient. The algorithm is efficient because it only requires two implementations of the ABC-SMC algorithm (which has quadratic complexity in the number of samples L): one for structure detection and a second for estimating the parameters of the final model structure. However, a disadvantage of the quantile test is that it does not directly assess the performance of the model simulations. Therefore, in this section we develop a second stage to the algorithm for term selection that directly measures simulation performance using the Bayesian

information criterion (BIC), where

$$\text{BIC} = -2\log\hat{\mathcal{L}} + N_\theta \log N_y \tag{18}$$

where $\hat{\mathcal{L}}$ is the log-likelihood estimate, approximated here from the sum-of-squared prediction error (Ljung, 1999), $\hat{\mathcal{L}} \approx \frac{1}{N_y}\sum_{k=1}^{N_y}\epsilon(t_k)^2$, where $\epsilon(t_k) = y(t_k) - y_M^*(t_k)$ and $y_M^*(t_k)$ is generated by the *maximum a posteriori* (MAP) parameter estimate $\hat{\boldsymbol{\theta}}_M$, i.e. the peak value of the sampled distribution obtained from the ABC-SMC algorithm, $\hat{\boldsymbol{\theta}}_M = \arg\max_{\boldsymbol{\theta}} p(\hat{\boldsymbol{\theta}}|\mathbf{y})$.

The two-stage algorithm proceeds as follows. At stage one we obtain the set of model terms $\mathcal{M}_1$ using the quantile test: the quantile test is only used on terms that pass an initial sensitivity test, where sensitivity is assessed as a parameter variance less than some constant β (here $\beta = 1$) - defined as pool 1 terms, the set $\mathcal{P}_1$. Terms with parameter variance greater than β are defined as pool 2 terms, the set $\mathcal{P}_2$. Terms in pool 2 have an ambiguous contribution to the model and therefore require further testing in the second stage. In the second stage, we obtain the final set of selected model terms $\mathcal{M}_2$ by first ordering pool 2 using the Cha-Srihari metric (defined below) (Cha & Srihari, 2002) and then iteratively testing pool 2 terms using the BIC.

The key step in the two-stage algorithm is the ordering of unselected terms by use of the Cha-Srihari distance metric (Cha & Srihari, 2002). The purpose of using Cha-Srihari is to detect which model parameter distributions have evolved the most from their uniform prior. We assume that the estimated parameter distributions that least resemble their uniform prior contribute the most to describing system dynamics. The ordering of terms makes the search through pool 2 much more efficient than taking the unselected terms at random. The Cha-Srihari distance, $D(A, B)$, measures how much effort it takes to transform a reference histogram, A (the prior), to a target histogram B (the posterior),

$$D(A,B) = \sum_{i=1}^{N_h} |s_i|, \text{ for } i = 1,\ldots,N_h, \tag{19}$$

where $s_i = \sum_{j=1}^{i} r_i$, for $i = 1,\ldots,N_h$; $r_i = A_i - B_i$, A_i and B_i are bar sizes of histograms A and B respectively, and N_h is the number of bars. We set $N_h = 5$ with bar centers at $[-2\boldsymbol{\gamma}, -\boldsymbol{\gamma}, 0, \boldsymbol{\gamma}, 2\boldsymbol{\gamma}]$. We sort the set of pool 2 terms in descending order of Cha-Srihari measure, which are then searched in order using the BIC. The two-stage structure detection algorithm is fully described in Algorithm 3.

