ORIGINAL PAPER

Approximate Bayesian computation sequential Monte Carlo via random forests

Khanh N. Dinh[1] · Cécile Liu[1,2] · Zijin Xiang[1,3] · Zhihan Liu[1,4] · Simon Tavaré[1]

Abstract

Approximate Bayesian Computation (ABC) is a popular inference method when likelihoods are hard to come by. Practical bottlenecks of ABC applications include selecting statistics that summarize the data without losing too much information or introducing uncertainty, and choosing distance functions and tolerance thresholds that balance accuracy and computational efficiency. Recent studies have shown that ABC methods using random forest (RF) methodology perform well while circumventing many of ABC's drawbacks. However, RF construction is computationally expensive for large numbers of trees and model simulations, and there can be high uncertainty in the posterior if the prior distribution is uninformative. Here we further adapt random forests to the ABC setting in two ways. The first exploits distributional random forests to provide a direct method for inferring the joint posterior distribution of parameters of interest, while the second describes a sequential Monte Carlo approach which updates the prior distribution iteratively to focus on the most likely regions in the parameter space. We show that the new methods can accurately infer posterior distributions for a wide range of deterministic and stochastic models in different scientific areas.

Contents

✉ Simon Tavaré
  st3193@columbia.edu

[1] Irving Institute for Cancer Dynamics and Department of Statistics, Columbia University, New York, NY, USA

[2] Department of Applied Mathematics, Ecole Nationale des Ponts et Chaussées, Champs-sur-Marne, France

[3] Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA, USA

[4] Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY, USA

1 Introduction

Mathematical modeling has played an important role in studying scientific phenomena. Its practical applications often depend on accurately extracting model parameters θ from experimentally observed data, y^{obs}. In the Bayesian framework, this entails inferring the posterior distribution

$$\pi(\theta \mid y^{\text{obs}}) \propto f(y^{\text{obs}} \mid \theta) \cdot \pi(\theta)$$

of the parameters from the data, exploiting the likelihood $f(y^{\text{obs}} \mid \theta)$ and the prior $\pi(\theta)$ of such observations under the model with given parameter values. However, the likelihood function is often intractable to derive theoretically, difficult to compute numerically, or too complex to optimize directly. Approximate Bayesian Computation (ABC) was proposed as an alternative method to approximate the posterior distributions in such scenarios (Tavaré et al. 1997; Fu and Li 1997; Pritchard et al. 1999; Beaumont et al. 2002). Using statistics to summarize model simulations, ABC seeks the parameters that result in minimal distance between the statistics and

those of the observed data. Different variations have been proposed to improve the performance of the original ABC method, such as ABC Monte Carlo Markov Chain (Marjoram et al. 2003) and ABC Sequential Monte Carlo (Toni et al. 2008), and they have collectively found remarkable success in applications across different scientific areas (Sisson and Fan 2018).

The accuracy and efficiency of ABC in practice depends on several factors. First, one requires a distance function to compare observed and simulated data. The metric is typically weighted so that different statistics contribute equally (Prangle 2017). However, an equally important consideration is the relevance of each statistic, defined as the amount of information that it carries toward identifying the underlying parameters. Optimizing these two criteria can be challenging (Jung and Marjoram 2011). Second, ABC requires a tolerance threshold to decide whether proposed parameters are accepted or rejected. The tolerance ϵ poses a trade-off between the computational efficiency, which increases with higher ϵ, and accuracy of the empirical posterior distributions, which improves as ϵ decreases. Choosing ϵ depends on the specific model, statistic choices and distance function, and it may require intuition or experimentation (Jung and Marjoram 2011). Finally, and most importantly, ABC's results depend on the choice of summary statistics. The approximated posterior distribution is only guaranteed to converge to the likelihood-based distribution as $\epsilon \to 0$ if the statistics are sufficient. For complex models, it is not always possible or realistic to find low-dimensional sufficient statistics (Sisson and Fan 2018, Chapter 5). This can sometimes be remedied by inclusion of many distinct statistics. However, the posterior estimation may become distorted if the selected statistics are noisy or uninformative. These factors in combination mean that optimizing the performance of ABC in practice can be challenging and require extensive experimentation.

Recent studies have shown that random forests (RF), a powerful non-parametric regression method (Breiman 2001), can be employed in the ABC context to infer posterior distributions (Raynal et al. 2019). Its appeal lies in reduced dependence on user-defined hyperparameters that are essential to traditional ABC implementations, including the metric function and tolerance threshold. Moreover, RF has been shown to perform well even if the majority of statistics are pure noise, indicating that it is significantly more tolerant to a wider selection of statistics to represent the data (Raynal et al. 2019).

Therefore, RF methods have the potential to approximate the posterior distributions well with lower computational cost. However, the inclusion of all simulations in inferring parameters, coupled with wide range in the prior distributions for some models, can sometimes result in higher uncertainty in the posterior distributions.

In this paper, we propose a new inference method, Approximate Bayesian Computation sequential Monte Carlo with random forests (ABC-SMC-(D)RF). It inherits the non-parametric nature of RF methods, but is embedded in the framework of Sequential Monte Carlo (Toni et al. 2008). The posterior distribution evolves with successive iterations to focus on the most likely regions in the parameter space, resulting in more relevant model simulations and lower uncertainty in the final distributions.

We use some traditional ABC methods in comparisons with ABC-SMC-(D)RF described in the following sections. Recent ABC implementations using random forests are reviewed in Section 2, with illustrative examples in Section 3. In Section 4, we describe ABC-SMC-(D)RF's methodology. We then demonstrate its performance, compared both to traditional ABC algorithms and previous RF-based methods, in a variety of deterministic and stochastic models (Section 5). We conclude with a discussion.

2 Random forests in the context of ABC

Although simple to implement, the ABC methods depend heavily on the choices of hyperparameters, including the summary statistic function $S(y)$, the distance function $d(s, s')$ and the tolerance threshold ϵ (Sisson and Fan 2018). Furthermore, the choices of Markov kernels to propose new particles in ABC Monte Carlo Markov Chain (ABC-MCMC) and ABC Sequential Monte Carlo (ABC-SMC) also affect the approximated posterior distributions. As random forests (RF) gain in popularity as a powerful non-parametric regression technique (Breiman 2001; Segal 2004; Rigatti 2017; Desai and Ouarda 2021), its applications in the context of ABC are also becoming more prominent. RF-based methods eliminate the need for choice of metric, tolerance level and the perturbation kernels, making them an attractive alternative to the previous ABC implementations. Importantly, the RF method is relatively robust to noise (Marin et al. 2018), in that $S(y)$ can include many poorly informative statistics without significant impact on the results.

