
This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Pesonen, Henri; Simola, Umberto; Köhn-Luque, Alvaro; Vuollekoski, Henri; Lai, Xiaoran; Frigessi, Arnoldo; Kaski, Samuel; Frazier, David T.; Maneesoonthorn, Worapree; Martin, Gael M.; Corander, Jukka

ABC of the future

Published in:

International Statistical Review

DOI:

[10.1111/insr.12522](https://doi.org/10.1111/insr.12522)

Published: 01/08/2023

Document Version

Publisher's PDF, also known as Version of record

Published under the following license:


CC BY-NC

Please cite the original version:

Pesonen, H., Simola, U., Köhn-Luque, A., Vuollekoski, H., Lai, X., Frigessi, A., Kaski, S., Frazier, D. T., Maneesoonthorn, W., Martin, G. M., & Corander, J. (2023). ABC of the future. *International Statistical Review*, 91(2), 243-268. <https://doi.org/10.1111/insr.12522>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

ABC of the future

**Henri Pesonen¹ , Umberto Simola², Alvaro Köhn-Luque¹,
Henri Vuollekoski³, Xiaoran Lai¹, Arnoldo Frigessi^{1,4},
Samuel Kaski^{3,5}, David T. Frazier⁶,
Worapree Maneesoonthorn⁷, Gael M. Martin⁶ and
Jukka Corander^{1,2,8}**

¹*Oslo Centre for Biostatistics and Epidemiology, University of Oslo, Oslo, Norway*

²*Helsinki Institute of Information Technology, Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland*

³*Helsinki Institute of Information Technology, Department of Computer Science, Aalto University, Helsinki, Finland*

⁴*Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway*

⁵*Department of Computer Science, University of Manchester, Manchester, UK*

⁶*Department of Econometrics & Business Statistics, Monash University, Clayton, Victoria, Australia*

⁷*Melbourne Business School, The University of Melbourne, Clayton, Victoria, Australia*

⁸*Parasites and Microbes, Wellcome Sanger Institute, Hinxton, UK*

*Henri Pesonen, Oslo Centre for Biostatistics and Epidemiology, University of Oslo, Oslo, Norway.
Email: henri.e.pesonen@gmail.com*

Summary

Approximate Bayesian computation (ABC) has advanced in two decades from a seminal idea to a practically applicable inference tool for simulator-based statistical models, which are becoming increasingly popular in many research domains. The computational feasibility of ABC for practical applications has been recently boosted by adopting techniques from machine learning to build surrogate models for the approximate likelihood or posterior and by the introduction of a general-purpose software platform with several advanced features, including automated parallelisation. Here we demonstrate the strengths of the advances in ABC by going beyond the typical benchmark examples and considering real applications in astronomy, infectious disease epidemiology, personalised cancer therapy and financial prediction. We anticipate that the emerging success of ABC in producing actual added value and quantitative insights in the real world will continue to inspire a plethora of further applications across different fields of science, social science and technology.

Key words: approximate Bayesian computation; Bayesian inference; likelihood-free inference; simulator-based inference.

1 Introduction

From its humble beginnings over two decades ago, approximate Bayesian computation, or ABC for short, has more recently been met with ever-increasing excitement and is now regarded as one of the most transformative ideas in the statistical sciences. For in-depth reviews of the history of ABC, its theoretical underpinnings and the most recent developments, see (Marin *et al.*, 2012; Sisson *et al.*, 2018; Beaumont, 2019). In particular, deeper theoretical insights into the large-sample behaviour of ABC inference have removed some of the main doubts regarding its statistical validity, and the field is moving rapidly towards a unified understanding of the key aspects that impinge on the asymptotic behaviour of ABC approximations. (Fearnhead & Prangle, 2012; Marin *et al.*, 2014; Green *et al.*, 2015; Lintusaari *et al.*, 2017; Frazier *et al.*, 2018; Li & Fearnhead, 2018a, 2018b; Frazier *et al.*, 2020). Nevertheless, the generic application potential of ABC and other likelihood-free inference (LFI) methods has been held back by the computational requirements of its standard inference algorithms and the lack of a suitable all-purpose software implementation. With the advent of more efficient inference strategies adopted from the field of machine learning (Gutmann & Corander, 2016; Gutmann *et al.*, 2018; Lueckmann *et al.*, 2018; Kokko *et al.*, 2019; Papamakarios *et al.*, 2019; Cranmer *et al.*, 2020; Grazian & Fan, 2020; Thomas *et al.*, 2022) and software platforms such as Engine for likelihood-free inference (ELFI) (Lintusaari *et al.*, 2018), ABCpy (Dutta *et al.*, 2017), BSL (An *et al.*, 2019) and sbi (Tejero-Cantero *et al.*, 2020), to name a few, the immediate prospect of both using and updating the ABC/LFI toolkits for challenging real-world applications certainly looks brighter.

For example, Bayesian optimisation for likelihood-free inference (BOLFI), has been shown in several benchmark examples to speed up ABC inference by 3–4 orders of magnitude (Gutmann & Corander, 2016), and multiple successful applications of it beyond typical benchmarks used in the statistical literature have emerged. These include applications in very diverse research fields, such as inverse reinforcement learning for cognitive user interface models (Kangasrääsiö *et al.*, 2017), brain task interleaving modeling (Gebhardt *et al.*, 2020) and more general computational models of cognition (Kangasrääsiö *et al.*, 2019), perturbation modeling and selection in bacterial populations (Corander *et al.*, 2017), direct dark matter detection (Simola *et al.*, 2019), pathogen outbreak modeling (Lintusaari *et al.*, 2019), sound source localisation (Forbes *et al.*, 2021), passenger flow estimation in airports (Ebert *et al.*, 2021), and the modeling of the genetic components that control the transmissibility of pathogens (Shen *et al.*, 2019). To inspire further methodological development, software engineering and dissemination of ABC and other LFI methods, we present here an array of real applications and discuss both the benefits and challenges that lie ahead for this exciting subfield of statistics.

2 Statistical Methodology

2.1 Preliminaries

We briefly introduce the main concepts and the notation used in the following sections, using the simplest form of ABC algorithm for the purpose of illustration. More detailed description of the methods used in the examples, that is, approximate Bayesian computation–population Monte Carlo (ABC–PMC) (Beaumont *et al.*, 2009) and BOLFI (Gutmann & Corander, 2016) can be found in the following sections.

Bayesian inference is based on calculating the posterior distribution

$$p(\theta | y_{\text{obs}}) \propto p(y_{\text{obs}} | \theta)p(\theta) \quad (1)$$

of the parameters θ given the observed data y_{obs} . The commonly employed methods of conducting inference based on the posterior (e.g. optimisation, importance sampling, Markov chain Monte Carlo (MCMC)) all require pointwise evaluation of the likelihood, $p(y_{\text{obs}}|\theta)$, at any θ . ABC provides an inferential framework for situations where the likelihood function is not available, or whose evaluation is too computationally challenging, by replacing likelihood *evaluation* with *simulation* of the data-generating process, where the latter task is often still feasible even when the former is not.

The rejection ABC algorithm in Algorithm 1 (Tavaré *et al.*, 1997; Pritchard *et al.*, 1999) is the basic formulation of an ABC method. Assuming that the simulator parameter $\theta \in \mathbb{R}^p$ is the target of the statistical inference, the rejection ABC algorithm produces N independent samples from an approximate ABC posterior distribution

$$\pi_{\epsilon}(\theta | y_{\text{obs}}) \propto \int p(y' | \theta) p(\theta) \mathbb{I}_{A_{\epsilon, y_{\text{obs}}}}(y') dy', \quad (2)$$

where y' is artificial data simulated from the generative model, $\mathbb{I}_{A_{\epsilon, y_{\text{obs}}}}(\cdot)$ is the indicator function for the set $A_{\epsilon, y_{\text{obs}}} = \{y' | d(y_{\text{obs}}, y') \leq \epsilon\}$, which is defined via a distance metric $d(\cdot, \cdot)$ and a threshold parameter ϵ .

The ABC posterior as defined by (2) is not conditioning on the data exactly, but on a set of artificial data following the distribution of the generative model that is within a tolerance ϵ from the observed data, as determined by the difference $d(y_{\text{obs}}, y')$ between observed and artificial data. Because the relative volume of $A_{\epsilon, y_{\text{obs}}}$ becomes vanishingly small when the dimension of the data y_{obs} increases, sampling-based algorithms such as rejection ABC reduce the sample space in order to perform adequately. Traditionally, this is performed by defining a set of summary statistics $s(\cdot)$. Ideally, the summarising function would be a sufficient statistic, but this is rarely available in problems with intractable likelihoods. In practice, $s(\cdot)$ are chosen according to a number of different principles, aimed to maximise the informativeness of the summaries in some sense (Joyce & Marjoram, 2008; Blum, 2010; Drovandi *et al.*, 2011; Fearnhead & Prangle, 2012; Drovandi *et al.*, 2015; Martin *et al.*, 2019). The core of Algorithm 1 is simple and straightforward to implement in most programming languages.

Algorithm 1 Rejection ABC algorithm for θ

```

for  $i = 1, \dots, N$  do
   $d' = \infty$ 
  while  $d' > \epsilon$  do
    Sample from the prior,  $\theta' \sim p(\theta)$ 
    Simulate from the generative model,  $y' \sim p(y | \theta')$ 
    Calculate distance,  $d' = d(s(y_{\text{obs}}), s(y'))$ 
  end while
  Set  $\theta^{(i)} = \theta'$ 
end for

```

2.2 ABC-PMC Algorithm

Although the rejection ABC algorithm is still being used frequently as a comparison method in the ABC and LFI literature, there are few applications where it would not be beneficial to instead take advantage of more sophisticated versions of this basic algorithm

(Beaumont *et al.*, 2009; Blum, 2010; Csilléry *et al.*, 2010; Drovandi & Pettitt, 2011; Marin *et al.*, 2012; Moral *et al.*, 2012; Rodrigues *et al.*, 2020; Simola *et al.*, 2020; Clarté *et al.*, 2021). The ABC–PMC approach by Beaumont *et al.* (2009) is an extension of the rejection ABC algorithm based on importance sampling, and aims to improve the efficiency of the procedure by retrieving a sequence of intermediate distributions. The steps of the method are summarised in Algorithm 2. The first iteration of the ABC–PMC algorithm corresponds to the four steps of the basic rejection ABC algorithm with ϵ_1 . Starting from the second iteration $t \geq 2$, a particle is sampled with replacement from the set of importance weighted particles at iteration $t - 1$, it is moved using, for example, a Gaussian kernel, and accepted or rejected based on $\epsilon_t < \epsilon_{t-1}$. The importance weight for particle $J = 1, \dots, N$ at iteration t is

$$W_t^{(J)} \propto p(\theta_t^{(J)}) / \sum_{K=1}^N W_{t-1}^{(K)} \phi \left[\tau_{t-1}^{-1/2} \left(\theta_t^{(J)} - \theta_{t-1}^{(K)} \right) \right],$$

where $\phi(\cdot)$ is a multivariate Gaussian kernel with mean 0 and identity covariance, and τ_{t-1} is twice the (weighted) sample covariance of the particles from iteration $t - 1$, as recommended in Beaumont *et al.* (2009).