3.3. *Derivative Order Model Selection*

Identifying the correct derivative order, n, of the nonlinear CT model is an important issue to address. Here we use the Bayes factor criterion to identify the correct derivative order, as this naturally fits with the ABC framework. ABC can be used in model selection by allocating competing models an index, and then treating this index selection as a parameter estimation problem (Toni et al., 2009). The Bayes factor for comparing evidence supporting two models with different derivative order $\mathcal{M}_i$ and $\mathcal{M}_j$ is

$$B_f = \frac{p(\mathcal{M}_i|\mathbf{y})/p(\mathcal{M}_j|\mathbf{y})}{p(\mathcal{M}_i)/p(\mathcal{M}_j)}, \tag{20}$$

Algorithm 3 Two-Stage Model Structure Detection

Require: derivative order n and polynomial order q
Define: superset of model terms $\mathcal{M}_0$, where $N_0 = |\mathcal{M}_0|$
Implement: Algorithm 1 for $\mathcal{M}_0$
Initialise: $\mathcal{P}_1 = \emptyset$ and $\mathcal{P}_2 = \emptyset$
for $j = 1 : N_0$ (determine $\mathcal{P}_1$ and $\mathcal{P}_2$)
  if variance$(\theta_j) \leq \beta$
    allocate term ϕ_j to $\mathcal{P}_1$
  else
    allocate term ϕ_j to $\mathcal{P}_2$
  end if
end for
Set: $N_1 = |\mathcal{P}_1|$ and $N_2 = |\mathcal{P}_2|$
Initialise: $\mathcal{M}_1 = \mathcal{P}_1$
for $j = 1 : N_1$ (quantile test)
  if $a_j \leq 0 \leq b_j$
    discard model term ϕ_j from $\mathcal{M}_1$
  else
    retain model term ϕ_j in $\mathcal{M}_1$
  end if
end for
Implement: Algorithm 1 for $\mathcal{M}_1$ (re-estimate parameters)

Order terms in $\mathcal{P}_2$ by descending Cha-Srihari metric
Initialise: $\mathcal{M}^{(0)} = \mathcal{M}_1$
for $j = 1 : N_2$ (BIC test for ordered $\mathcal{P}_2$)
  Form model $\mathcal{M}^{(j)}$ by adding term $\mathcal{P}_2(j)$ to $\mathcal{M}^{(j-1)}$
  Implement: Algorithm 1 for $\mathcal{M}^{(j)}$ (re-estimate params)
  if BIC$(\mathcal{M}^{(j)})$ < BIC$(\mathcal{M}^{(j-1)})$
    retain $\mathcal{P}_2(j)$ in $\mathcal{M}^{(j)}$
  else
    break
  end if
end for
Set: $\mathcal{M}_2 = \mathcal{M}^{(j)}$
end Algorithm 3

which for equal uniform priors, $p(\mathcal{M}_i) = p(\mathcal{M}_j)$, simplifies to

$$B_f = \frac{p(\mathcal{M}_i|\mathbf{y})}{p(\mathcal{M}_j|\mathbf{y})} \tag{21}$$

In practice, for derivative order selection, Algorithm 3 is performed independently for models of different derivative order and then models are compared using the Bayes factor in a final iteration of the basic ABC algorithm.

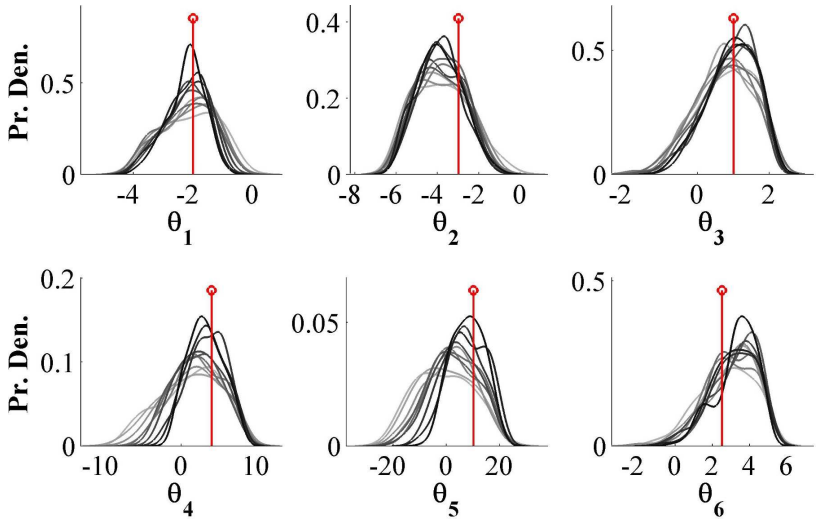

Figure 2. Parameter estimation using ABC-SMC (Algorithm 1). True parameters are shown as red stem plots. Estimated sample distributions are shown over 10 iterations of the ABC-SMC procedure (iteration 1 in grey; iteration 10 in black).