2.1 One-dimensional ABC random forests

Raynal et al. introduced ABC random forest (ABC-RF) to approximate the posterior distribution for ABC inference problems with one parameter (Raynal et al. 2019), based on the random forest formulation (Algorithm 1) developed by Breiman (2001). The root node for each tree T_t is the reference table, consisting of parameters drawn from the prior distribution and corresponding statistics. Each node R_* in the decision tree is split into two nodes R_{*1} and R_{*2} by selecting one statistic $k_* \in \{1, \ldots, |S|\}$, where $|S|$ is the total number of summary statistics, and a threshold s_* such that the L^2-loss

between the divided sets

$$\frac{1}{|R_*|}\left(\sum_{(\theta,s)\in R_{*1}}\left(\theta-\bar{\theta}_1\right)^2 + \sum_{(\theta,s)\in R_{*2}}\left(\theta-\bar{\theta}_2\right)^2\right) \quad (1)$$

is minimized. Here $|R_*|$ is the number of particles in R_*, and $\bar{\theta}_1$, $\bar{\theta}_2$ are the parameter means in R_{*1} and R_{*2} respectively. The tree is finished when each leaf either consists of less than $N_{\min}$ particles, or all of the leaf's particles have the same statistics. To guarantee forest diversity, the root node of each tree (Step 5 in Algorithm 1) is bootstrapped from the reference table, and the statistics considered for splitting each node (Step 9 in Algorithm 1) is limited to a randomly selected subset of size n_{try} among the available statistics S.

Algorithm 1: Growing regression trees for ABC-RF (Raynal et al. 2019)

---

1 Sample $\theta^{(1)},\ldots,\theta^{(N)} \sim \pi(\theta)$
2 Simulate $y^{(1)},\ldots,y^{(N)}$ and compute $s^{(1)},\ldots,s^{(N)}$
3 Form reference table $R = \left\{(\theta^{(1)},s^{(1)}),\ldots,(\theta^{(N)},s^{(N)})\right\}$
4 **for** $t=1,\ldots,B$ **do**
5     $R_0 \leftarrow$ bootstrapped from R
6   Tree T_t is rooted in node R_0
7   **while** $\max_{\text{leaf } R_* \in T_t} |R_*| > N_{\min}$ **do**
8   Find a leaf R_* in T_t with $|R_*| > N_{\min}$
9   Statistic candidates for splitting $R_* \leftarrow$ sample of size n_{try} among S
10   Choose statistic k_* among the candidates and threshold $s_* \in \mathbb{R}$ to split node R_* into new leaves $R_{*1} = \left\{(\theta,s)\in R_* : s_{k_*} \leq s_*\right\}$ and $R_{*2} = \left\{(\theta,s)\in R_* : s_{k_*} > s_*\right\}$, such that the L^2-loss criterion (Eq. (1)) is minimized

---

Given an observation y^{obs}, predicting its parameter θ based on tree T_t involves following from the root node with the statistics s^{obs} and comparing $s_{k_*}^{\text{obs}}$ with s_* at each node R_* to select the next node, until reaching a leaf $L_t(s^{\text{obs}})$. The prediction for θ based on T_t is then the average parameter among particles in $L_t(s^{\text{obs}})$:

$$\frac{\sum_{i=1}^{N}\theta^{(i)}\cdot n_t^{(i)}\cdot \mathbf{1}_{s^{(i)}\in L_t(s^{\text{obs}})}}{\sum_{i=1}^{N} n_t^{(i)}\cdot \mathbf{1}_{s^{(i)}\in L_t(s^{\text{obs}})}}$$

where $n_t^{(i)}$ is the number of times $\theta^{(i)}$ is duplicated in the bootstrapped sample R_0, and $\mathbf{1}_{s^{(i)}\in L_t(s^{\text{obs}})}$ is the indicator for whether $s^{(i)}$ falls into the same leaf as s^{obs}. The prediction for θ based on the whole forest is the average of predictions based on each tree:

$$\sum_{i=1}^{N}\theta^{(i)}\cdot\frac{1}{B}\sum_{t=1}^{B}\frac{n_t^{(i)}\cdot\mathbf{1}_{s^{(i)}\in L_t(s^{\text{obs}})}}{\sum_{i=1}^{N} n_t^{(i)}\cdot\mathbf{1}_{s^{(i)}\in L_t(s^{\text{obs}})}}$$

Raynal et al. (2019) argued that this weighted estimate implies that the density of particles $\theta^{(i)}$ with corresponding weights $w^{(i)}$ (Algorithm 2) forms the approximation $\pi_{ABC-RF}(\theta|s^{\text{obs}})$ for the posterior distribution.

Algorithm 2: Posterior distribution from ABC-RF (Raynal et al. 2019)

---

1 **for** $t=1,\ldots,B$ **do**
2   Follow the tree T_t from root node with s^{obs} until locating leaf $L_t(s^{\text{obs}})$
3 **for** $i=1,\ldots,N$ **do**
4   Weight for particle
    $\theta^{(i)} \leftarrow w^{(i)} = \frac{1}{B}\sum_{t=1}^{B}\frac{n_t^{(i)}\cdot\mathbf{1}_{s^{(i)}\in L_t(s^{\text{obs}})}}{\sum_{i=1}^{N} n_t^{(i)}\cdot\mathbf{1}_{s^{(i)}\in L_t(s^{\text{obs}})}}$

Marin et al. (2022) developed an R package `abcrf` based on the ABC-RF method. For multivariate problems, ABC-RF is usually applied for each parameter marginally (Raynal et al. 2019). The authors also provided suggested values for RF hyperparameters; by default, $N_{\min}=5$ and $n_{\text{try}}=|S|/3$. Among the most important hyperparameters is the count of trees B, for which they recommend analyzing whether the out-of-bag mean squared error stabilizes around the selected B (Raynal et al. 2019; Pudlo et al. 2016).

2.2 Distributional random forests

Ćevid et al. (2022) developed distributional random forests (DRF) for multivariate regression problems. The approach considers two criteria to split tree nodes. The first criterion extends the L^2-loss formula (Eq. 1), which the authors rewrite as

$$\frac{1}{|R_*|}\sum_{(\theta,s)\in R_*}\left(\theta-\bar{\theta}\right)^2 - \frac{|R_{*1}|\cdot|R_{*2}|}{|R_*|^2}\left(\bar{\theta}_2 - \bar{\theta}_1\right)^2$$

where $\bar{\theta}$ is the parameter mean in the parent node R_*. Because the first term does not depend on the split, minimizing the L^2-loss is equivalent to maximizing

$$\frac{|R_{*1}|\cdot|R_{*2}|}{|R_*|^2}\left(\bar{\theta}_2 - \bar{\theta}_1\right)^2 \quad (2)$$

Choosing (k_*,s_*) for multivariate problems then involves maximizing Eq. 2, aggregated across different parameters. Because this condition, denoted as the CART criterion, only considers the difference in mean parameters in the child nodes and not the whole distributions, Ćevid et al. developed a second criterion based on the maximal mean discrepancy (MMD) metric (Ćevid et al. 2022). The MMD defines the difference between distributions of particles in two different sets

by employing a positive-definite kernel and its embedding into a Reproducing Kernel Hilbert Space (RKHS) (Gretton et al. 2007). Due to the computational expense of MMD and the need to compute it for many different statistic and threshold candidates, the authors replace the kernel with its Fourier approximation. By default, they use the Gaussian kernel $k(\theta, \theta') = \exp(-\|\theta - \theta'\|_2^2/2\sigma^2)/(\sigma\sqrt{2\pi})^{|\Theta|}$ with parameter count $|\Theta|$ and bandwidth σ, the median pairwise L^2 distance between the parameters of the particles in the parent node. With Fourier features $\omega_1, \ldots, \omega_L$ randomly chosen from the multivariate Gaussian distribution $\text{Normal}_{|\Theta|}\left(\mathbf{0}, \sigma^{-2}\mathbf{I}_{|\Theta|}\right)$ where $\mathbf{0}$ and $\mathbf{I}_{|\Theta|}$ are square zero and identity matrices of size $|\Theta| \times |\Theta|$, Ćevid et al. derive the approximate MMD as

$$
\mathcal{D}(R_{*1}, R_{*2}) = \frac{1}{L} \sum_{l=1}^{L} \frac{|R_{*1}| \cdot |R_{*2}|}{|R_*|^2} \times
$$
$$
\left| \frac{1}{|R_{*1}|} \sum_{\theta \in R_{*1}} \varphi_{\omega_l}(\theta) - \frac{1}{|R_{*2}|} \sum_{\theta \in R_{*2}} \varphi_{\omega_l}(\theta) \right|^2
\tag{3}
$$

where $\varphi_\omega(u) = e^{i\omega^T u}$ are the approximate kernels. Constructing the random forest using the MMD criterion (Algorithm 3) consists of splitting each node, such that (3) is maximized.