Algorithm 2 ABC–PMC algorithm for θ

```

Set  $\epsilon_1 > \dots > \epsilon_T$ 
if  $t = 1$  then
  for  $J = 1, \dots, N$  do
    while  $d_1^{(J)} > \epsilon_1$  do
      Propose  $\theta^{(J)} \sim p(\theta)$ 
      Generate  $y' \sim p(y | \theta^{(J)})$ 
      Calculate  $d_1^{(J)} = d(s(y_{\text{obs}}), s(y'))$ 
    end while
    Set weight  $W_1^{(J)} = N^{-1}$ 
  end for
else if  $2 \leq t \leq T$  then
  Set  $\tau_t = 2 \cdot \text{cov} \left( \{\theta_{t-1}^{(J)}, W_{t-1}^{(J)}\}_{J=1}^N \right)$ 
  for  $J = 1, \dots, N$  do
    while  $d_t^{(J)} > \epsilon_t$  do
      Select  $\theta_t^*$  from  $\theta_{t-1}^{(J)}$  with probabilities  $\left\{ W_{t-1}^{(J)} / \sum_{K=1}^N W_{t-1}^{(K)} \right\}_{J=1}^N$ 
      Propose  $\theta_t^{(J)} \sim N(\theta_t^*, \tau_t)$ 
      Generate  $y' \sim p(y | \theta_t^{(J)})$ 
      Calculate  $d_t^{(J)} = d(s(y_{\text{obs}}), s(y'))$ 
    end while
    Set weight  $W_t^{(J)} \propto p(\theta_t^{(J)}) / \sum_{K=1}^N W_{t-1}^{(K)} \phi \left[ \tau_{t-1}^{-1/2} \left( \theta_t^{(J)} - \theta_{t-1}^{(K)} \right) \right]$ 
  end for
end if
```

As a final remark on the ABC–PMC algorithm, both the total number of iterations T and the series of decreasing tolerances $\epsilon_1 > \epsilon_2 > \dots > \epsilon_T$ must be selected in advance by the

researcher, which might have an impact on the computational performance of the procedure, as well as on the attainability of a suitably accurate approximation of the exact posterior distribution (Silk *et al.*, 2013; Simola *et al.*, 2020). For this reason, rather than defining in advance the series of decreasing tolerances, other choices are possible. In particular, adaptively selecting the tolerance for some iteration t based on a previously defined quantile of the distances of the accepted particles from the previous iteration $t - 1$ improves the efficiency of the algorithm in terms of how many times the forward model is used (Lenormand *et al.*, 2013; Weyant *et al.*, 2013; Ishida *et al.*, 2015; Cisewski-Kehe *et al.*, 2019; Simola *et al.*, 2020). For these reasons also a tuning step is often necessary in order to balance the trade-off between computational efficiency and the reliability of the inferential results.

2.3 BOLFI Algorithm

Although potentially orders of magnitude more efficient than the basic rejection ABC algorithm, ABC-PMC is based on the idea of identifying relevant regions of the parameter space by finding proposal distributions such that the corresponding simulated datasets are similar within a tolerance to the observed dataset, according to pre-defined summary statistics. As often only weak information is available in advance about what regions are relevant, this requires a large number of datasets to be simulated through the forward model. Lintusaari *et al.* (2018) notes, for example, that the number of simulated datasets required in the implementation of a sequential ABC sampler like Algorithm 2 is usually at least of order 10^6 . If model simulation itself is heavy, the total computational cost of the sequential algorithm becomes very high.

To reduce the need for a large number of potentially costly model simulations, active learning methods adapt the querying process according to different strategies. BOLFI uses Bayesian optimisation (Gutmann & Corander, 2016) to construct iteratively a probabilistic surrogate model for the relationship between parameters θ' and discrepancies $d(\theta') = d(s(y'), s(y_{\text{obs}}))$ using the growing evidence set \mathcal{E}_t , which consists of pairs $\{(\theta_i, d(\theta_i))\}_{i=1}^t$. The original formulation of BOLFI uses a Gaussian process (GP) as the surrogate model for the discrepancy function, and new evidence $\{(\theta_{t+1}, d(\theta_{t+1}))\}$ is sampled from relevant regions of the parameter space. Relevant regions are determined to be parts of the space where the discrepancy is small. The probabilistic model is defined as $d(\theta) | \mathcal{E}_t \sim GP(\mu_t(\theta), v_t(\theta) + \sigma^2)$, where GP is a Gaussian process with mean and variance functions,

$$\mu_t(\theta) = k_t(\theta)^T K_t^{-1} [d(\theta_1), \dots, d(\theta_t)]^T \quad (3)$$

$$v_t(\theta) + \sigma^2 = k(\theta, \theta) - k_t(\theta)^T K_t^{-1} k_t(\theta) + \sigma^2. \quad (4)$$

The vector $k_t(\theta) = [k(\theta, \theta_1), \dots, k(\theta, \theta_t)]^T$ and matrix $K_t = [k(\theta_i, \theta_j)] \in \mathbb{R}^{t \times t}$ are defined via covariance functions $k(\theta', \theta'')$.

A common choice for a covariance function is the squared exponential covariance function

$$k(\theta', \theta'') = \sigma_f^2 \exp \left(- \sum_{j=1}^d \frac{1}{\lambda_j^2} (\theta'_j - \theta''_j)^2 \right), \quad (5)$$

where parameters θ'_j and θ''_j are the j th elements of θ' and θ'' . When fitting a GP to the evidence set, the hyperparameters σ , σ_f , and λ_j can be optimised iteratively (Rasmussen & Williams, 2006). In practice, it is not necessary to update the hyperparameters given each additional evidence point; instead, a more efficient strategy is to update them with every T_{update} additional points, for some specified value for T_{update} .

Finally, from $d(\theta)$ we can retrieve a suitable pointwise approximation to the likelihood function

$$L(\theta) \approx \Phi\left(\frac{h - \mu_t(\theta)}{\sqrt{v_t(\theta) + \sigma^2}}\right), \quad (6)$$

where $\Phi(\cdot)$ is the Gaussian cumulative density, and h is a threshold parameter which we choose to be the minimum of the mean discrepancy $\mu_t(\theta)$, although other choices are reasonable. The discrepancies in the evidence set can be log-transformed for possibly improving the GP-fit as advised in Gutmann & Corander (2016). The rest of the steps of the algorithm remain the same.

In practice, an acquisition function is used to determine the locations of relevant parameter space, and there are several reasonable choices for it. Some are directly based on optimising the density fitting (Järvenpää *et al.*, 2019) while some are efficient for finding the optimum of the function, such as the lower confidence bound selection criterion (LCBSC) (Brochu *et al.*, 2010; Srinivas *et al.*, 2010). LCBSC is defined as

$$\mathcal{A}_t(\theta) = \mu_t(\theta) - \sqrt{\eta_t^2 v_t(\theta)}, \quad (7)$$

where $\eta_t^2 = 2\log\left(t^{\frac{d}{2}} + 2\pi^2/3\epsilon_\eta\right)$ is a coefficient depending on the iteration t , the dimension of the parameter space d and the tunable parameter ϵ_η . The parameter value is obtained by minimising (7) and varying it randomly to further balance exploration and exploitation of the probabilistic function $d(\theta)$. Our approach is to query the simulator at θ_t , where

$$\theta_t \sim TN(\hat{\theta}_t, \sigma_{\text{acq}}^2, a_L, a_U), \quad (8)$$

where $TN(\hat{\theta}_t, \sigma_{\text{acq}}^2, a_L, a_U)$ is a normal distribution truncated to interval $[a_L, a_U]$ with mean $\hat{\theta}_t = \text{argmin}_\theta \mathcal{A}_t(\theta)$ and tunable variance σ_{acq}^2 . Variance σ_{acq}^2 control the variation from the LCBSC minimum. The interval $[a_L, a_U]$ is the optimisation region for the parameter.

BOLFI can be thought as an extension of either an ABC- or a synthetic likelihood (SL)-type method (Wood, 2010). As an SL-type method, the likelihood surrogate and the prior can be used as a target for an MCMC or a Hamiltonian Monte Carlo (HMC) sampling algorithm to draw an approximate posterior sample of size N_{sample} . A popular HMC sampling algorithm is the no-u-turn sampler (NUTS) (Hoffman & Gelman, 2014) which avoids certain sensitivity issues regarding user-defined parameters which the algorithm sets automatically. In this study we use the BOLFI implementation available in ELFI, and the specific version of it is presented in Algorithm 3.

Algorithm 3 BOLFI

```

for  $t = 1, \dots, N_{\text{init}}$  do
    Generate  $\theta_t \sim p(\theta)$ ,
    Generate  $y' \sim p(y \mid \theta_t)$ 
    Calculate  $d_t = d(s(y_{\text{obs}}), s(y'))$ 
end for
Set  $\mathcal{E}_{N_{\text{init}}} = \{(\theta_t, d_t)\}_{t=1}^{N_{\text{init}}}$ 
Fit  $d(\theta) \mid \mathcal{E}_{N_{\text{init}}} \sim GP(\mu_{N_{\text{init}}}(\theta), v_{N_{\text{init}}} + \sigma^2)$ 
Set  $k = 0$ 
for  $t = N_{\text{init}} + 1, \dots, N_{\mathcal{E}}$  do
    if  $t \equiv T_{\text{update}} \pmod{T_{\text{update}}}$  then
        Optimize GP-hyperparameters
    end if
    Calculate  $\eta_t^2 = 2 \log \left( t^{\frac{d}{2}+2} \pi^2 / 3 \epsilon_{\eta} \right)$ 
    Calculate  $\hat{\theta}_t = \arg \min_{\theta} \mu_t(\theta) - \sqrt{\eta_t^2 v_t(\theta)}$ 
    Generate  $\theta_t \sim TN(\hat{\theta}_t, \sigma_{\text{acq}}^2, a_L, a_U)$ 
    Generate  $y' \sim p(y \mid \theta_t)$ 
    Calculate  $d_t = d(s(y_{\text{obs}}), s(y'))$ 
    Set  $\mathcal{E}_t = \mathcal{E}_{t-1} \cup \{(\theta_t, d_t)\}$ 
    Fit  $d(\theta) \mid \mathcal{E}_t \sim GP(\mu_t(\theta), v_t + \sigma^2)$  based on  $\mathcal{E}_t$ 
    Set  $k = k + 1$ 
end for
Define  $L(\theta) \approx \Phi \left( \frac{h - \mu_t(\theta)}{\sqrt{v_t(\theta) + \sigma^2}} \right)$ 
Generate  $N_{\text{sample}}$  draws from approximate posterior  $\propto L(\theta)p(\theta)$ 

```

2.4 Recent Progress in Method Development

The features of rejection ABC, that is, the distance function, the distance threshold and the summarising statistics are still common in many modern ABC algorithms, even if newer methods have achieved superior performance relative to the original formulation of the method.

The basic rejection ABC uses samples generated from the prior, which can be extremely ineffective depending on the informativeness of the prior. ABC-PMC and ABC-sequential Monte Carlo (ABC-SMC) methods (Toni *et al.*, 2009) improve on rejection ABC by sequentially constructing better proposal distributions. Sequential ABC methods have further been improved upon, for example, by introducing adaptivity to the threshold selection (Simola *et al.*, 2020) and the distance function (Prangle, 2017). Other approaches have been introduced to help with summary statistic design and selection (Fearnhead & Prangle, 2012; Thomas *et al.*, 2022). MCMC procedures can also be used to generate samples from the approximate posterior distributions (Marjoram *et al.*, 2003; Marjoram, 2013; Vihola & Franks, 2020), and improving their applicability to high-dimensional problems is an on-going research problem (Rodrigues *et al.*, 2020; Clarté *et al.*, 2021).

Parametric approaches such as SL trade the requirement for a distance function and a threshold for a pointwise approximation of the likelihood by a parametric distribution such as a normal distribution. The moments of the approximating likelihood are, in turn, estimated using simulator draws that are possibly summarised (Wood, 2010; Price *et al.*, 2018). The parametric approximation of the posterior enables the use of a variety of methods to draw a representative

sample of the posterior (An *et al.*, 2019). SL methods have also been extended to high-dimensional parameter spaces (Ong *et al.*, 2018).

Recently, density estimation techniques based on deep neural network architectures have been developed for likelihood-free inference (Grazian & Fan, 2020; Lueckmann *et al.*, 2018; Papamakarios *et al.*, 2019). These approaches are similar to the SL-type methods where the likelihood is approximated with a parametric surrogate, but in the place of the parametric distribution, neural network architectures such as autoregressive flows or emulator networks are used to fit either a surrogate model for the likelihood (local approach) or for the simulator (global approach). Furthermore, these methods have also been combined with active learning approaches to minimise the required number of simulator queries. Lueckmann *et al.* (2018) proposed to minimise the variance of the posterior surrogate as suggested by Järvenpää *et al.* (2019) in local inference and to maximise information gain in global inference (Houlsby *et al.*, 2011; Gal *et al.*, 2017; Depeweg *et al.*, 2017).