Table 1. Identified models from the numerical example. ABC1 refers to the one-stage ABC identification method, and ABC2 refers to the two-stage ABC identification method.

	SNR = 20 dB			SNR = 10 dB		
True System	dCTM	ABC1	ABC2	dCTM	ABC1	ABC2
$-2.00y(t)$	$-2.00y(t)$	$-2.54y(t)$	$-1.79y(t)$	$-9.25y(t)$	$-2.40y(t)$	$-1.76y(t)$
$-3.00\dot{y}(t)$	$-1.25\dot{y}(t)$	$-4.46\dot{y}(t)$	$-2.46\dot{y}(t)$	–	–	$-2.56\dot{y}(t)$
$1.00u(t)$	$0.77u(t)$	$1.12u(t)$	$1.13u(t)$	$0.65u(t)$	$1.14u(t)$	$1.14u(t)$
$4.00\dot{y}^2(t)$	$3.53\dot{y}^2(t)$	–	$3.88\dot{y}^2(t)$	$4.02\dot{y}^2(t)$	–	$4.11\dot{y}^2(t)$
$10.00y(t)u(t)$	$5.09y(t)u(t)$	–	$9.20y(t)u(t)$	–	–	$9.14y(t)u(t)$
$2.50u^3(t)$	$0.52u^3(t)$	$3.77u^3(t)$	$3.42u^3(t)$	–	$3.07u^3(t)$	$3.35u^3(t)$
–	–	–	–	$26.71y^2(t)$	–	–
–	–	–	–	$-0.01u(t)\dot{u}^2(t)$	–	–

4. Numerical Example

To investigate the performance and accuracy of the proposed ABC identification framework, we applied the one-stage and two-stage algorithms to a test system, with increasing measurement noise. The results were compared in each case to the derivative continuous-time method (dCTM) that uses Kalman smoothing to estimate signal derivatives developed for continuous-time systems by Coca and Billings (1999).

The test system used was

$$\begin{aligned}\ddot{z}(t) &= \theta_1 z(t) + \theta_2 \dot{z}(t) + \theta_3 u(t) + \theta_4 \dot{z}^2(t) \\ &\quad + \theta_5 z(t)u(t) + \theta_6 u^3(t),\end{aligned} \tag{22}$$

$$y(t) = z(t) + e(t), \tag{23}$$

where $e(t)$ was defined as zero-mean Gaussian random noise with variance (i) $\lambda^2 = 4 \times 10^{-5}$ for SNR of 20dB, (ii) $\lambda^2 = 4 \times 10^{-4}$ for SNR of 10dB. The parameter vector was set to $\boldsymbol{\theta} = (-2, -3, 1, 4, 10, 2.5)$. The excitation signal was set to a zero-mean uniform random sequence in the range $(-1, 1)$ band-limited to 20 Hz. For parameter estimation using ABC-SMC the parameter size was set to $L = 200$ and the number of population iterations to $K = 10$. The nonlinear order was set to $q = 3$ and the derivative order to $n = 2$ (except for the derivative selection test described below).