Algorithm 3: Growing regression trees for DRF (Ćevid et al. 2022)

1　Form reference table $R = \left\{\left(\theta^{(1)}, s^{(1)}\right), \ldots, \left(\theta^{(N)}, s^{(N)}\right)\right\}$, similar to Algorithm 1
2　**for** $t = 1, \ldots, B$ **do**
3　　$R_t \leftarrow$ subsample of R of size n_{sample}; R_t is split into R_{t1} and R_{t2}
4　　Tree T_t is rooted in node R_{t1}
5　　**while** $\max_{\text{leaf } R_* \in T_t} |R_*| > N_{\min}$ **do**
6　　　Find a leaf R_* in T_t with $|R_*| > N_{\min}$
7　　　Statistic candidates $\leftarrow$ subsample of S with size $\min(\max(\tilde{n}_{\text{try}}, 1), |S|)$, where $\tilde{n}_{\text{try}} \sim \text{Poisson}(n_{\text{try}})$
8　　　**if** criterion = CART **then**
9　　　　Choose statistic k_* and threshold s_* to split node R_* into new leaves $R_{*1} = \left\{(\theta, s) \in R_* : s_{k_*} \leq s_*\right\}$ and $R_{*2} = \left\{(\theta, s) \in R_* : s_{k_*} > s_*\right\}$, such that the CART criterion (Eq. 2), aggregated across all parameters, is maximized
10　　　**else if** criterion = MMD **then**
11　　　　$\sigma \leftarrow$ median pairwise distance between θ's in R_*
12　　　　Fourier features $\omega_1, \ldots, \omega_L \sim \text{Normal}_{|S|}\left(\mathbf{0}, \sigma^{-2}\mathbf{I}_{|S|}\right)$
13　　　　Choose k_* and s_* to split R_* into R_{*1} and R_{*2}, such that the MMD criterion (Eq. 3), defined with ω_l's, is maximized
14　　**for** $i = 1, \ldots, |R_{t2}|$ **do**
15　　　Follow T_t from root node with $s^{(i)}$ to locate leaf $L_t\left(s^{(i)}\right)$

Once the random forest is constructed, computing the posterior distribution for an observation y^{obs} consists of finding the leaf $L_t\left(s^{\text{obs}}\right)$ in each tree T_t that the observed statistics s^{obs} falls into. Because the samples for tree construction are not bootstrapped, the weight for each particle $\theta^{(i)}$ in the reference table is simply the normalized number of times that it ends up in the same leaf as s^{obs} across the entire forest (Algorithm 4).

The CDF of the joint posterior distribution $\pi_{ABC-DRF}$ $(\theta|s^{\text{obs}})$ is then approximated as

$$
\mathbb{P}\left(\theta_1 \leq x_1, \ldots, \theta_{|\Theta|} \leq x_{|\Theta|}\right) = \sum_{i=1}^{N} w^{(i)} \mathbf{1}_{\theta_1^{(i)} \leq x_1, \ldots, \theta_{|\Theta|}^{(i)} \leq x_{|\Theta|}}
$$

Algorithm 4: Multivariate distribution from DRF (Ćevid et al. 2022)

1　**for** $t = 1, \ldots, B$ **do**
2　　Follow the tree T_t from root node with s^{obs} until locating leaf $L_t(s^{\text{obs}})$
3　**for** $i = 1, \ldots, N$ **do**
4　　Weight for particle $\theta^{(i)} \leftarrow w^{(i)} = \frac{1}{B} \sum_{t=1}^{B} \frac{\mathbf{1}_{L_t(s^i)=L_t(s^{\text{obs}})}}{|L_t(s^{\text{obs}})|}$

Ćevid et al. (2022) developed an R package `drf` based on Algorithms 3 and 4. Compared to `abcrf`, the authors employ different methods to maximize forest diversity (Algorithm 3). Instead of a fixed number of n_{try} randomly selected statistics from which the splitting condition is chosen, the number of candidate statistics in DRF varies between nodes, following $\min(\max(\tilde{n}_{\text{try}}, 1), |S|)$ where $\tilde{n}_{\text{try}}$ is sampled from $\text{Poisson}(n_{\text{try}})$. Furthermore, instead of bootstrapping the reference table, the samples R_{t1} to form each tree's root node are subsampled from R. After the tree is constructed, the particles actually stored in the leaves for computing posterior weights are from a disjoint subsample R_{t2} of R, in order to minimize overfitting. By default, the total number of particles used for tree construction R_{t1} and weights R_{t2} is $N/2$.

In the next section we describe how DRF may be adapted for use in approximate Bayesian computation.

3 ABC-DRF

The main limitation of `abcrf` is that it is typically used for inference of one parameter at a time, resulting in difficulty studying the *joint* posterior of the elements of θ. We noted in Dinh et al. (2024) that DRF may be exploited for use in ABC by constructing the reference table just as for ABC-RF, and then using `drf` to generate observations from the

joint posterior. We give an illustration of this approach, called ABC-DRF, in this section.

3.1 A hierachical normal mean example

We illustrate ABC-DRF with an adaptation of an example from Raynal et al. (2019) and discussed in Dinh et al. (2024). The model has

$$y_1, y_2, \ldots, y_n \mid \theta_1, \theta_2 \sim \text{Normal}(\theta_1, \theta_2)$$
$$\theta_1 \mid \theta_2 \sim \text{Normal}(0, \theta_2)$$
$$\theta_2 \sim \text{IG}(\alpha, \beta) \tag{4}$$

where $\text{IG}(\alpha, \beta)$ denotes the inverse gamma distribution with shape α and rate β and density

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} (1/x)^{\alpha+1} \exp(-\beta/x), \quad x > 0.$$

The observed data y^{obs} consists of n data points $y_1^{\text{obs}}, \ldots, y_n^{\text{obs}}$.