3 ABC in Infectious Disease Epidemiology with Application to EBOLA Outbreaks

Simulator-based inference is well suited to infectious disease epidemiology. For example, it has been used to resolve the outbreak dynamics of stochastic birth-death-mutation models (Lintusaari *et al.*, 2019), and to infer the transmission dynamics of the Ebola haemorrhagic fever outbreak in 1995 in the Democratic Republic of Congo (McKinley *et al.*, 2009) and the COVID-19 pandemic (Chinazzi *et al.*, 2020). In this case study, we demonstrate the application of simulator models to gain insight into the epidemic caused by the emergence of the Ebola virus in West Africa in 2014 as reported by the WHO Ebola Response Team (Team WER) (WHO Ebola Response Team, 2014) to infer the basic reproduction number R_0 , that is, the mean value of secondary infections caused by an infectee when no control measures are in place. The computational complexity of the recent, individual-based simulator models can be substantial and the pace at which inference can be delivered is of particular importance to determine the severity of the situation, and the predicted progress under different hypotheses. Therefore, active learning-based methods such as BOLFI in Algorithm 3 provide a strategy to minimise the number of simulator queries and to quickly approximate the posterior for R_0 .

3.1 Team WER Model

WHO Ebola Response Team (2014) estimated the basic reproduction number by modeling the incidence of onset $I(\cdot)$ at time t with a Poisson process

$$I(t) \sim \text{Poisson} \left(R_0 \sum_{s=1}^{T_{\text{exp}}} \omega I(t-s) \right), \quad t = 1, \dots, T_{\text{exp}} \quad (9)$$

which is defined via the basic reproduction number R_0 , the serial interval parameter ω , which is the time difference between the onset of symptoms of the infector and the infectee, and the incidence history $I(0), \dots, I(t-1)$. The serial interval parameter ω was determined heuristically from the contact tracing data. It was approximated as gamma-distributed with mean 15 and coefficient of variation 0.66. Finally, the time interval $[0, T_{\text{exp}}]$ is the initial period of exponential growth of infections.

In this case, the likelihood for the observed time series can be written as

$$p(I(1), \dots, I(T_{\text{exp}}) | R_0, \omega) = \prod_{t=1}^{T_{\text{exp}}} P(I(t) = k | R_0, I(1), \dots, I(t-1), \omega), \quad (10)$$

and the posterior distribution can be solved analytically when R_0 is modeled as gamma-distributed *a priori*. However, Britton & Tomba (2019) argued that even though this population level model has good predictive power, it would tend to lead to underestimation of R_0 . The main causes of the underestimation in the model are the use of the serial intervals instead of generation times which indicate the time between infections of the infectee and infector, and the handling of the contact tracing when there are multiple possible infectors. Both of these aspects, plus other complex factors in the observation process, are simpler to take into account when using a forward simulator-based modeling approach, such as the one described in the next subsection.

3.2 Description of the Simulator

We simulate an outbreak of a virus in a homogeneous and infinite community. The general simulation model follows closely the description in Britton & Tomba (2019). The main difference from the model used by WHO Ebola Response Team (2014), is that the forward simulation program generates the state of each infected individual as described by the diagram in Figure 1 instead of via a population level Poisson-process model. Each infected individual will be in one of four possible states: *latent*, *infectious*, *recovered* or *perished*. After the initial infection, an individual enters the latent period t_{lat} of the infection. It is modeled as $t_{\text{lat}} \sim \Gamma(2, 5)$, where $\Gamma(\alpha, \beta)$ is a gamma-distribution with shape α and scale β . The latent period is followed by the infectious period $t_{\text{inf}} \sim \Gamma(1, 5)$ in which the transfer of the infection to other individuals is possible. The probability of a new infection after s time units because the initial infection follows the infection rate $R_0 f_G(s)$ defined by R_0 and the generation time distribution $f_G(\cdot)$. The individual survives the infectious period with probability $p_{\text{reco}} = 0.3$ after a recovery period of $t_{\text{reco}} \sim \Gamma(4, 3)$, or perishes with probability $(1 - p_{\text{reco}})$ after a period of $t_{\text{die}} \sim \Gamma(4/9, 9)$.

For simplicity, we assume that there is no delay in reporting the infection once symptoms arise, and that all of the cases are reported. Assumptions concerning possible reporting bias could be built into the model in a similar fashion as for the other quantities. The time period between infection and symptoms is defined as the incubation period which is the latent period multiplied by an incubation factor $\gamma \sim \text{Uni}(0.8, 1.2)$.

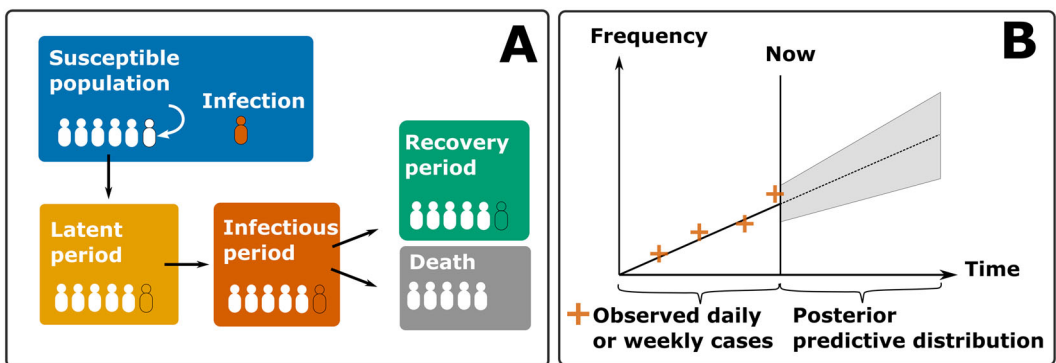


FIGURE 1. Panel A: In the generative model, each infected subject is in one of the four stages. After the initial infection, the subject is in latent stage and becomes infectious after that. The infectious period ends either in recovery or in death. Panel B: The generative model can be used with the approximate posterior distribution of the basic reproduction number to probabilistically forecast how the infection will spread in the following weeks

The simulation always starts with a single infected individual and iterates in steps of $\Delta t = 0.2$ days until either 104 weeks pass or 100,000 individuals have been infected. During each time step the status of each infected individual is checked, and if the current status is *infectious*, a new individual is infected with probability $p_{\text{inf}} = \Delta t / \Delta T$, where $\Delta T = \hat{t}_{\text{inf}} / R_0$ is the mean time between infections and \hat{t}_{inf} is the mean duration of infectivity.

Notice that in principle it could be possible to construct a likelihood function in this example. However, as the state of each of the infected individual is modeled separately and the number of infected individuals grows exponentially, the resulting combinatorial complexity very quickly prohibits the construction of the likelihood function outside of very small scale simulations.

3.3 Inference with BOLFI

The observed data consisted of the daily cumulative count of confirmed cases from the begining of the 2014 epidemic for which we infer R_0 . We used the same initial time periods $[0, T_{\text{exp}}]$ for estimating R_0 as used in WHO Ebola Response Team (2014). The inference was performed separately for Guinea and Liberia using data from 22 March to 30 March 2014 and from 16 June to 20 August 2014, respectively. Subsequent data from the following days were used to illustrate the progression of the virus. In the original study, the authors determined that the initial growth period would be over during the later time period.

The fact that we do not know the true onset of the virus is taken into account by generating artificial counts until we obtain one that exceeds the first observed count, and then continuing the simulation for the same number of days that is in the observed dataset. This way the artificial datasets will have a similar level of variability to the observed one.

The prior for R_0 is modeled as a truncated normal distribution $TN(1.7, 0.5, 1.05, 4)$. As summary statistic we used the median of slopes of log-transformed case counts that were calculated with consecutive datapoints. As the discrepancy measure $d(\cdot, \cdot)$ we used the logarithm of the euclidean distance. The BOLFI parameters are listed in Table 1.

To verify R_0 inference performance, a sample of size 500 was drawn from the approximate posterior distribution obtained by BOLFI using NUTS sampler. The approximate posterior sample was propagated using the forward simulator, which allows for the investigation of the prediction intervals as a function of time. The propagation of the sample for the following days was visually compared with the actual observed cumulative case count.

3.4 Results Using Empirical Data

The inference results for Guinea and Liberia are illustrated in Figure 2. For Guinea and Liberia, the inferred estimates and 95% credible intervals (CIs) of the basic reproduction number are quite consistent with those reported by WHO Ebola Response Team (2014).

Table 1. BOLFI parameters for the inference of the epidemiological model

BOLFI-parameter	value
N_{init}	5
$N_{\mathcal{E}}$	100
T_{update}	5
σ^2_{acq}	0.1
N_{sample}	2000

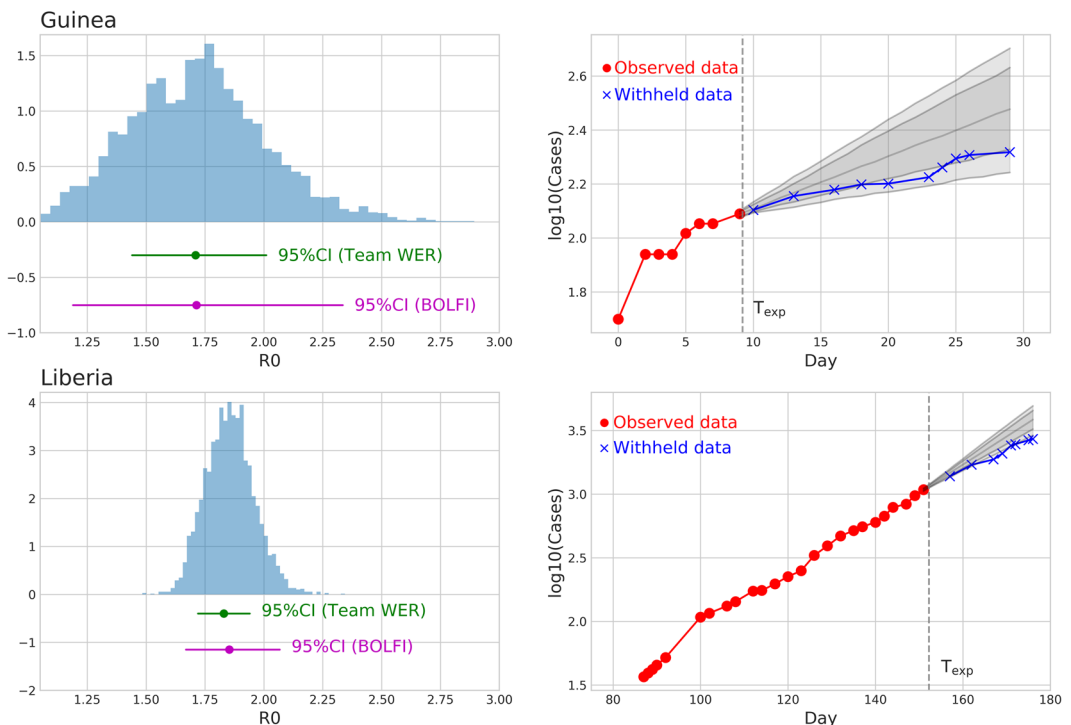


FIGURE 2. Left: Approximate posterior distributions for R_0 using data from the initial unconstrained growth period. Results indicate that the subject-level modeling increases the uncertainty about the R_0 value relative to results reported in WHO Ebola Response Team (2014). Right: The forecast of the number of cumulative infections along with observed data that were used in inference and data withheld from the inference. Ribbons on the plot indicate pointwise 80% and 95% probability intervals of the forecast. Comparison of forecast and withheld data indicates that period of initial growth T_{exp} as reported in WHO Ebola Response Team (2014) is reasonable

For Guinea, they reported an estimate 1.71 with 95% CI of 1.44–2.01, whereas the BOLFI posterior had a mean of 1.72 and a 95% CI of 1.19–2.33. For Liberia, Team WER reported an estimate 1.83 with 95% CI, 1.72–1.94 and BOLFI provided an approximate posterior with mean 1.87 and 95% CI 1.49–2.18. Reference Althaus (2014) reported maximum likelihood estimates 1.51 (95%-confidence interval: 1.50 – 1.52) and 1.59 (95%-confidence interval: 1.57 – 1.60), for Guinea and Liberia, respectively. The wider uncertainty about R_0 in our case study CIs reflects the more complex underlying modeling assumptions, and taking such uncertainty appropriately into account may be critical in epidemic management situations.

Forward propagated posterior prediction results consisting of the pointwise median of the simulated trajectories and the 80%- and 95%-probability bands are also reported in Figure 2 along with the observations used in the inference and some out-of-sample observations that were not used in inference, but are used to illustrate forecasting accuracy. WHO Ebola Response Team (2014) assumed, on a basis of a visual inspection, that the initial period of exponential growth would be 30 March for Guinea and 24 August for Liberia which are illustrated with the gray vertical lines on Figure 2. Dates are reported as counts because the beginning of the epidemic (March 22). Our model fit is consistent with the initial growth period assumption as the observed data counts after the assumed initial growth period diverge quickly from the probable prediction curves.