We found that the parameters of the nonlinear system were accurately estimated using the ABC-

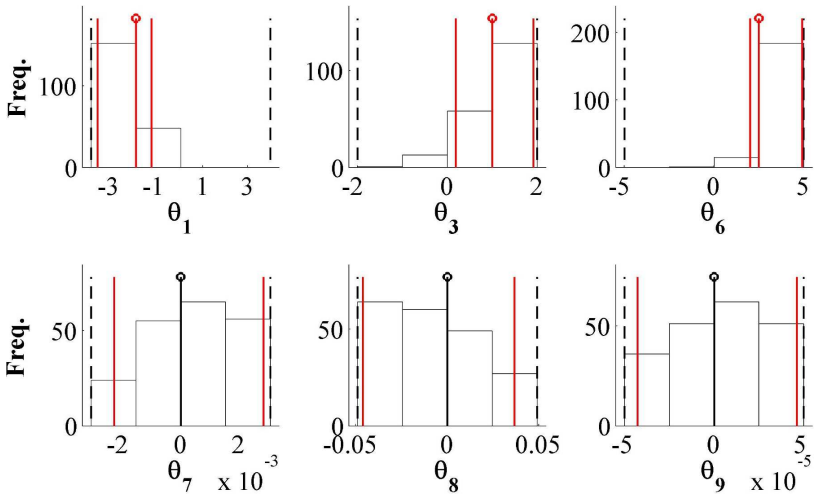

Figure 3. Structure detection using the one-stage procedure (Algorithm 2: $K = 3, L = 200$). Top row: True system terms, correctly selected. Bottom row: False system terms, correctly not selected. The black dotted and red solid vertical lines indicates the prior and quantile values respectively.

SMC procedure defined in Algorithm 1 (Figure 2). The dCTM and one-stage ABC identification algorithms performed well at high SNR (20 dB) with the correct model terms being chosen (illustrative examples are shown in Figure 3) but worsened with increasing noise levels (SNR=10 dB). The two-stage ABC algorithm performed well, however, even at higher noise levels, correctly identifying all terms (Table 1). The much improved performance of the two-stage ABC algorithm is highlighted by a comparison of simulations (Figure 4).

To demonstrate the selection of derivative order, two models of derivative order $n = (2, 3)$, were obtained using the two-stage ABC algorithm and were compared as explained in section 3.3 (using the 10 dB input-output data). From 200 samples, the model with $n = 2$ was selected 187 times and the other model with $n = 3$ was selected 13 times. The Bayes factor in this case, $B_f(1, 2) = \frac{187}{13} = 14.4$, correctly provided strong evidence in favour of $n = 2$ rather than $n = 3$, demonstrating the effectiveness of this model selection approach.

5. Summary

We have developed a computational Bayesian identification framework for nonlinear continuous-time systems. The identification framework makes use of the ABC-SMC algorithm for parameter estimation, which is a rejection sampling technique driven by model simulations (as opposed to one-step-ahead predictions commonly used in system identification). A simple one-stage structure detection algorithm is used to drive term selection by significance testing. A two-stage algorithm has also been developed, which augments the significance test with term selection based on simulation performance, which enhances the robustness of the scheme in the presence of noise.

The main advantages of these algorithms, compared to the current methods available for CT nonlinear system identification, are: (i) parameter distributions are intrinsically generated, giving the user a clear description of uncertainty, (ii) the framework uses a simulation method, which avoids signal derivative estimation and (iii) the simulation approach is more suited to modelling scenarios with non-persistently exciting inputs and fast-sampling. The numerical evaluation of the structure detection algorithms demonstrates the high fidelity of the two-stage algorithm in the presence of increased noise levels in the observed system output.