The joint posterior distribution of (θ_1, θ_2) is determined by

$$\theta_2 \mid y^{\text{obs}} \sim \text{IG}\left(\frac{n}{2} + \alpha, \ B\right) \tag{5}$$

$$\theta_1 \mid \theta_2, y^{\text{obs}} \sim \text{Normal}\left(\frac{n \cdot \bar{y}^{\text{obs}}}{n+1}, \frac{2 \cdot \theta_2}{n+1}\right), \tag{6}$$

where

$$B = \frac{1}{2}\left((S^{obs})^2 + 2\beta + \frac{n \cdot (\bar{y}^{obs})^2}{n+1}\right),$$
$$\bar{y}^{obs} = n^{-1} \sum_{i=1}^{n} y_i^{\text{obs}}, \ S^{obs} = \sqrt{\sum_{i=1}^{n} \left(y_i^{\text{obs}} - \bar{y}^{obs}\right)^2}.$$

The marginal posterior of θ_1 is

$$\theta_1 \mid y^{\text{obs}} \sim \frac{n \cdot \bar{y}^{\text{obs}}}{n+1} + \sqrt{\frac{2B}{(n+1)(n+2\alpha)}} \cdot \text{T}_{n+2\alpha}$$

where T_m denotes the t-distribution with m degrees of freedom. θ_1 and θ_2 are uncorrelated under the posterior.

To illustrate the behavior of ABC-DRF, we generated a test set $y^{\text{obs}} = (y_1^{\text{obs}}, \ldots, y_{10}^{\text{obs}})$ from the model, and let $s^{\text{obs}} = S(y^{\text{obs}})$ consist of the 61 summary statistics described in Raynal et al. (2019). The first three statistics $s_1^{\text{obs}}, s_2^{\text{obs}}, s_3^{\text{obs}}$ are the mean, variance and median absolute deviation of y^{obs}. The next eight statistics $s_4^{\text{obs}}, \ldots, s_{11}^{\text{obs}}$ are sums and products of either two or all values of $s_1^{\text{obs}}, s_2^{\text{obs}}, s_3^{\text{obs}}$. The final 50 statistics are noise: $s_{12}^{\text{obs}}, \ldots, s_{61}^{\text{obs}} \sim \text{Uniform}(0, 1)$.

We first infer the posterior distribution from drf, following the example in Dinh et al. (2024). The reference table

consists of $N = 20,000$ entries, each of which results from sampling θ and $y_1, \ldots, y_n$ from (4), then computing s similarly to s^{obs}. The algorithm drf then infers the joint posterior distribution for θ, which we compare against the true posterior distribution.

Figure 1a shows the comparison for $\alpha = 4, \beta = 5$. The posterior distributions from ABC-DRF are in agreement with the true density. Furthermore, the variable importance analysis detects information in $s_1, \ldots, s_{11}$, which indeed contain signals for the distribution of y^{obs} (Figure 1b). In contrast, the pure noise statistics $s_{12}, \ldots, s_{61}$ are deemed unimportant, as expected.

Compared to abcrf, drf's ability to infer the joint distribution is crucial for inference problems where the parameters are known to be dependent. We will experiment with such models in Section 5.

4 ABC sequential Monte Carlo with random forests

Despite their promise as relatively simple ABC approaches, there are some drawbacks in using abcrf and drf in practice. First, their performance typically improves as the forest size increases. The number of splitting criterion computations in growing a forest of B balanced trees from a reference table of size N, where each split considers n_{try} statistics, is $\mathcal{O}(B \cdot n_{\text{try}} \cdot N \cdot \log N)$. For complex problems involving many statistics and requiring large reference tables, the forest can become too computationally expensive to construct. Second, similar to other ABC implementations, the accuracy of the estimated posterior distributions depends on the number of simulations N. However, because the number of splits is approximately $N \log N$ in balanced trees and more in unbalanced trees, the application of abcrf or drf in difficult inference problems where many simulations are necessary can demand more memory than available resources.

To alleviate these problems, we propose integrating the random forest methods within the ABC sequential Monte Carlo (ABC-SMC) framework (Algorithm 5), which was introduced by Toni et al. (2008) as a novel approach to solve the problem of low acceptance rates in the ABC rejection (ABC-REJ) method. ABC-SMC refines the particle population drawn from the prior distribution iteratively, where the posterior distribution from iteration $t-1$ becomes the prior distribution for iteration t. A common choice for the sampling distribution in iteration t is $g_t(\theta) = \sum_{k=1}^{N_{t-1}} w_{t-1}^{*(k)} \cdot K_t(\theta_t^{(k)} \mid \theta_{t-1}^{(k)})$, which samples a random particle from iteration $t-1$ then perturbs it with a Markov kernel $K_t(\theta \mid \theta')$ (Sisson and Fan 2018). The statistics simulated from this new particle are compared against data statistics with a kernel function H_h centered at 0. The kernel scale parameters
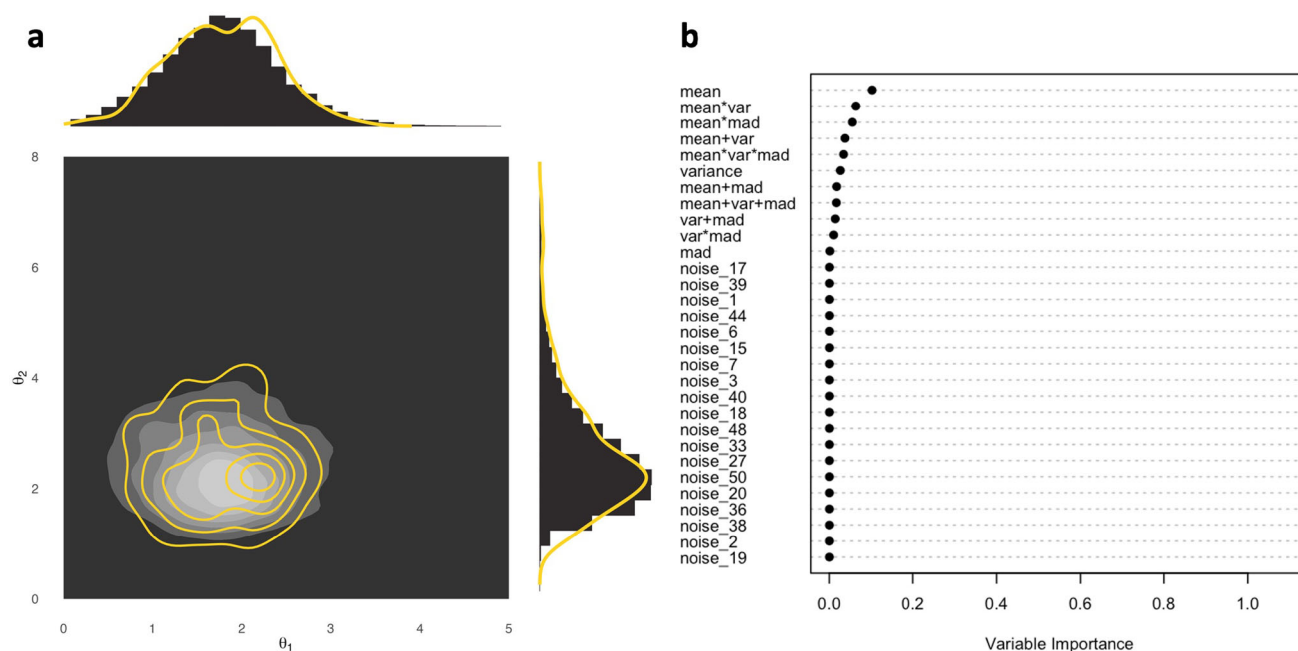
Fig. 1 Inference of $\theta = (\theta_1, \theta_2)$ in the hierarchical model, with $\alpha = 4$, $\beta = 5$. **a**: Joint posterior distributions for θ_1 and θ_2, inferred from ABC-DRF with $N = 20,000$ simulations (yellow contours) against ground truth (density heatmap in gray-scale), with marginal distributions for each parameter from ABC-DRF (yellow line) against ground truth (black histogram). **b**: Variable importance analysis for summary statistics from ABC-DRF, performed with $N = 1,000$ simulations