4 ABC In Personalised Cancer Treatment With Application To Breast Tumor Modeling

Personalised cancer treatment is an application area where simulator-based inference has a lot of potential, given the constantly improving biological generative models for the evolution of the disease under treatment (Sottoriva & Tavaré, 2010; Kozłowska *et al.*, 2018; Lai *et al.*, 2019; Lai *et al.*, 2021). Realistic biological generative tumor models that are built up from the molecular and cellular processes result typically in a level of complexity that renders analytical solutions infeasible. Using simulator-based methods we are able to conduct inference on such detailed models, which could in the future be used to optimise personalised treatments to achieve the best possible therapy outcomes. However, the complexity of the biological modeling often makes the generative modeling computationally demanding and only active learning-based methods such as BOLFI (Algorithm 3) enable the inference to be performed within a reasonable time.

4.1 Description of the Simulator

We consider an example of breast cancer tumor modeling that is a combination of empirical data and detailed computer simulation. The system is partly initialised and tuned based on a biopsy from a real patient, but we use only simulated data to demonstrate the inference and prediction procedure.

The cancer simulator we use is a multi-scale pharmacokinetic and pharmacodynamic model describing the response of a cross-section of breast tumor tissue to a combination of chemotherapeutic and anti-angiogenic agents(Lai *et al.*, 2019). Mathematically, it consists of a hybrid cellular automaton model (Ribba *et al.*, 2004; Alarcón *et al.*, 2010) that couples stochastic and discrete model formalisms with deterministic and continuous components accounting for biological processes at different spatio-temporal scales; see Figure 3. There are multiple

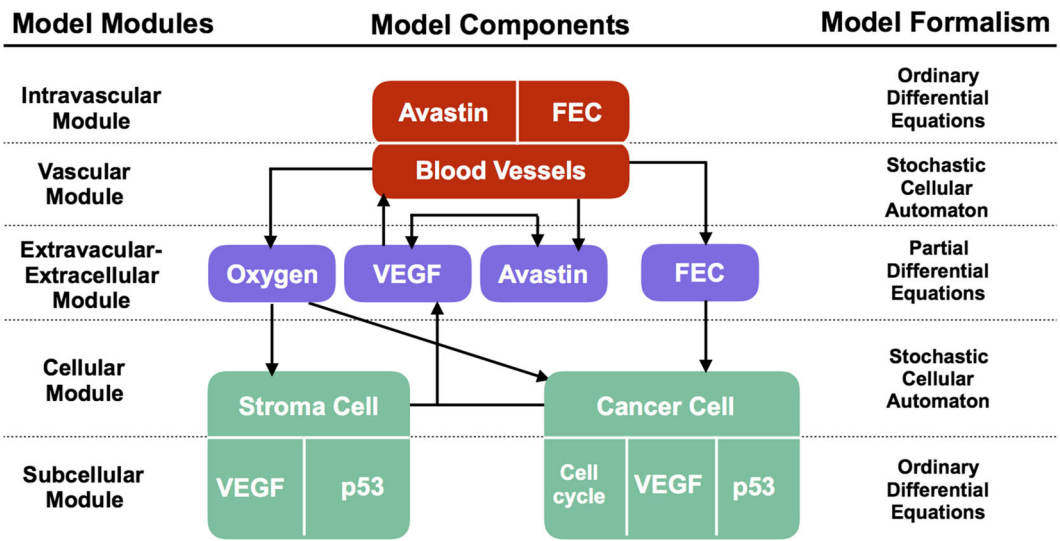


FIGURE 3. Modular structure of the hybrid cellular automaton describing the response of breast tumor tissue to a cocktail of chemotherapies (FEC) and Avastin. The diagram shows the main components of each module and the interactions among them. The right column shows the different model formalisms used for each of the model modules. See (Lai *et al.*, 2019) for full model details and its patient-specific parameterisation

17518823, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/insr.12522 by Aalto University, Wiley Online Library on [09/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

parameters in the model influencing the evolution of cancer cells, stromal cells and tumour vessels, which are all modeled as individual agents. Many of the parameters can be inferred and fixed via various means but some of the unknown parameters have to be inferred based on the reaction of the tumour to the treatment. Here we focus on the inference of two key parameters determining the outcome of each patient to the drug treatment. Those parameters account for the *sensitivity of cancer cells to the chemotherapy*, α , and the *minimal cell cycle length of cancer cells*, T_c . We can write the simulator as a nonlinear time series model for the evolution of state x_t

$$x_t = f_t(\alpha, T_c, x_{t-1}, v_t), \quad t = 1, 2, \dots, T, \quad (11)$$

where f_t is the nonlinear transition model at t , which is the time for the temporally discretised simulator with 30 minute increments, and v_t is the stochastic component of the simulator. The simulated state x_t consists of cells, vessels, and extracellular concentrations of oxygen, Avastin and vascular endothelial growth factor (VEGF) within the simulation grid, which here is a 33×20 rectangular grid representing a specific two-dimensional cross-section of the tumor.

The time series model is not available for constant monitoring but some of the components can be indirectly measured at sparse time points using methods such as magnetic resonance imaging. For simplicity we assume that it is possible to measure the true state of the system

$$y_{t_k} = x_{t_k}, \quad k = 1, 2, \dots, \quad (12)$$

In this proof-of-concept study we assume that observations can be collected every three days $t_k = k \cdot 3 \cdot 48$, $k = 1, 2, \dots, 6$.

The drug is administered for the patient every 3 weeks, that is, at times $t_k^{\text{drug}} = k \cdot 21 \cdot 48$, $k = 0, 1, 2, 3$ and we aim to infer the parameters within the first 3-week treatment period, as we would want to forecast whether or not the chosen treatment is effective before the second dose. This dose could be adjusted, given the inferred parameters, to achieve an optimal outcome. Being able to reliably infer the parameters using as little data as possible from the beginning of the cancer evolution curve would enable simulated testing of personalised treatment strategies that are anticipated to be efficient for the specific patient.

4.2 Inference with BOLFI

We generate the fake observed data using fixed parameter values $T_c = 14.69$ and $\alpha = 3.0$, and predetermined treatment protocol, and investigate how well we can infer the set parameters and how accurately we are able to forecast the disease progress based on the data collected within the first treatment period. If we were able to infer the patient-specific parameter values based on the observed data, then we would be able to forecast how the patient defined by the parameter values would react also to *different* treatment protocols within the accuracy of the simulator. In our inference framework the results are approximations to the posterior distribution of the parameters $p(\alpha, T_c | y_{t_1}, \dots, y_{t_6})$, conditioned on the state observed at t_1, \dots, t_6 . Prior distributions for the unknown parameters are modeled as uniform $\alpha \sim \text{Uni}(1, 4)$ and $T_c \sim \text{Uni}(1, 21)$

As summary statistics we use the proportions of cancerous cells in the simulation grid at each observation time,

Table 2. BOLFI parameters for the inference of the breast cancer model

BOLFI-parameter	value
N_{init}	30
N_{ε}	150
T_{update}	10
σ_{acq}^2	[0.5, 0.1]
N_{sample}	2000

$$s_k := s(y_{t_k}) = \frac{1}{660} \sum_{i=1}^{660} y_{t_k, \text{cells}}^{(i)}, \quad k = 1, \dots, 6 \quad (13)$$

where $660 = 33 \cdot 20$, is the size of the simulation grid and $y_{t_k, \text{cells}}^{(i)}$, $i = 1, \dots, 660$ is the vectorised simulation grid with binary value 1 when the i th grid location contains a cancer cell. The simulations are computationally very demanding and we use BOLFI to produce the approximation of the posterior. The BOLFI parameters are listed in Table 2. We use the log-transformed euclidean distance for the response of the surrogate model within BOLFI.

After obtaining an approximate posterior curve based on the likelihood surrogate provided by BOLFI, NUTS is used to draw a sample from it, and the generative model is used to propagate the posterior sample in time under the selected treatment. This results in a simulation-based estimate of the posterior predictive distribution of the state $p(x_{t_{\text{end}}} | s_1, \dots, s_6)$. Here we use the twelve week mark $t_{\text{end}} = 4032$ as the end point. This time interval contains four full treatment periods.

4.3 Results

The inference results are illustrated in Figure 4. We see that the posterior distributions have the probability mass peaks located closely around the true values. Importantly, we are able to

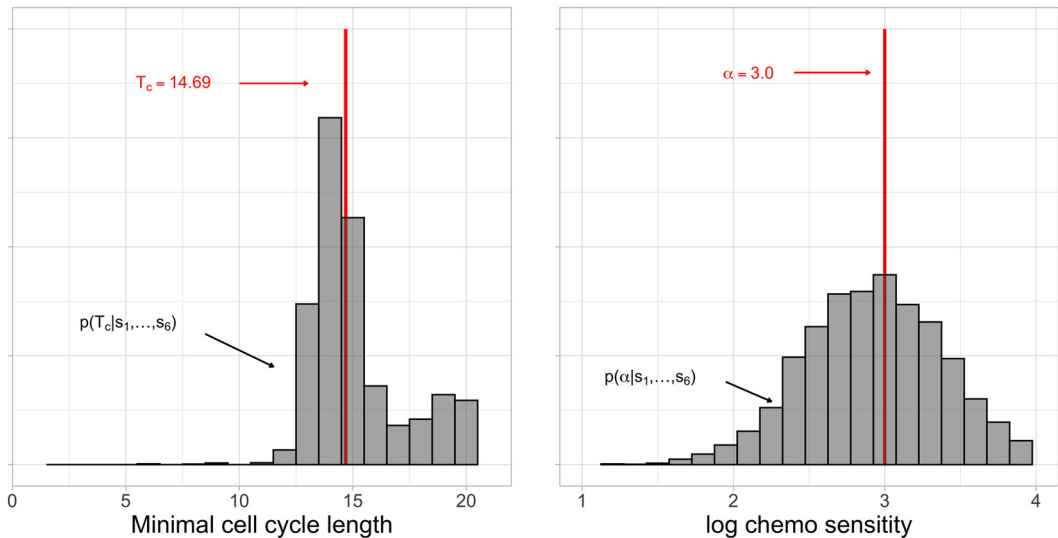


FIGURE 4. Approximated posterior distributions of the posterior marginals. The red vertical lines on the first two subplots indicate the true simulation values which were used for simulating the cancer cell growth trajectory

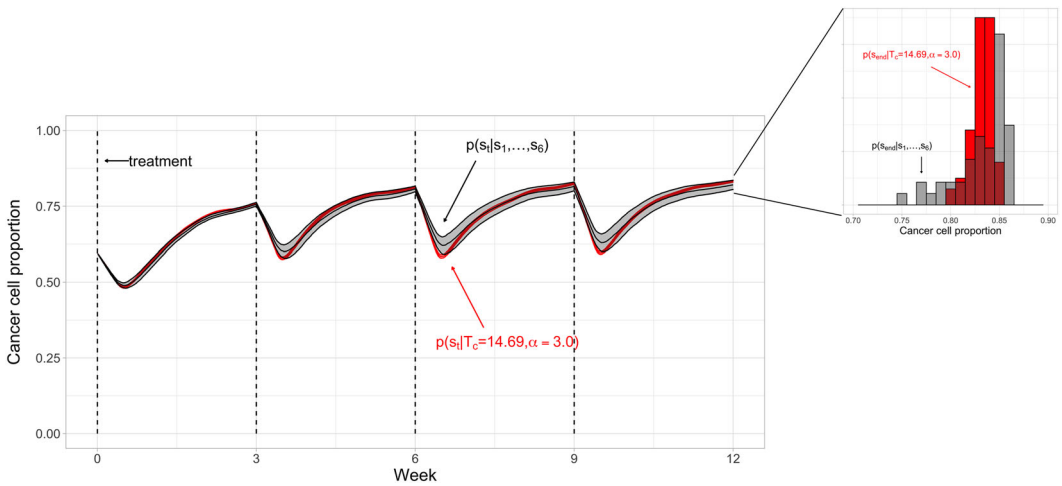


FIGURE 5. Simulated trajectories of the evolution of cancer cell proportions in the simulation grid given the true parameter values, and the forecast of the trajectory as simulated given the posterior distribution of the parameter values. The simulation end point is highlighted as the histogram on the right

use the posterior distributions to probabilistically investigate how the disease will evolve under treatment. In Figure 5 we plot the summarised state of the cancer cells for 12 week treatment period and compare it to the trajectory simulated given the true parameter values. The results indicate that the forecast is quite consistent given the inferred posterior distributions and that the current treatment most likely will not erase the cancerous growth, and other treatments should be investigated for a better outcome.