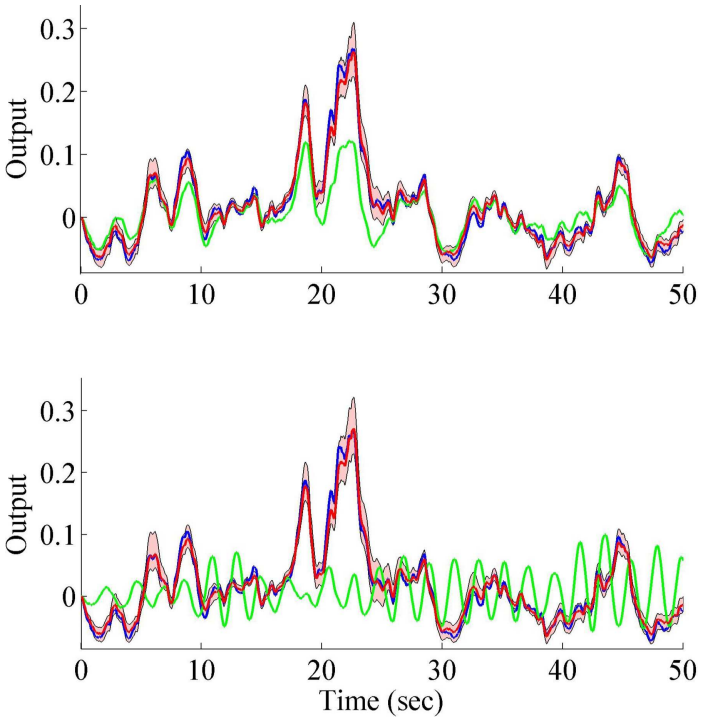

Figure 4. Comparison of noise free output (blue), dCTM model (green) and two-stage ABC (red - the shaded region indicates uncertainty from ABC parameter range). Top: SNR = 20 dB. Bottom: SNR = 10 dB.

Funding

The authors gratefully acknowledge that this work was supported by the Engineering and Physical Sciences Research Council (EPSRC) UK, and a European Research Council Advanced Investigator Award.

References

Anderson, S. R., & Kadirkamanathan, V. (2007). Modelling and identification of non-linear deterministic systems in the delta-domain. *Automatica*, *43*(11), 1859-1868.

Anderson, S. R., Lepora, N. F., Porrill, J., & Dean, P. (2010). Nonlinear dynamic modeling of isometric force production in primate eye muscle. *IEEE Transactions on Biomedical Engineering*, *57*(7), 1554–1567.

Baldacchino, T., Anderson, S. R., & Kadirkamanathan, V. (2012). Structure detection and parameter estimation for NARX models in a unified EM framework. *Automatica*, *48*(5), 857–865.

Baldacchino, T., Anderson, S. R., & Kadirkamanathan, V. (2013). Computational system identification for Bayesian NARMAX modelling. *Automatica*, *49*, 2641-2651.

Beaumont, M. A. (2010). Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, *41*, 379–406.

Beaumont, M. A., Zhang, W., & Balding, D. J. . (2002). Approximate Bayesian computation in population genetics. *Genetics*, *162*, 2025-2035.

Billings, S. A. (2013). *Nonlinear system identification: Narmax, methods in the time, frequency, and spatio-temporal domains*. Wiley.

Cha, S.-H., & Srihari, S. N. (2002). On measuring the distance between histograms. *Pattern Recognition*, *35*(6), 1355–1370.

Chen, S., Billings, S. A., & Luo, W. (1989). Orthogonal least squares methods and their application to non-linear system identification. *International Journal of control*, *50*(5), 1873–1896.

Coca, D., & Billings, S. (1999). A direct approach to identification of nonlinear differential models from discrete data. *Mechanical Systems and Signal Processing*, *13*(5), 739–755.

Falsone, A., Piroddi, L., & Prandini, M. (2015). A randomized algorithm for nonlinear model structure selection. *Automatica*, *60*, 227–238.

Garnier, H., & Wang, L. (2008). *Identification of continuous-time models from sampled data*. London: Springer.

Gevers, M. (2005). Identification for control: From the early achievements to the revival of experiment design. *European Journal of Control*, *11*(4), 335–352.

Guo, Y., Guo, L. Z., Billings, S. A., & Wei, H.-L. (2015). Identification of continuous-time models for nonlinear dynamic systems from discrete data. *International Journal of Systems Science*(ahead-of-print), 1–11.