decrease in successive iterations (i.e., $h_1 \geq \cdots \geq h_T$), hence only particles increasingly closer to s_{obs} are retained. The thresholds $c_2, \ldots, c_T$ further control the acceptance or rejection and resampling of particles in each iteration. The final set of particles $\left\{ \theta_T^{(1)}, \ldots, \theta_T^{(N)} \right\}$ from iteration T is drawn from

$$\pi_{\text{ABC-SMC}}(\theta|s^{\text{obs}}) \propto \int H_{h_T}(\|s_t^{(i)} - s^{\text{obs}}\|) \cdot p(s|\theta) \cdot \pi(\theta) \, ds,$$

where $p(s|\theta)$ is the probability distribution of observed statistics s from the model, given parameter θ (Sisson and Fan 2018).

We develop ABC sequential Monte Carlo with random forests (ABC-SMC-(D)RF), which incorporate `abcrf` and `drf` into ABC-SMC. The first version, ABC-SMC-DRF (Algorithm 6), is applicable for problems with multiple parameters. The first iteration samples the parameter sets directly from the prior distribution, then applies ABC-DRF on the reference table $R \in \mathbb{R}^{N_1 \times (|S|+|\Theta|)}$ that combines all parameter sets and corresponding statistics. In subsequent iterations, parameter sets are drawn from particles in the previous iteration, with weights estimated from ABC-DRF for s^{obs}. The parameter sets are then perturbed with a multivariate Markov kernel, before ABC-DRF is similarly applied. Finally, the ABC-DRF weights are re-calibrated (Step 15, Algorithm 6), similar to ABC-SMC (Step 12, Algorithm 5).

The second version, ABC-SMC-RF (Algorithm 7), is built around ABC-RF, and designed for univariate problems or finding marginal distributions for multivariate models. Each

---

Algorithm 5: ABC-SMC (adapted from Alg. 4.7, page 111, Sisson and Fan 2018)

1 **for** $t = 1, \ldots, T$ **do**
2   **if** $t = 1$ **then**
3   **for** $i = 1, \ldots, N_t$ **do**
4   Generate $\theta_1^{(i)} \sim g_1(\theta)$, an initial sampling distribution
5   Simulate data $y_1^{(i)}$ with $\theta_1^{(i)}$, and compute $s_1^{(i)} = S\left(y_1^{(i)}\right)$
6   Compute weight $w_1^{*(i)} = H_{h_1}\left(\left\|s_1^{(i)} - s^{\text{obs}}\right\|\right) \pi\left(\theta_1^{(i)}\right) \Big/ g_1\left(\theta_1^{(i)}\right)$
7   **else**
8   Construct sampling distribution $g_t(\theta)$ from $\left\{\left(\theta_{t-1}^{(k)}, w_{t-1}^{*(k)}\right)\right\}$
9   **for** $i = 1, \ldots, N_t$ **do**
10   Generate $\theta_t^{(i)} \sim g_t(\theta)$
11   Simulate data $y_t^{(i)}$ with $\theta_t^{(i)}$, and compute $s_t^{(i)} = S\left(y_t^{(i)}\right)$
12   Compute weight $w_t^{(i)} = H_{h_t}\left(\left\|s_t^{(i)} - s^{\text{obs}}\right\|\right) \pi\left(\theta_t^{(i)}\right) \Big/ g_t\left(\theta_t^{(i)}\right)$
13   Reject $\theta_t^{(i)}$ with probability $1 - r_t^{(i)}$ with $r_t^{(i)} = \min\left(1, \frac{w_t^{(i)}}{c_t}\right)$, and go to Step 10
14   Otherwise, accept $\theta_t^{(i)}$ and set modified weight $w_t^{*(i)} = w_t^{(i)}/r_t^{(i)}$

Algorithm 6: ABC-SMC-DRF for multiple parameters

1 **for** $t = 1, \ldots, T$ **do**
2   **for** $i = 1, \ldots, N_t$ **do**
3   **if** $t = 1$ **then**
4   Generate $\theta_1^{(i)} \sim \pi(\theta)$
5   **else**
6   Sample θ^* from $\left\{ \left(\theta_{t-1}^{(k)}, w_{t-1}^{*(k)} \right) \right\}_{k=1,\ldots,N_{t-1}}$
7   Generate $\theta_t^{(i)} \sim K_t(\theta | \theta^*)$
8   If $\pi\left(\theta_t^{(i)}\right) = 0$, then return to Step 6
9   Simulate data $y_t^{(i)}$ with $\theta_t^{(i)}$, and compute $s_t^{(i)} = S\left(y_t^{(i)}\right)$
10   Form $R = \left\{ \left(\theta_t^{(1)}, s_t^{(1)} \right), \ldots, \left(\theta_t^{(N_t)}, s_t^{(N_t)} \right) \right\}$
11   Perform DRF (Algorithms 3 and 4) with reference table R to compute weights $w_t^{(1)}, \ldots, w_t^{(N_t)}$ for observed statistics s^{obs}
12   **if** $t = 1$ **then**
13      $w_t^{*(i)} = w_t^{(i)}$
14   **else**
15      $w_t^{*(i)} = w_t^{(i)} \dfrac{\pi\left(\theta_t^{(i)}\right)}{\sum_{k=1}^{N_{t-1}} w_{t-1}^{*(k)} \cdot K_t\left(\theta_t^{(i)} \big| \theta_{t-1}^{(k)}\right)}$
16   Normalize $\left\{ w_t^{*(i)} \right\}$

Algorithm 7: ABC-SMC-RF for single parameters

1 **for** $t = 1, \ldots, T$ **do**
2   **for** $i = 1, \ldots, N_t$ **do**
3   **if** $t = 1$ **then**
4   Generate $\theta_1^{(i)} \sim \pi(\theta)$
5   **else**
6   **for** $j = 1, \ldots, |\Theta|$ **do**
7   Sample θ_j^* from $\left\{ \left(\theta_{j,t-1}^{(k)}, w_{j,t-1}^{*(k)} \right) \right\}_{k=1,\ldots,N_{t-1}}$
8   Generate $\theta_{j,t}^{(i)} \sim K_{j,t}(\theta | \theta_j^*)$
9   If $\pi_j\left(\theta_{j,t}^{(i)}\right) = 0$, then return to Step 7
10        $\theta_t^{(i)} \leftarrow \left(\theta_{1,t}^{(i)}, \ldots, \theta_{|\Theta|,t}^{(i)} \right)$
11   Simulate data $y_t^{(i)}$ with $\theta_t^{(i)}$, and compute $s_t^{(i)} = S\left(y_t^{(i)}\right)$
12   **for** $j = 1, \ldots, |\Theta|$ **do**
13   Form $R_j = \left\{ \left(\theta_{j,t}^{(1)}, s_t^{(1)} \right), \ldots, \left(\theta_{j,t}^{(N_t)}, s_t^{(N_t)} \right) \right\}$
14   Perform ABC-RF (Algorithms 1 and 2) with reference table R_j to compute weights $w_{j,t}^{(1)}, \ldots, w_{j,t}^{(N_t)}$ for observed statistics s^{obs}
15   **if** $t = 1$ **then**
16        $w_{j,t}^{*(i)} = w_{j,t}^{(i)}$
17   **else**
18        $w_{j,t}^{*(i)} = w_{j,t}^{(i)} \dfrac{\pi_j\left(\theta_{j,t}^{(i)}\right)}{\sum_{k=1}^{N_{t-1}} w_{j,t-1}^{*(k)} \cdot K_{j,t}\left(\theta_{j,t}^{(i)} \big| \theta_{j,t-1}^{(k)}\right)}$
19   Normalize $\left\{ w_{j,t}^{*(i)} \right\}$