5 ABC In Astronomy With An Application To Supernova Models

Likelihood-free inference is becoming increasingly important in astronomy, where physical models cannot often be fully characterised in terms of a tractable likelihood function (Cameron & Pettitt, 2012; Schafer & Freeman, 2012; Weyant *et al.*, 2013; Ishida *et al.*, 2015; Leclercq, 2018; Picchini *et al.*, 2020). Here we evaluate the performance improvement arising from using BOLFI (Algorithm 3) instead of the ABC-PMC algorithm (introduced in Section 2.2) on an astronomical model from Jennings & Madigan (2017). The ABC-PMC version proposed for this example is slightly different from Algorithm 2. In particular, the version employed for this example requires us to properly tune the final number of iterations T , the first tolerance ϵ_1 and the quantile q_t used for adaptively decreasing the series of tolerances. We did so by following the suggestions in Lenormand *et al.* (2013) and Weyant *et al.* (2013). The choices for these three quantities are highlighted in Section 5.1.

Starting with the SuperNova ANALysis (SNANA) light curve package by Kessler *et al.* (2009) and the corresponding implementation of the SALT-II light curve fitter presented in Guy *et al.* (2010), a sample of 400 supernovae with redshift range $z \in [0.5, 1.0]$ have been simulated and then binned into $i = 20$ redshift bins. A model that describes the distance modulus as a function of redshift z is defined as:

$$\mu_i^{\text{model}}(z_i; \Omega_m, w_0) \propto 5 \log_{10} \left(\frac{c(1+z_i)}{h_0} \right) \int_0^{z_i} \frac{dz'}{E(z')}, \quad (14)$$

where $E(z) = \sqrt{\Omega_m(1+z)^3 + (1 - \Omega_m) \exp\left(3 \int_0^z d \ln(1+z') [1 + w(z')]\right)}$.

The true cosmological parameters used to generate the ‘observed’ data μ are the matter density of the universe $\Omega_m = 0.3$, the dark energy density of the universe $\Omega_\Lambda = (1 - \Omega_m) = 0.7$, the present value of the dark energy equation $\omega_0 = -1.0$ and, finally, the current Hubble constant $h_0 = 0.7$. In the following, Ω_Λ and h_0 are considered known and fixed at their input values. The goal is to estimate the two cosmological parameters Ω_m and ω_0 . The original example as presented in Jennings & Madigan (2017) is available in the *astroABC* Python package.

Jennings & Madigan (2017) added artificial noise to the data simulated through (14) by using a skew-normal distribution (Azzalini, 1985) with location, scale and skewness parameters fixed at -0.1 , 0.3 and 5.0 , respectively. By doing so, the commonly used MCMC algorithm for Bayesian statistical inference is not applicable. In fact, in order to perform the analyses by using the MCMC algorithm, Jennings & Madigan (2017) tried to add artificial noise to the data simulated through Eq. (14) by using a normal distribution with location and scale parameters fixed at -0.1 and 0.3 , respectively. The results obtained by the MCMC algorithm (see Section 5 in Jennings & Madigan (2017)) led to a poor estimation of the parameters of interest and therefore an ABC based approach is preferred (Jennings & Madigan, 2017).

The goal of the analysis presented in Jennings & Madigan (2017) was to present a comparison between this slightly more complicated model (for which the ABC-PMC algorithm is required) and a simplified version for which artificial noise is not added (for which the likelihood function is tractable and therefore MCMC is possible). The contour plot of the joint distribution (Ω_m, ω_0) obtained by using the MCMC algorithm and the marginal posterior means for Ω_m and ω_0 are available in Jennings & Madigan (2017). Beyond retrieving reliable summaries such as marginal posterior means and marginal highest posterior density (HPD) intervals for the parameters of interest, it is of relevance from a physics standpoint to evaluate how well a likelihood-free inference approach preserves the so-called ‘banana-shape’ (Kessler *et al.*, 2013; Hinton *et al.*, 2019) that describes the relation between Ω_m and ω_0 . The ‘banana-shape’ is not expected to significantly change after the artificial noise is added to the data, as shown by Jennings & Madigan (2017).

In order to use BOLFI (introduced in Section 2.3) and the ABC-PMC sampler for the estimation of the matter density of the universe parameter Ω_m and the dark energy equation parameter ω_0 , two additional quantities must be specified. As highlighted in Section 2, the distance function used to compare the observed and the simulated data and the prior distributions for the parameters of interest (i.e. for this example Ω_m and ω_0) must be defined. Because in this section we want to compare the performance of BOLFI with the ABC-PMC sampler, we performed the analysis for both methods by using the same specifications recommended in Jennings & Madigan (2017). The metric $d(\cdot, \cdot)$ that compares the observed data μ with the simulated data $\mu_{\text{sim}}(z)$ is defined as:

$$d(\mu, \mu_{\text{sim}}(z)) = \sum_i \frac{(\mu_i - \mu_{\text{sim}}(z_i))^2}{2\sigma_i^2}, \quad (15)$$

where σ_i is the error on the data point μ_i , estimated by calculating the sample variance of the observation in the i^{th} bin. We note that, in this example, Jennings & Madigan (2017) do not formally define a summary statistic for the data. Therefore the summary statistic is equivalent to the 20 dimensional vector of the data. The prior distributions, $\Omega_m \sim N(0.3, 0.5)$ and $\omega_0 \sim N(-1.0, 0.5)$, are chosen, where the prior for Ω_m is consistent with the $(0, 1)$ range for this parameter.

Table 3. BOLFI parameters for the estimation of the supernova model

BOLFI-parameter	value
N_{init}	50
$N_{\mathcal{E}}$	300
T_{update}	1
σ_{acq}^2	1
N_{sample}	1000

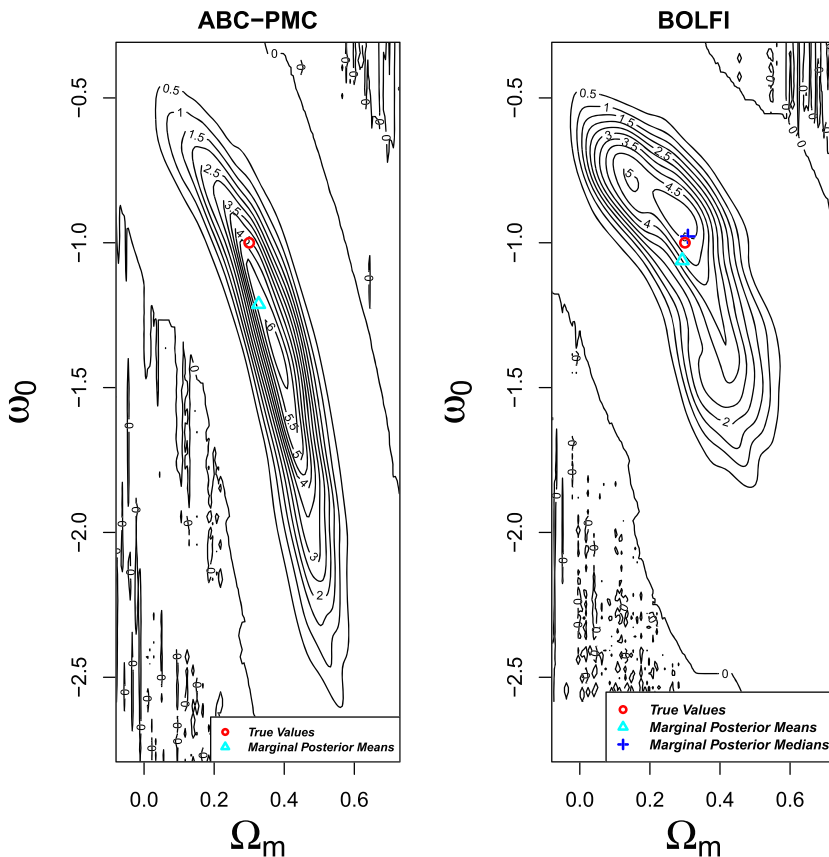


FIGURE 6. (left) Contour plot of the joint distribution (Ω_m, ω_0) obtained by the ABC-PMC sampler and (right) contour plot of the joint distribution (Ω_m, ω_0) obtained by BOLFI. The true values $(\Omega_m = 0.3, \omega_0 = 1)$ are highlighted with a red circle. Relevant regions of the parameter space are inferred and for both methods the marginal posterior means (and for BOLFI also the marginal posterior medians) are highlighted (see also Table 4). The point estimates obtained by BOLFI are closer to the true values for Ω_m and ω_0 compared with the corresponding estimates provided by ABC-PMC. The expected 'banana-shape' is reconstructed by both methods, although the contour plot obtained by BOLFI presents a smaller lower tail compared with the 'banana-shape' retrieved by using the ABC-PMC algorithm

5.1 ABC-PMC Inference and Acceleration by BOLFI

The ABC-PMC sampler from the *astroABC* package was run using $N = 1000$ particles, a total number of iterations $T = 20$ and a quantile equal to $q_t = 0.75$, which is used to reduce the ABC tolerance parameter through the iterations. It follows that, with respect to the

Table 4. Results obtained by using the ABC–PMC sampler and BOLFI. Together with the obtained point estimates of Ω_m and ω_0 (the weighted marginal posterior means for the ABC–PMC analyses and the marginal posterior means and marginal posterior medians for BOLFI), the marginal HPD credible intervals 90% are displayed. BOLFI provides point estimates for the parameters that are closer to the true values compared with the corresponding estimates retrieved by the ABC–PMC sampler. HPD credible intervals 90% for Ω_m and ω_0 indicate that the probability mass in the ‘banana-shape’ produced by BOLFI is within a more compact region with a smaller lower tail compared with the ‘banana-shape’ retrieved by using the ABC–PMC algorithm.

	True values	ABC–PMC (90% HPD)	BOLFI mean (90% HPD)	BOLFI median (90% HPD)
Ω_m	0.3	0.36 (0.18, 0.54)	0.29 (0.055, 0.48)	0.31 (0.055, 0.48)
ω_0	−1	−1.22 (−2.11, −0.62)	−1.06 (−1.56, −0.61)	−0.98 (−1.56, −0.61)

ABC–PMC sampler defined in Algorithm 2, the vector of the ABC tolerances $\epsilon_1, \dots, \epsilon_{20}$ is not tuned in advance by the researcher but instead defined through the approach suggested by Lenormand *et al.* (2013), among others. The perturbation kernel used from the second iteration onwards in Algorithm 2 follows the recommended Gaussian distribution, having variance equal to twice the empirical coverage amongst the particles for both Ω_m and ω_0 . Following the choices by Jennings & Madigan (2017), the first tolerance was fixed to $\epsilon_1 = 500$ and the final tolerance was $\epsilon_{20} = 29.82$. We tried different combinations of T, q_t in order to identify a suitable level for the final tolerance at which to stop the ABC–PMC algorithm, resulting in ϵ_T close to 30.

The computational efficiency of both the ABC–PMC sampler and BOLFI was investigated. Our parameter choices for BOLFI are summarised in Table 3. We used a Metropolis-Hasting algorithm to produce the N_{sample} approximate posterior draws. With the selected particle sample size of $N = 1000$, the ABC–PMC sampler takes 90 minutes to produce the final ABC posterior distribution. In comparison, BOLFI produces the posterior distribution in 3 minutes. The gain in computational efficiency is a clear advantage obtained by using BOLFI over the ABC–PMC sampler. Figure 6 displays the contour plots of the joint distribution (Ω_m, ω_0) obtained by the ABC–PMC sampler and by BOLFI, while the point estimates for Ω_m and ω_0 obtained by the ABC–PMC analysis (the weighted marginal posterior means) and the estimates retrieved by BOLFI (the marginal posterior means and marginal posterior medians) are summarised in Table 4. It is possible to note that BOLFI provides marginal posterior means closer to the true values ($\Omega_m = 0.29$ and $\omega_0 = -1.06$) compared with the corresponding estimates provided by the ABC–PMC ($\Omega_m = 0.36$ and $\omega_0 = -1.22$). Both procedures are able to reconstruct the expected ‘banana-shape’, although the contour plot obtained by BOLFI presents a smaller lower tail compared with the ‘banana-shape’ retrieved by using the ABC–PMC algorithm. This observation is also confirmed by looking at the marginal 90% HPD credible intervals for Ω_m and ω_0 , reported in parentheses in Table 4. However, repeated experiments would be required to quantify how well the methods based on different approximations actually estimate the posterior distributions.