Henriksen, S. J., Wills, A., Schon, T., & Ninness, B. (2012). Parallel implementation of particle MCMC methods on a GPU. In *16th ifac symposium on system identification, brussels, belgium* (pp. 1143–1148).

Holmes, G. R., Anderson, S. R., Dixon, G., Robertson, A. L., Reyes-Aldasoro, C. C., Billings, S. A., ... Kadirkamanathan, V. (2012). Repelled from the wound, or randomly dispersed? Reverse migration behaviour of neutrophils characterized by dynamic modelling. *Journal of The Royal Society Interface*, *9*(77), 3229–3239.

Krishnanathan, K., Anderson, S. R., Billings, S. A., & Kadirkamanathan, V. (2012). A data-driven framework for identifying nonlinear dynamic models of genetic parts. *ACS Synthetic Biology*, *1*(8), 375–384.

Kukreja, S. L., Galiana, H. L., & Kearney, R. E. (2003). NARMAX representation and identification of ankle dynamics. *IEEE Transactions on Biomedical Engineering*, *50*(1), 70–81.

Kukreja, S. L., Galiana, H. L., & Kearney, R. E. (2004). A bootstrap method for structure detection of NARMAX models. *International Journal of Control*, *77*(2), 132–143.

Lee, A., Yau, C., Giles, M., Doucet, A., & Holmes, C. (2010). On the utility of graphics cards to perform massively parallel simulation of advanced Monte Carlo methods. *Journal of Computational and Graphical Statistics*, *19*, 769-789.

Li, K., Peng, J.-X., & Bai, E.-W. (2006). A two-stage algorithm for identification of nonlinear dynamic systems. *Automatica*, *42*(7), 1189–1197.

Li, L., & Billings, S. (2001). Continuous time non-linear system identification in the frequency domain. *International Journal of Control*, *74*(11), 1052–1061.

Liepe, J., Taylor, H., Barnes, C., Huvet, M., Bugeon, L., Thorne, T., ... Stumpf, M. (2012). Calibrating spatio-temporal models of leukocyte dynamics against in vivo live-imaging data using approximate Bayesian computation. *Integrative Biology*, *4*, 335–345.

Ljung, L. (1999). *System identification - theory for the user* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Ninness, B., & Henriksen, S. (2010). Bayesian system identification via Markov chain Monte Carlo techniques. *Automatica*, *46*(1), 40–51.

Peterka, V. (1981). Bayesian system identification. *Automatica*, *17*(1), 41–53.

Piroddi, L., & Spinelli, W. (2003). An identification algorithm for polynomial narx models based on simulation error minimization. *International Journal of Control*, *76*(17), 1767–1781.

Rao, G. P., & Unbehauen, H. (2006). Identification of continuous-time systems. *IEE Proc. Control, Theory and Applications*, *153*(2), 185–220.

Sisson, S. A., Fan, Y., & Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, *104*(6), 1760–1765.

Sisson, S. A., Fan, Y., & Tanaka, M. M. (2009). Sequential Monte Carlo without likelihoods. erratum 1041760. *Proceedings of the National Academy of Sciences*, *106*, 16889-16889.

Sjoberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P., ... Juditsky, A. (1995). Nonlinear black box modeling in system identification: a unified overview. *Automatica*, *31*(12), 1691–1724.

Tavare, S., Balding, D. J., Griffiths, R. C., & Donnelly, P. (1997). Inferring coalescence times from DNA sequence data. *Genetics*, *145*, 505–518.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., & Stumpf, M. P. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, *6*(31), 187–202.

Tsang, K., & Billings, S. (1994). Identification of continuous time nonlinear systems using delayed state variable filters. *International Journal of Control*, *60*(2), 159–180.

Unbehauen, H., & Rao, G. P. (1990). Continuous-time approaches to system identification - a survey. *Automatica*, *26*(1), 23-35.

Yuz, J., & Goodwin, G. C. (2005). On sampled-data models for nonlinear systems. *IEEE Transactions on Automatic Control*, *50*(10), 1477–1489.