---

parameter in an iteration is sampled independently from the previous iteration and then perturbed. The statistics are then computed for the whole parameter set. The reference table consists of these statistics and values corresponding to a given parameter, from which ABC-RF predicts the parameter's marginal distribution based on s^{obs}. The ABC-RF weights are then re-calibrated and normalized.

The kernel values $H_h\left(\|s - s^{\text{obs}}\|\right)$ in ABC-SMC (Algorithm 5) increase as $s \rightarrow s^{\text{obs}}$. In the ABC-SMC-(D)RF algorithms, these kernels are replaced by the weights from ABC-DRF in Algorithm 6 and ABC-RF in Algorithm 7. These weights behave similarly to the ABC-SMC kernels: particles with statistics s closer to s^{obs} fall into the same leaves as the data more often across the forest, therefore the corresponding parameters θ are more heavily weighted. Furthermore, in ABC-SMC, smaller scale parameters h_t reduce the kernel diffusion, that is, only particles with $\|s - s^{\text{obs}}\|$ increasingly close to 0 have high weights. Decreasing the minimum node sizes N_{min} in ABC-(D)RF has the same effect, as it reduces the accepted particles across the forest to only those with statistics very close to those of the data. Finally, ABC-SMC-(D)RF does not reject particles before (D)RF is performed in each iteration except for those outside the prior distribution, and hence is analogous to ABC-SMC with rejection thresholds $c_2 = \cdots = c_T = 0$.

We have developed an R package abcsmcrf based on Algorithms 6 and 7. It offers several practical advantages when compared to direct applications of abcrf and drf. First, constructing iterative random forests of size B from ref-

erence tables of sizes $N_1, \ldots, N_T$ requires $\mathcal{O}(B \cdot \sum_{t=1}^{T}(N_t \cdot \log N_t))$ splits, assuming balanced trees. This requires less computational resources than growing a random forest based on a reference table of size $N = \sum_{i=1}^{T} N_i$, which contains $\mathcal{O}(B \cdot N \cdot \log N))$ splits. Therefore, with the same total number of model simulations, ABC-SMC-(D)RF requires less memory and computational runtime. Second, similar to the behavior of ABC-SMC compared to ABC-REJ, ABC-SMC-(D)RF is likely to converge to the posterior distribution faster than ABC-RF or ABC-DRF, because the parameter distributions are constantly updated to focus on regions in the parameter space Θ that best explain s^{obs}. We will test this with some examples in the next section. Finally, unlike ABC-RF and ABC-DRF which require the full reference table before forest construction and parameter prediction, the approximated posterior distribution from each iteration of ABC-SMC-(D)RF can be compared against the previous iteration to assess whether it has converged, potentially further lowering computational expense.

5 Results

In this section, we compare the performance of ABC-SMC-(D)RF (Algorithm 6 or 7) across different infer-
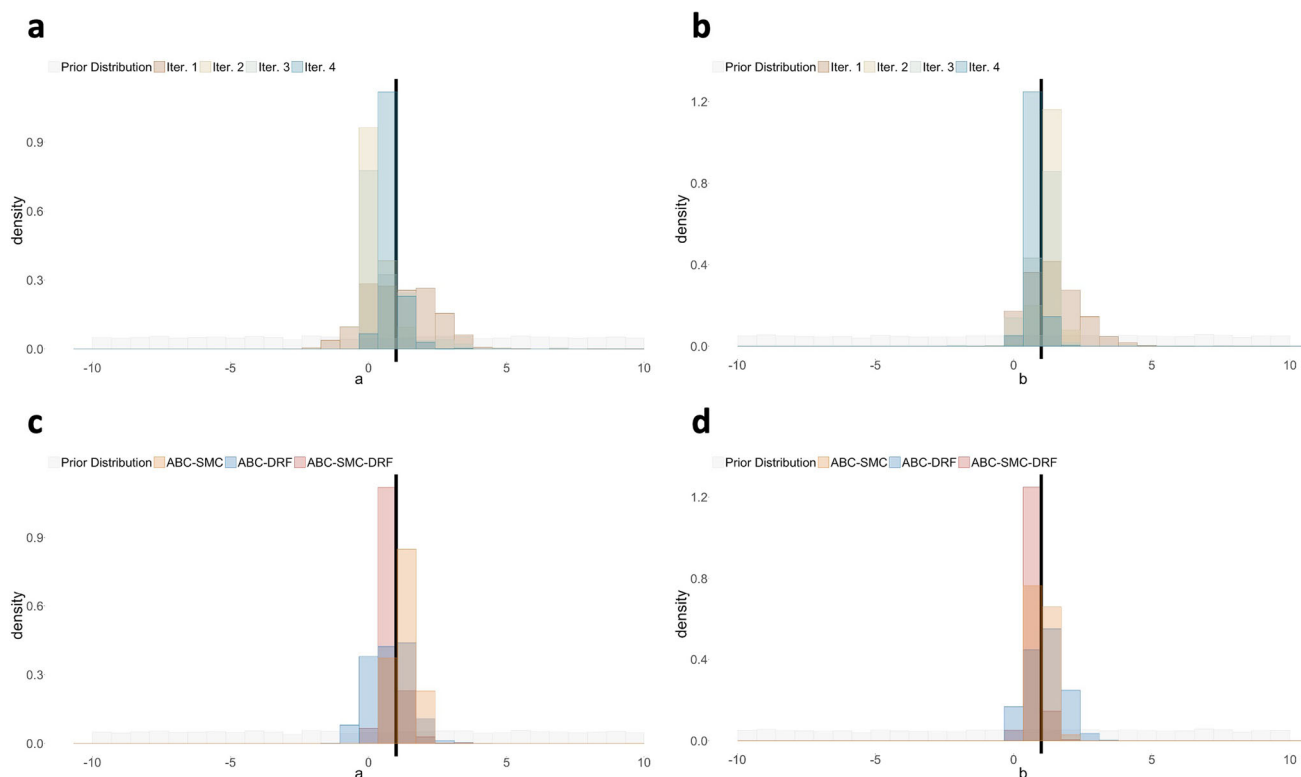
a

b

c

d

Fig. 2 Parameter inference for the deterministic Lotka-Volterra model. **a, b**: Iterative posterior distributions for reaction rates a (**a**) and b (**b**) from ABC-SMC-RF. **c, d**: Marginal posterior distributions for a (**c**) and b (**d**), from ABC-SMC, ABC-DRF and ABC-SMC-DRF. Black vertical lines denote true parameter values

ence problems, against ABC-SMC (Algorithm 5), ABC-RF (Algorithms 1, 2) and ABC-DRF (Algorithms 3, 4). Unless specified otherwise, for ABC-RF, ABC-DRF and ABC-SMC-(D)RF, we use the default parameters in abcrf and drf for tree count B, candidate statistic count n_{try}, leaf size threshold $N_{\min}$, root node subsample size n_{sample}, etc.