6 ABC Forecasting With An Application To Optimal Portfolio Allocation

Thus far the discussion has centred primarily on the use of ABC as an inferential method, and on improving the performance of more basic versions of ABC via BOLFI. Whilst BOLFI has been used for *prediction* in two of the three previous illustrations, any comparison with predictions that *would* have been produced via exact (likelihood-based) Bayesian inference has not been possible, due simply to the fact that the likelihood function is inaccessible (or, at the very least, challenging) in the given examples.

In this section, we step back from the illustration of ABC in situations where it is essential, to an artificial situation in which the exact posterior and, hence, the exact predictive, is available.

The aim of the exercise is to illustrate that an ABC-based predictive distribution can be a very accurate approximation of the exact predictive and, hence, yield equally reliable forecasts. This then provides some reassurance that prediction based on an LFI method has value in those cases where it is indeed the only option, such as those illustrated in this paper. We revert here to the simplest form of rejection ABC in order to emphasise that predictive accuracy does not necessarily depend on using an optimal version of the inferential algorithm.

Let Y_{n+1} denote a scalar random variable, observed at time $n+1$, and generated according to $p(y_{n+1}|\theta, y_{\text{obs}})$, where $y_{\text{obs}} = [y_1, y_2, \dots, y_n]^T$. The exact Bayesian predictive (or forecast distribution - we use the terms ‘forecast’ and ‘prediction’, and their variants, synonymously) is $p(y_{n+1}|y_{\text{obs}}) = \int p(y_{n+1}|\theta, y_{\text{obs}})p(\theta|y_{\text{obs}})d\theta$, where $p(\theta|y_{\text{obs}})$ is the exact posterior, defined in the usual way, and y_{n+1} denotes a value in the support of Y_{n+1} . In cases where $p(y_{\text{obs}}|\theta)$ and, hence, $p(\theta|y_{\text{obs}})$, is inaccessible, $p(y_{n+1}|y_{\text{obs}})$ is also inaccessible, and a natural solution is to define the *approximate* Bayesian predictive, $g(y_{n+1}|y_{\text{obs}}) = \int p(y_{n+1}|\theta, y_{\text{obs}})\pi_{\epsilon}(\theta|y_{\text{obs}})d\theta$, with $g(y_{n+1}|y_{\text{obs}})$ produced using the ABC posterior in (2), $\pi_{\epsilon}(\theta|y_{\text{obs}})$, which could be accessed using either Algorithm 1 or Algorithm 2. Alternatively, BOLFI (Algorithm 3) could be used to produce an approximate posterior sample, as described in Section 2.3, noting that a notational change would be required to represent this approximate posterior in the expression for $g(y_{n+1}|y_{\text{obs}})$.

In an extensive exploration, Frazier *et al.* (2019) demonstrate, both theoretically and in practical situations, that the differences between $g(y_{n+1}|y_{\text{obs}})$ and $p(y_{n+1}|y_{\text{obs}})$ can be *negligible*, despite there sometimes being *substantial* differences between the approximate and exact posteriors. Moreover, the authors also demonstrate that ABC-based forecasting can produce reliable forecasts in a fraction of the time required for exact methods, owing to the speed with which $\pi_{\epsilon}(\theta|y_{\text{obs}})$ can be constructed, relative to $p(\theta|y_{\text{obs}})$.

The simplicity and computational speed of ABC-based forecasting is important in the sphere of economics and finance, where the need to predict the actions, and interactions, of large numbers of economic ‘agents’ leads to complex dynamic models that often challenge the MCMC toolkit and exact Bayesian forecasting. To illustrate this novel use of ABC, we document the performance of ABC-based forecasting, relative to exact forecasting, in a particular empirical example: the production of a utility-optimising financial portfolio. We reiterate that our illustration is based on the simplest version of ABC, as given in Algorithm 1.

6.1 Optimal Portfolio Allocation: The Role of Prediction

At time n , an investor chooses to allocate her wealth W_n across m possible investment choices, where at time $(n+1)$ the different investment choices yield random returns denoted by $R_{n+1,i}$, $i = 1, 2, \dots, m$, with $R_{n+1} = [R_{n+1,1}, \dots, R_{n+1,m}]^T$. For $\Delta_m := \{\alpha \in [0, 1]^m : \sum_{i=1}^m \alpha_i = 1\}$, the individual’s wealth at time $(n+1)$ is given by $W_{n+1} = W_n[1 + \alpha^T R_{n+1}]$. The goal of portfolio analysis is to discern an “optimal” allocation rule $\alpha^{\text{opt}} \in \Delta_m$ for the portfolio $\alpha^T R_{n+1}$.

The portfolio allocation problem exhibits different solutions depending on the definition of “optimality”. A common approach in economics and finance is to find α^{opt} by maximising expected utility (of wealth) using von Neumann-Morgenstern expected utility (EU) theory (von Neumann & Morgenstern, 1953): For $u(\cdot) : [0, \infty) \rightarrow \mathbb{R}$ a utility function, and $\mathbb{E}_n(\cdot)$ denoting expectation conditional on information available at time n , the optimal allocation rule is

$$\alpha_{n+1}^{\text{opt}} = \arg \max_{\alpha_{n+1} \in \Delta_m} \mathbb{E}_n[u(W_{n+1})]. \quad (16)$$

For this illustration we consider the canonical risk-averse investor with power utility function, $u(W_{n+1}) = \frac{W_{n+1}^{1-\gamma}}{1-\gamma}$, where $\gamma > 1$ denotes the risk aversion parameter. Given a model for conditional returns, $p(R_{n+1}|R_{\text{obs}})$, where $R_{\text{obs}} = [R_1, R_2, \dots, R_n]^T$, the allocation rule that maximises the expected utility is given by

$$\alpha_{n+1}^{\text{opt}} = \arg \max_{\alpha_{n+1} \in \Delta_m} E_n(u(W_{n+1})) = \arg \max_{\alpha_{n+1} \in \Delta_m} \int u(W_{n+1}) p(R_{n+1}|R_{\text{obs}}) dR_{n+1}. \quad (17)$$

Given that $p(R_{n+1}|R_{\text{obs}})$ is typically unavailable in closed-form, we must resort to simulation to compute the integral in (17): if we can obtain M draws from $p(R_{n+1}|R_{\text{obs}})$, denoted as $R_{n+1}^{(j)}$, $j = 1, \dots, M$, the value of $\alpha_{n+1}^{\text{opt}}$ can be approximated numerically by solving

$$\hat{\alpha}_{n+1}^{\text{opt}} \equiv \arg \max_{\alpha_{n+1} \in \Delta_m} \sum_{j=1}^M u\left(W_n \left[1 + \alpha_{n+1}^T R_{n+1}^{(j)}\right]\right) / M. \quad (18)$$

The optimally allocated portfolio at time $n+1$ is then given by $W_{n+1}^{\text{opt}} = W_n[1 + \hat{\alpha}_{n+1}^{\text{opt}T} R_{n+1}^{\text{obs}}]$, with utility, $u(W_{n+1}^{\text{opt}})$, where R_{n+1}^{obs} denotes the observed value of R_{n+1} . Repeating this exercise over an evaluation period yields a series of such utility values, which can be averaged to produce an estimate of $\mathbb{E}_n[u(W_{n+1}^{\text{opt}})]$, $\hat{\mathbb{E}}_n[u(W_{n+1}^{\text{opt}})]$. In what follows, we demonstrate that, in this particular representative example, there is negligible difference between the values of $\hat{\mathbb{E}}_n[u(W_{n+1}^{\text{opt}})]$ produced via *exact* and *approximate* predictives associated with a given model for R_{n+1} .

6.2 Empirical Example

For the illustration we consider a very simple portfolio, $W_{n+1} = W_n[1 + \alpha_{n+1}^T R_{n+1}]$, where $R_{n+1} = [\exp(y_{n+1}), \exp(rf_{n+1})]^T$ and $\alpha_{n+1} = [\alpha_{n+1}, 1 - \alpha_{n+1}]^T$, with y_{n+1} the logarithmic return on the S&P500 market portfolio and rf_{n+1} the (known) logarithmic return on a one-month constant maturity US Treasury Bill, over the period n to $(n+1)$. The monthly S&P500 Total Returns index is sourced from the Chicago Board of Options Exchange (CBOE), and captures both capital gains as well as dividend yields. The Treasury Bill is sourced from the Federal Reserve St Louis (FRED) database. The weight $\alpha_{n+1} \in [0, 1]$ determines the proportion of wealth allocated to the risky market portfolio, and is to be chosen according to (18). (See, for example, (Billio *et al.*, 2013).) The monthly data extends from June 1986 to June 2018, totaling 384 observations. The first 324 observations are reserved for inference/training and the final 60 observations (five years) used to estimate expected utility.

We produce the 60 one-step-ahead exact and approximate predictive distributions over the evaluation period based on the following stochastic volatility model for y_t ,

$$y_t = \mu + \sqrt{V_t} \varepsilon_t \quad (19)$$

$$\ln V_t = \omega + \rho \ln V_{t-1} + \sigma v_t, \quad (20)$$

where $\varepsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ and $v_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$, with ε_t and v_t independent for all $t = 1, 2, \dots, n$. The unknowns comprise the static parameter vector, $\theta = [\rho, \sigma, \omega, \mu]^T$, the vector of (in-sample) latent volatilities, $V = [V_1, V_2, \dots, V_n]^T$, and the unknown V_{n+1} on which y_{n+1} is conditioned. This model is adequate for capturing the behaviour of monthly returns, and allows for a ready application of an exact algorithm, for this comparative exercise.

Expanding windows are used to produce five predictives - one exact and four approximate - for the 60 time points in the evaluation period, with draws of θ updated only yearly in all cases. Draws from $p(\theta, V|y_{\text{obs}})$, used to estimate the *exact* predictive, are produced using a particle Metropolis Hastings (PMH) scheme (Andrieu & Doucet, 2010). The four approximate predictives are produced using the auxiliary likelihood-based ABC approach of Martin *et al.* (2019) to specify the summary statistics in Algorithm 1. Four alternative models from the generalised autoregressive conditionally heteroscedastic (GARCH) family of volatility models are chosen to define the auxiliary likelihood: a GARCH model with normal errors, a GARCH model with Student's t errors, an asymmetric GARCH with normal errors, and an asymmetric GARCH model with Student's t errors. We refer the reader to Chapter 9 of Brooks (2014) for a description of these, and related, models; we simply highlight here that such conditionally deterministic volatility models are suitable for an auxiliary likelihood-based version of ABC, given the closed-form nature of their likelihoods. A nearest-neighbour version of rejection ABC is used, with $N = 34992$ and a selection probability of 0.86%. See Frazier *et al.* (2018) for an explanation of the rule adopted to determine these values for a given sample size, n .

For each predictive, $M = 2000$ draws are used to produce an optimal weight, as per (18), the associated optimal portfolio, and its utility. The 60 values of utility are then averaged for a particular predictive method, with five such numbers produced.

Before discussing the resulting (estimated) expected utilities, we present representative exact and approximate predictives in Panels A and B of Figure 7, for two particular months. We label the predictive estimated using PMH as 'Exact' and the approximate predictives constructed using each of these four auxiliary models listed earlier as: 'ABC1', 'ABC2', 'ABC3' and 'ABC4', respectively. The plots demonstrate that there is very little to distinguish between the approximate and exact predictive methods, in particular for June 2018. The similarity between the exact and approximate predictives is further borne out in the expected utility calculations.

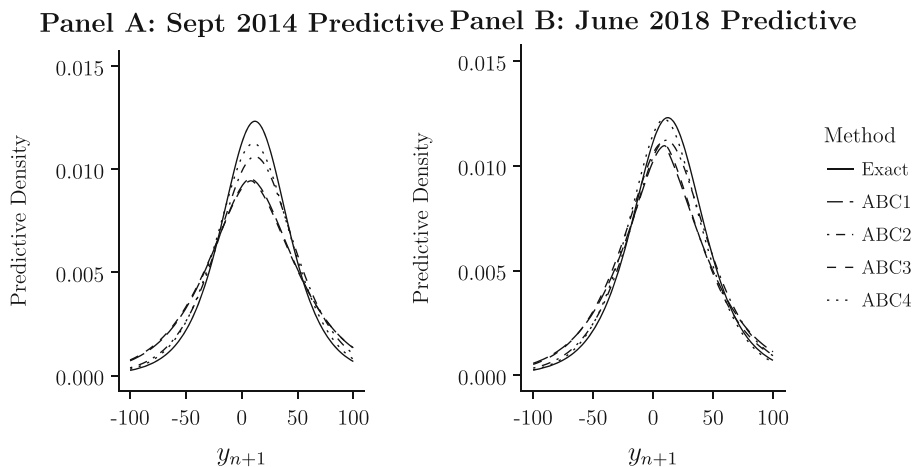


FIGURE 7. One-step-ahead predictive densities for September 2014 (Panel A) and June 2018 (Panel B): exact (PMH-based) and approximate (ABC-based)

For a fixed risk aversion parameter of $\gamma = 4$, the exact approach yields expected utility of -4.72 , while the four approximate methods yield expected utilities of -4.69 , -4.70 , -4.69 and -4.69 . Moreover, the computation time required to produce the exact expected utility is ten times larger than the time required to produce the expected utility via the *slowest* of the approximate methods, and fifteen times greater than the computation time for the *fastest* approximate method. A comprehensive set of expected utility estimates were produced, for different degrees of risk aversion, with the same negligible difference between exact and approximate results being in evidence.

We close this section by noting that the close match of the exact and approximate predictives is not simply an artefact of the particular predictive model chosen. In Frazier *et al.* (2019) (cited earlier), comparable numerical results are documented for a range of different models, including for both continuous and discrete data. Moreover, under the required regularity conditions for Bayesian consistency of both the exact and ABC posteriors, the one-step-ahead exact and ABC-based predictives are shown to merge in the sense that, for a large enough sample size, and for models of *any* fixed dimension, the predictive distributions are identical.

7 Discussion

ABC and similar likelihood-free inference methods are emerging as an important part of the analysis toolbox in various application areas where we are able to simulate realistic data, but the models are too complicated for likelihood-based inference. Here we have demonstrated various aspects of using this approach in challenging real-world applications that go beyond the typical benchmark examples used in the literature.

A particularly interesting combination of inference and generative modeling is the possibility of learning the parameters of a system based on observed data and simulating the performance under various scenarios, treatments or interventions. For example, one of our case studies involved predicting a patient-specific cancer treatment outcome, which could in the future be an important part of treatment planning, especially as the mathematical modeling of disease evolution under treatments is rapidly improving (Kozłowska *et al.*, 2018). A similar application is found in policy planning for epidemics, as the models of infection control are improving, and different ‘what-if’-scenarios can be explored. Such activities have recently been extensively performed across the world during the Covid-19 pandemic and the iteratively improving understanding about the epidemiology of the virus illustrates the need for proper representations of uncertainties in the model components.

As the modeling proficiency increases, we will be faced with even more difficult inference tasks in higher dimensional domains, with computationally heavier simulators and, in some applications, even requirements for carrying out the inference (and prediction) in real-time. These challenges call for further advances from the statistics and computation community on both the inference algorithms and their efficient implementation via user-friendly software.

ACKNOWLEDGEMENT

This work was supported by the European Research Council grant number 742158, European Union’s Horizon 2020 Research and Innovation Programme, Grant Agreement number 847912, Research Council of Norway, Grant Numbers 237718, 311188 and 309273, the Academy of Finland grant number 1316602, Australian Research Council Discovery Grants DP170100729 and DP200101414 and Australian Research Council Early Career Researcher Award DE200101070.

REFERENCES

- Alarcón, T., Byrne, H.M. & Maini, P.K. (2010). A multiple scale model for tumor growth. *Multiscale Model. Simul.*, **3**(2), 440.
- Althaus, C.L. (2014). Estimating the reproduction number of ebola virus (ebov) during the 2014 outbreak in west africa. *PLoS currents*, **6**.
- An, Z., South, L.F. & Drovandi, C. 2019. BSL: An R Package for Efficient Parameter Estimation for Simulation-Based Models via Bayesian Synthetic Likelihood. arXiv preprint 1907.10940.
- An, Z., South, L.F., Nott, D.J. & Drovandi, C.C. (2019). Accelerating Bayesian synthetic likelihood with the graphical lasso. *J. Comput. Graph. Stat.*, **28**(2), 471–475.
- Andrieu, C. & Doucet, A. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B*, **72**, 263–342.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian J. Stat.*, **12**(2) 171–178.
- Beaumont, M.A. (2019). Approximate Bayesian Computation. *Ann. Rev. Stat. Appl.*, **6**(1), 379–403. <https://doi.org/10.1146/annurev-statistics-030718-105212>
- Beaumont, M.A., Cornuet, J.-M., Marin, J.-M. & Robert, C.P. (2009). Adaptive approximate Bayesian computation. *Biometrika*, **96**(4), 983–990.
- Billio, M., Casarin, R., Ravazzola, F. & van Dijk, H.K. (2013). Time-varying combinations of predictive densities using nonlinear filtering. *J. Econ.*, **177**, 213–232.
- Blum, M.G.B. (2010). Approximate Bayesian computation: A nonparametric perspective. *J. Am. Stat. Assoc.*, **105**(491), 1178–1187.
- Britton, T. & Tomba, G.S. (2019). Estimation in emerging epidemics: Biases and remedies. *J. R. Soc. Interface*, **16**(150), 20180670.
- Brochu, E., Cora, V.M. & de Freitas, N. 2010. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv preprint arXiv:1012.2599.
- Brooks, C. (2014). *Introductory Econometrics for Finance*. Cambridge University Press.
- Cameron, E. & Pettitt, A.N. (2012). Approximate bayesian computation for astronomical model analysis: A case study in galaxy demographics and morphological transformation at high redshift. *Monthly Notices R. Astronomical Soc.*, **425**, 44–65.
- Chinazzi, M., Davis, J.T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Pastore y Piontti, A., Mu, K., Rossi, L., Sun, K., Viboud, C., Xiong, X., Yu, H., Halloran, M.E., Longini, I.M. & Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science*, **368**(6489), 395–400. <https://science.sciencemag.org/content/368/6489/395.full.pdf>
- Cisewski-Kehe, J., Weller, G., Schafer, C. et al. (2019). A preferential attachment model for the stellar initial mass function. *Electronic J. Stat.*, **13**(1), 1580–1607.
- Clarté, G., Robert, C.P., Ryder, R.J. & Stoehr, J. (2021). Componentwise approximate Bayesian computation via Gibbs-like steps. *Biometrika*, **108**(3), 591–607. <https://academic.oup.com/biomet/advance-article-pdf/doi/10.1093/biomet/asaa090/36332711/asaa090.pdf>
- Corander, J., Fraser, C., Gutmann, M.U., Arnold, B., Hanage, W.P., Bentley, S.D., Lipsitch, M. & Croucher, N.J. (2017). Frequency-dependent selection in vaccine-associated pneumococcal population dynamics. *Nature Ecology Evol.*, **1**(12), 1950–1960.
- Cranmer, K., Brehmer, J. & Louppe, G. (2020). The frontier of simulation-based inference. *Proc. Nat. Acad. Sci.*, **117**(48), 30055–30062. <https://www.pnas.org/content/117/48/30055.full.pdf>
- Csilléry, K., Blum, M.G.B., Gaggiotti, O.E. & François, O. (2010). Approximate Bayesian computation (ABC) in practice. *Trends Ecology Evol.*, **25**(7), 410–418.
- Depeweg, S., Hernández-Lobato, J.M., Doshi-Velez, F. & Udluft, S. 2017. Decomposition of uncertainty in bayesian deep learning for efficient and risk-sensitive learning. arXiv preprint arXiv:1710.07283.
- Drovandi, C.C. & Pettitt, A.N. (2011). Estimation of parameters for macroparasite population evolution using approximate bayesian computation. *biometrics. Stat. Comput.*, **67**(1), 225–233.
- Drovandi, C.C., Pettitt, A.N. & Faddy, M.J. (2011). Approximate bayesian computation using indirect inference. *J. R. Stat. Soc. Ser. C*, **60**, 1–21.
- Drovandi, C.C., Pettitt, A.N. & Lee, A. (2015). Bayesian indirect inference using a parametric auxiliary model. *Stat. Sci.*, **30**, 72–95.
- Dutta, R., Schoengens, M., Onnela, J.-P. & Mira, A. (2017). Abcpy: A user-friendly, extensible, and parallel library for approximate Bayesian computation. In *Proceedings of the Platform for Advanced Scientific Computing Conference, PASC '17*, pp. 8:1–8:9. ACM, Association for Computing Machinery: New York, NY, USA.
- Ebert, A., Dutta, R., Mengersen, K., Mira, A., Ruggeri, F. & Wu, P. (2021). Likelihood-free parameter estimation for dynamic queueing networks: Case study of passenger flow in an international airport terminal. *J. R. Stat. Soc.: Ser. C (Appl. Stat.)*, **70**(3), 770–792. <https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/rssc.12487>