5.1 Deterministic Lotka-Volterra model

The Lotka-Volterra model describes the interaction dynamics between predators and prey (Lotka 1925; Volterra 1928). The deterministic model for the number of prey x and predators y is in the form of paired nonlinear differential equations:

$$\frac{dx}{dt} = ax - cxy$$
$$\frac{dy}{dt} = bxy - dy$$

where a is the prey's birth rate, d is the predator's death rate, and c and b are the prey's death rate and predator's birth rate due to predation, respectively. We fix $c = d = 1$ and seek to infer $\theta = (a, b)$ from prior distribution $a, b \sim$ Uniform$(-10, 10)$. Similar to Toni et al. (2008), we solve the ODE system with initial condition $(x(0), y(0)) = (1, 0.5)$

with $\theta = (1, 1)$, and sample (x, y) at eight time points between $t = 0$ and $t = 15$, then add Normal$(0, 0.5^2)$ noise to each data point to form the observed statistics $s^{\text{obs}} = \{x_1, y_1, \ldots, x_8, y_8\}$.

We first implement ABC-SMC-(D)RF for multiple parameters. Algorithm 6 is applied with $T = 4$ iterations, each with $N_t = 5,000$ simulations in the reference table, and perturbation kernels $K_t(\theta|\theta^*) = \theta^* + $ Uniform$(-0.1, 0.1)$. Figures 2a, b show the distributions of a and b after each iteration t in abcsmcrf. The distributions become more concentrated as t progresses, but stay centered around the true values of (a, b) (Table 1).

To evaluate abcsmcrf's performance, we compare the final posterior distributions with ABC-DRF and ABC-SMC. Algorithm drf is performed on a reference table of size $N = 20,000$, to match the total number of simulations in abcsmcrf. We use the R package EasyABC (Jabot et al. 2023) to implement ABC-SMC with $T = 5$ iterations. EasyABC uses the distance $d(s, s^{\text{obs}}) = \sum_{i=1}^{8} \left[(x_i - x_i^{\text{obs}})^2 + (y_i - y_i^{\text{obs}})^2 \right]$ to compare simulated statistics s against observation s^{obs}, and adaptive perturbation kernels $K_t(\theta|\theta') = $ Normal$\left(\theta', 2 \cdot \text{var}(\theta_{t-1}) \right)$ as proposed in Beaumont et al. (2009). Referencing Toni et al. (2008), we impose the series of tolerance thresholds $\epsilon_1 = 30.0, \epsilon_2 = $

Table 1 Means and variances of marginal posterior distributions for the deterministic Lotka-Volterra model from ABC-SMC-DRF, ABC-DRF, and ABC-SMC. The best result for each statistic across different algorithms is in bold ($\mathbb{E}(a)$, $\mathbb{E}(b)$ closest to true values $(a, b) = (1, 1)$, and lowest Var(a) and Var(b))

Statistics	ABC-SMC-DRF iterations				ABC-DRF	ABC-SMC
	1	2	3	4 (Final posterior)		
$\mathbb{E}(a)$	1.1699	0.2772	0.6376	**0.9177**	0.7727	1.3174
Var(a)	1.5548	0.1618	0.7091	**0.0941**	0.5470	0.1177
$\mathbb{E}(b)$	1.4605	1.3320	1.0431	0.8603	1.1808	**1.0178**
Var(b)	1.0441	0.0806	0.1920	**0.0468**	0.3951	0.1154

$16.0, \epsilon_3 = 6.0, \epsilon_4 = 5.0, \epsilon_5 = 4.3$. To produce 1,000 accepted particles, `EasyABC` requires $N = 56,850$ simulations.

Figures 2c, d present the posterior distributions approximated by the ABC implementations. The marginal distributions from ABC-SMC-DRF are as centered and concentrated around the true values for (a, b) as ABC-SMC (Table 1). In contrast, even though the distributions from DRF are centered around the ground truth, the variances are consistently higher, indicating higher uncertainty.

5.2 Stochastic biochemical reaction systems

We illustrate ABC-SMC-DRF's application in inferring reaction rates for biochemical systems.

We first examine the Michaelis-Menten model, which describes the kinetics between an enzyme E that binds to a substrate S to form a complex ES, leading to product P:

$$ES \rightarrow E + P \tag{7}$$
$$E + S \rightarrow ES \tag{8}$$
$$ES \rightarrow E + S \tag{9}$$
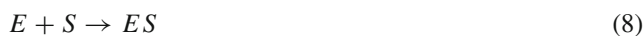
The state of the system at time t can be described as $X(t) = [E(t), S(t), ES(t), P(t)]$, the counts of each type of molecule. The time for the next reaction (7), (8) and (9) to occur is exponentially distributed, with rates defined by their propensity functions:

$$\alpha_1(X(t)) = \bar{c}_1 \cdot ES(t); \quad \bar{c}_1 = c_1$$
$$\alpha_2(X(t)) = \bar{c}_2 \cdot E(t) \cdot S(t); \quad \bar{c}_2 = \frac{10^{c_2}}{\text{nA} \cdot \text{vol}}$$
$$\alpha_3(X(t)) = \bar{c}_3 \cdot ES(t); \quad \bar{c}_3 = 10^{c_3}$$

respectively, where the Avagadro's constant nA $= 6.023 \cdot 10^{23}$ approximates the number of molecules in a mole, vol $= 10^{-15}$ is the volume of the system, and $\theta = (c_1, c_2, c_3)$ parameterizes the Michaelis-Menten model for different biochemical systems (Wilkinson 2018, Example 7.3).

The initial state is defined by $E(0) = 2 \cdot 10^{-7} \cdot \text{nA} \cdot \text{vol}$, $S(0) = 5 \cdot 10^{-7} \cdot \text{nA} \cdot \text{vol}$, and $ES(0) = P(0) = 0$ (Wilkinson 2018). We seek to infer θ from $s^{\text{obs}} = \{X(t), t = 1, \ldots, 10\}$, simulated with true parameters $c_1 = 0.1, c_2 = 6, c_3 = -4$. We compare the results from `drf` with $N = 20,000$ simulations, and `abcsmcrf` with $T = 5$ iterations, each with reference tables of size $N_t = 4,000$, from uniform prior distributions $\pi(c_1) = \text{U}(0, 1)$, $\pi(c_2) = \text{U}(5, 7)$, $\pi(c_3) = \text{U}(-5, -3)$. The perturbation kernels in `abcsmcrf` in each iteration are $K_t(c_1|c_1') = c_1' + \text{U}(-0.05, 0.05)$, $K_t(c_2|c_2') = c_2' + \text{U}(-0.1, 0.1)$, and $K_t(c_3|c_3') = c_3' + \text{U}(-0.1, 0.1)$. All simulations are generated by applying Gillespie's algorithm (Gillespie 1977).