- Fearnhead, P. & Prangle, D. (2012). Constructing summary statistics for approximate bayesian computation: semi-automatic approximate bayesian computation. *J. R. Stat. Soc.: Ser. B (Stat. Methodol.)*, **74**(3), 419–474. <https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2011.01010.x>
- Forbes, F., Nguyen, H.D., Nguyen, T.T. & Arbel, J. 2021. Approximate Bayesian computation with surrogate posteriors. working paper or preprint.
- Frazier, D.T., Maneesoonthorn, W., Martin, G.M. & McCabe, B.P.M. (2019). Approximate Bayesian forecasting. *Int. J. Forecast.*, **35**(2), 521–539.
- Frazier, D.T., Martin, G.M., Robert, C.P. & Rousseau, J. (2018). Asymptotic properties of approximate bayesian computation. *Biometrika*, **105**, 593–607.
- Frazier, D.T., Robert, C.P. & Rousseau, J. (2020). Model misspecification in approximate Bayesian computation: consequences and diagnostics. *J. R. Stat. Soc. Ser. B*, **82**(2), 421–444.
- Gal, Y., Islam, R. & Ghahramani, Z. 2017. Deep Bayesian active learning with image data. arXiv preprint arXiv:1703.02910.
- Gebhardt, C., Oulasvirta, A. & Hilliges, O. (2020). Hierarchical reinforcement learning explains task interleaving behavior. *Comput. Brain Behavior*, **4**(3), 284–304.
- Grazian, C. & Fan, Y. (2020). A review of approximate Bayesian computation methods via density estimation: Inference for simulator-models. *WIREs Comput. Stat.*, **12**(4), e1486. <https://onlinelibrary.wiley.com/doi/pdf/10.1002/wics.1486>
- Green, P.J., Latuszyński, K., Pereyra, M. & Robert, C.P. (2015). Bayesian computation: A summary of the current state, and samples backwards and forwards. *Stat. Comput.*, **25**(4), 835–862.
- Gutmann, M.U. & Corander, J. (2016). Bayesian optimization for likelihood-free inference of simulator-based statistical models. *J. Machine Learn. Res.*, **17**(125), 1–47.
- Gutmann, M.U., Dutta, R., Kaski, S. & Corander, J. (2018). Likelihood-free inference via classification. *Stat. Comput.*, **28**(2), 411–425.
- Guy, J., Sullivan, M., Conley, A., Regnault, N., Astier, P., Balland, C., Basa, S., Carlberg, R.G., Fouchez, D., Hardin, D. & Hook, I.M. (2010). The supernova legacy survey 3-year sample: Type ia supernovae photometric distances and cosmological constraints. *Astron. Astrophys.*, **523**, A7.
- Hinton, S.R., Davis, T.M., Kim, A.G., Brout, D., D'Andrea, C.B., Kessler, R., Lasker, J., Lidman, C., Macaulay, E., Möller, A. & Sako, M. (2019). Steve: A hierarchical Bayesian model for supernova cosmology. *Astrophys. J.*, **876**(1), 15.
- Hoffman, M.D. & Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *J. Machine Learn. Res.*, **15**(47), 1593–1623.
- Houlsby, N., Huszar, F., Ghahramani, Z. & Lengyel, M. 2011. Bayesian Active Learning for Classification and Preference Learning. arXiv preprint arXiv:1112.5745.
- Ishida, E.E.O., Vienti, S.D.P., Penna-Lima, M., Cisewski, J., de Souza, R.S., Trindade, A.M.M., Cameron, E. & Busti, V.C. (2015). cosmoabc: Likelihood-free inference via population monte carlo approximate Bayesian computation. *Astron. Comput.*, **13**, 1–11.
- Järvenpää, M., Gutmann, M.U., Pleska, A., Vehtari, A. & Marttinen, P. (2019). Efficient Acquisition Rules for Model-Based Approximate Bayesian Computation. *Bayesian Anal.*, **14**(2), 595–622.
- Jennings, E. & Madigan, M. (2017). astroabc: An approximate bayesian computation sequential monte carlo sampler for cosmological parameter estimation. *Astron. Comput.*, **19**, 16–22.
- Joyce, P. & Marjoram, P. (2008). Approximately sufficient statistics and Bayesian computation. *Stat. Appl. Gen. Molec. Biol.*, **7**(1), 1–16.
- Kangasrääsiö, A., Athukorala, K., Howes, A., Corander, J., Kaski, S. & Oulasvirta, A. (2017). Inferring cognitive models from data using approximate bayesian computation. In *Proceedings of the 2017 chi conference on human factors in computing systems*, CHI'17, pp. 1295–1306. Association for Computing Machinery, Association for Computing Machinery: New York, NY, USA.
- Kangasrääsiö, A., Jokinen, J.P.P., Oulasvirta, A., Howes, A. & Kaski, S. (2019). Parameter inference for computational cognitive models with approximate Bayesian computation. *Cognitive Sci.*, **43**(6), e12738. <https://onlinelibrary.wiley.com/doi/pdf/10.1111/cogs.12738>
- Kessler, R., Bernstein, J.P., Cinabro, D., Dilday, B., Frieman, J.A., Jha, S., Kuhlmann, S., Miknaitis, G., Sako, M., Taylor, M. & Vanderplas, J (2009). Snana: A public software package for supernova analysis. *Publications Astron. Soc. Pacific*, **121**(883), 1028.
- Kessler, R., Guy, J., Marriner, J., Betoule, M., Brinkmann, J., Cinabro, D., El-Hage, P., Frieman, J.A., Jha, S., Mosher, J. & Schneider, D.P. (2013). Testing models of intrinsic brightness variations in type ia supernovae and their impact on measuring cosmological parameters. *Astrophys. J.*, **764**(1), 48.
- Kokko, J., Remes, U., Thomas, O., Pesonen, H. & Corander, J. (2019). Pyfire: Python implementation of likelihood-free inference by ratio estimation. *Wellcome Open Res.*, **4**.

- Kozłowska, E., Färkkilä, A., Vallius, T., Carpen, O., Kemppainen, J., Grénman, S., Lehtonen, R., Hynninen, J., Hietanen, S. & Hautaniemi, S. (2018). Mathematical modeling predicts response to chemotherapy and drug combinations in ovarian cancer. *Cancer Res.*, **78**(14), 4036–4044. <https://cancerres.aacrjournals.org/content/78/14/4036.full.pdf>
- Lai, X., Geier, O., Fleischer, T., Garred, O., Borgen, E.F., Funke, S., Kumar, S., Rognes, M.E., Seierstad, T., Boerresen-Dale, A.-L., Kristensen, V.N., Engebraaten, O., Köhn-Luque, A. & Frigessi, A. (2019). Towards personalized computer simulation of breast cancer treatment: A multi-scale pharmacokinetic and pharmacodynamic model informed by multi-type patient data. *Cancer Res.*, **73**, 41–49.
- Lai, X., Taskén, H.A., Mo, T., Funke, S.W., Frigessi, A., Rognes, M.E. & Köhn-Luque, A. (2021). A scalable solver for a stochastic, hybrid cellular automaton model of personalized breast cancer therapy. *Int. J. Numer. Methods Biomed. Eng.*, **38**(1), e3542. <https://onlinelibrary.wiley.com/doi/pdf/10.1002/cnm.3542>
- Leclercq, F. (2018). Bayesian optimization for likelihood-free cosmological inference. *Phys. Rev. D*, **98**(6), 63511.
- Lenormand, M., Jabot, F. & Deuant, G. (2013). Adaptive approximate bayesian computation for complex models. *Comput. Stat.*, **6**(28), 2777–2796.
- Li, W. & Fearnhead, P. (2018a). Convergence of regression-adjusted approximate Bayesian computation. *Biometrika*, **105**(2), 301–318.
- Li, W. & Fearnhead, P. (2018b). On the asymptotic efficiency of approximate Bayesian computation estimators. *Biometrika*, **105**(2), 285–299.
- Lintusaari, J., Blomstedt, P., Rose, B., Sivula, T., Gutmann, M.U., Kaski, S. & Corander, J. (2019). Resolving outbreak dynamics using Approximate Bayesian Computation for stochastic birth-death models [version 2; peer review: 2 approved]. *Wellcome Open Res*, **4**(14).
- Lintusaari, J., Gutmann, M.U., Dutta, R., Kaski, S. & Corander, J. (2017). Fundamentals and Recent Developments in Approximate Bayesian Computation. *System. Biol.*, **66**, 66–82.
- Lintusaari, J., Vuollekoski, H., Kangasrääsiö, A., Skytén, K., Järvenpää, M., Martinen, P., Gutmann, M.U., Vehtari, A., Corander, J. & Kaski, S. (2018). ELFI: Engine for Likelihood-Free Inference. *J. Machine Learn. Res.*, **19**(16), 1–7.
- Lueckmann, J.-M., Bassetto, G., Karaletsos, T. & Macke, J.H. 2018. Likelihood-free inference with emulator networks. arXiv preprint arXiv:1805.09294.
- Marin, J.-M., Pillai, N.S., Robert, C.P. & Rousseau, J. (2014). Relevant statistics for Bayesian model choice. *J. R. Stat. Soc.: Ser. B (Stat. Methodol.)*, **76**, 833–859.
- Marin, J.-M., Pudlo, P., Robert, C.P. & Ryder, R.J. (2012). Approximate Bayesian computational methods. *Stat. Comput.*, **22**, 1167–1180.
- Marjoram, P. (2013). Approximate Bayesian Computation. *OA Gen.*, **1**, 853.
- Marjoram, P., Molitor, J., Plagnol, V. & Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *Proc. Nat. Acad. Sci. United States Am.*, **100**, 15324–15328.
- Martin, G.M., McCabe, B., Rendan, P.M., Frazier, D.T., Maneesoonthorn, W. & Robert, C.P. (2019). Auxiliary likelihood-based approximate Bayesian computation in state space models. *J. Comput. Graph. Stat.*, **28**(3), 508–522.
- McKinley, T., Cook, A.R. & Deardon, R. (2009). Inference in epidemic models without likelihoods. *Int. J. Biostat.*, **5**(1), 1–37.
- Moral, P.D., Doucet, A. & Jasra, A. (2012). An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Stat. Comput.*, **22**(5), 1009–1020.
- Ong, V.M.-H., Nott, D.J., Tran, M.-N., Sisson, S.A. & Drovandi, C.C. (2018). Likelihood-free inference in high dimensions with synthetic likelihood. *Comput. Stat. Data Anal.*, **128**, 271–291.
- Papamakarios, G., Sterratt, D.C. & Murray, I. (2019). Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS) 2019*, Vol. **89**, pp. 837–848. PMLR, Society for Artificial Intelligence and Statistics.
- Picchini, U., Simola, U. & Corander, J. 2020. Adaptive mcmc for synthetic likelihoods and correlated synthetic likelihoods. arXiv preprint arXiv:2004.04558.
- Prangle, D. (2017). Adapting the ABC Distance Function. *Bayesian Anal.*, **12**, 289–309.
- Price, L.F., Drovandi, C.C., Lee, A. & Nott, D.J. (2018). Bayesian synthetic likelihood. *J. Comput. Graph. Stat.*, **27**(1), 1–11.
- Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A. & Feldman, M.W. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molec. Biol. Evol.*, **16**(12), 1791–1798.
- Rasmussen, C.E. & Williams, K.I. (2006). *Gaussian processes for machine learning*. MIT Press.
- Ribba, B., Alarcón, T., Marron, K., Maini, P.K. & Agur, Z. (2004). The use of hybrid cellular automaton models for improving cancer therapy. In *International Conference on Cellular Automata*, pp. 444–453, Springer.

- Rodrigues, G.S., Nott, D.J. & Sisson, S.A. (2020). Likelihood-free approximate Gibbs sampling. *Stat. Comput.*, **30**, 1057–1073.
- Schafer, C.M. & Freeman, P.E. (2012). *Statistical challenges in modern astronomy V*, Lecture Notes in Statistics. Springer.
- Shen, P., Lees, J.A., Bee, G.C.W., Brown, S.P. & N., W.J. (2019). Pneumococcal quorum sensing drives an asymmetric owner-intruder competitive strategy during carriage via the competence regulon. *Nature Microbiol.*, **4**, 198–208.
- Silk, D., Filippi, S. & Stumpf, M.P.H. (2013). Optimizing threshold-schedules for sequential approximate bayesian computation: Applications to molecular systems. *Stat. Appl. Gen. Molecul. Biol.*, **5**(12), 603–618.
- Simola, U., Cisewski-Kehe, J., Gutmann, M.U. & Corander, J. (2020). Adaptive approximate bayesian computation tolerance selection. *Bayesian Anal.*, **16**(2), 397–423.
- Simola, U., Pelssers, B., Barge, D., Conrad, J. & Corander, J. (2019). Machine learning accelerated likelihood-free event reconstruction in dark matter direct detection. *J. Instrument.*, **14**(3), P03004.
- Sisson, S.A., Fan, Y. & Beaumont, M. (2018). *Handbook of approximate bayesian computation*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, Taylor & Francis Group.
- Sottoriva, A. & Tavaré, S. (2010). Integrating approximate Bayesian computation with complex agent-based models for cancer research. In *Proceedings of Compstat'2010*, pp. 57–66. Springer.
- Srinivas, N., Krause, A., Kakade, S.M. & Seeger, M. (2010). Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the 27th International Conference on Machine Learning (ICML)*, Israel, ICML.
- Tavaré, S., Balding, D.J., Griffiths, R.C. & Donnelly, P. (1997). Inferring coalescence times from dna sequence data. *Genetics*, **145**(2), 505–518.
- Tejero-Cantero, A., Boelts, J., Deistler, M., Lueckmann, J.-M., Durkan, C., Gonçalves, P.J., Greenberg, D.S. & Macke, J.H. (2020). SBI: A toolkit for simulation-based inference. *J. Open Source Softw.*, **5**(52), 2505.
- Thomas, O., Dutta, R., Corander, J., Kaski, S. & Gutmann, M.U. (2022). Likelihood-free inference by ratio estimation. *Bayesian Anal.*, **17**(1), 1–31. Advance publication.
- Toni, T., Welch, D., Strelkowa, N., Ipsen, A. & Stumpf, M.P.H. (2009). Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interf.*, **6**(31), 187–202.
- Vihola, M. & Franks, J. (2020). On the use of approximate Bayesian computation Markov chain Monte Carlo with inflated tolerance and post-correction. *Biometrika*, **107**, 381–395.
- von Neumann, J. & Morgenstern, O. (1953). *Theory of Games and Economic Behavior*. Princeton University Press.
- WHO Ebola Response Team (2014). Ebola virus disease in West Africa – The first 9 months of the epidemic and forward projections. *New England J. Med.*, **371**(16), 1481–1495.
- Weyant, A., Schafer, C. & Wood-Vasey, W.M. (2013). Likelihood-free cosmological inference with type Ia supernovae: approximate Bayesian computation for a complete treatment of uncertainty. *Astrophys. J.*, **764**, 116.
- Wood, S.N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, **466**, 1102–1107.

[Received January 2022; accepted September 2022]