Comparing the results for c_1 and c_2 shows that ABC-SMC-DRF's posterior distributions (c_1: mean = 0.102, 95% CI = [0.070, 0.138]; c_2: mean = 5.986, 95% CI = [5.886, 6.079]) are centered closely around the true values of $c_1 = 0.1, c_2 = 6$, similarly to ABC-DRF, but with narrower confidence intervals (ABC-DRF c_1: mean = 0.095, 95% CI = [0.026, 0.171]; c_2: mean = 5.963, 95% CI = [5.766, 6.127]) (Figure 3a, b). In contrast, both methods are uncertain about the true distribution for c_3 (ABC-SMC-DRF: mean = -3.882, 95% CI = [$-4.945, -3.040$]; ABC-DRF: mean = -4.011, 95% CI = [$-4.928, -3.102$]) (Figure 3c).

Analyzing the reaction system reveals a likely reason: assuming true values of θ, reactions (7) and (9) occur with rates $\bar{c}_1 = 0.1$ and $\bar{c}_3 = 10^{-4}$, respectively. Because of the big difference in magnitudes, complex ES overwhelmingly undergoes reaction (7), while reaction (9) rarely occurs and does not significantly impact the observed molecule counts.

We apply Morris's global sensitivity analysis (Morris 1991) to study the parameter identifiability (R package `sensitivity` (Iooss et al. 2024; Monari and Strachan 2017)). The method involves computing the elementary effect of each parameter, defined in this case as the change in the statistics divided by the change in the parameter, across different sampling schemes, and analyzing its absolute mean μ^* and standard deviation σ. The study shows that c_1 ($\mu^* = 0.019, \sigma = 0.012$) and c_2 ($\mu^* = 0.018, \sigma = 0.002$) have significantly higher absolute mean elementary effects than c_3 ($\mu^* = 0.001, \sigma = 0.002$) (Figure 3d). This implies

that varying c_3 has little to no effect on the observed statistics, consistent with earlier findings (Degasperi and Gilmore 2008).

Indeed, despite the uncertainty in c_3, simulations performed with θ values sampled from the posterior distributions of ABC-DRF and ABC-SMC-DRF are centered around the observed values s^{obs} across different molecules and time points (Figure 3e, f, g, h). Moreover, consistent with ABC-SMC-DRF's higher certainty in the inference for c_1 and c_2, molecule counts simulated from its posterior distribution have significantly reduced range compared to ABC-DRF.

5.3 Stochastic birth-death models

5.3.1 Linear birth-death process

We next examine the performance of ABC-SMC-DRF for the homogeneous linear birth-death process illustrated in Tavaré (2018), for which exact results may be found. The model starts at time 0 with $Z(0) = z_0$ individuals, each of which divides at rate $\lambda > 0$ and dies at rate $\mu > 0$, independently for each individual. The number of individuals at time $t > 0$ is denoted $Z(t)$, with expected value $\mathbb{E}[Z(t)] = Z(0)e^{(\lambda-\mu)t}$.

The probability $\mathbb{P}(Z(t) = j \mid Z(0) = i)$ that there are j individuals at time t, given that there are i individuals at time 0, follows from Keiding (1975) as

$$
\begin{aligned}
&\mathbb{P}(Z(t) = j \mid Z(0) = i) \\
&= \sum_{l=0}^{\min(i,j)} \binom{i}{l} (1 - \alpha(t))^l \alpha(t)^{i-l} \binom{j-1}{i-l} \times \\
&\quad (1 - \beta(t))^l \beta(t)^{j-l}, \, j = 1, 2, \ldots
\end{aligned}
\tag{10}
$$

and

$$
\mathbb{P}(Z(t) = 0 \mid Z(0) = i) = \alpha(t)^i,
\tag{11}
$$

where $\alpha(t)$ and $\beta(t)$ are defined by

$$
\begin{aligned}
\alpha(t) &= \frac{\mu\left(e^{(\lambda-\mu)t} - 1\right)}{\lambda e^{(\lambda-\mu)t} - \mu}, \quad \beta(t) = \frac{\lambda}{\mu}\alpha(t), \quad \text{if } \lambda \neq \mu; \\
\alpha(t) &= \beta(t) = \frac{\lambda t}{1 + \lambda t}, \quad \text{if } \lambda = \mu,
\end{aligned}
$$

Kendall (1948). The Markov property implies that the likelihood for λ and μ, given an observation $y^{obs} = (Z(t_1) = z_1, \ldots, Z(t_n) = z_n)$ with $z_n > 0$, is

$$
f\left(\lambda, \mu \middle| y^{obs}\right) = \frac{\prod_{i=0}^{n-1} \mathbb{P}(Z(t_{i+1}) = z_{i+1} \mid Z(t_i) = z_i)}{\mathbb{P}(Z(t_n) > 0 \mid Z(0) = z_0)},
\tag{12}
$$

which may be calculated from (10) and (11). The observations used in the example below are generated on a grid of time points $t_1 = 1/25$, $t_2 = 2/25$, ..., $t_{25} = 1$, and $z_0 = 10$.

Instead of approximating the posterior distribution of $\theta = (\lambda, \mu)$, we approximate that of the extinction parameter μ/λ and the Malthusian parameter $\lambda - \mu$, so that $\theta = (\mu/\lambda, \lambda - \mu)$. The following prior distributions were used:

$$
\begin{aligned}
\lambda - \mu &\sim \text{Gamma}(\text{shape} = 3, \text{rate} = 1) \\
\mu/\lambda &\sim \text{Beta}(5, 5)
\end{aligned}
$$

These priors guarantee that $\lambda > \mu$, implying that the probability of extinction is less than 1. We evaluate ABC-SMC-RF and ABC-SMC-DRF with $T = 4$ iterations, each with $N_t = 5,000$ simulations in the reference table and the random forest in each iteration is constructed from $B = 2,500$ trees. We implement the normal perturbation kernels $K_t(\lambda - \mu | \lambda^* - \mu^*) = \mathcal{N}\left(\lambda^* - \mu^*, 2 \cdot \sigma_{t-1}^{\lambda-\mu}\right)$ and $K_t(\mu/\lambda | \mu^*/\lambda^*) = \mathcal{N}\left(\mu^*/\lambda^*, 2 \cdot \sigma_{t-1}^{\mu/\lambda}\right)$, where $\sigma_{t-1}^{\lambda-\mu}$ and $\sigma_{t-1}^{\mu/\lambda}$ are the empirical variances of the Malthusian and extinction parameters from iteration $t-1$. This perturbation framework was proven to be efficient in the context of ABC-SMC (Beaumont et al. 2009). The values of $\sigma_{t-1}^{\lambda-\mu}$ and $\sigma_{t-1}^{\mu/\lambda}$ are reported in Table 2. Figure 4 present the joint and marginal posterior distributions from ABC-SMC-RF and ABC-SMC-DRF. The results from both methods are close to the true posterior distribution.

5.3.2 Controlled branching process

We also investigate the application of ABC-SMC-(D)RF for a discrete-time controlled branching process developed by González et al. (2022). Let Z_n be the population size at generation $n = 0, 1, \ldots$, which is defined recursively via

$$
\begin{aligned}
Z_0 &= 1 \\
Z_{n+1} &= \sum_{j=1}^{\phi_n(Z_n)} X_{n,j}
\end{aligned}
$$

where $X_{n,j}$ is the number of offspring from the jth individual at generation n and the control variable $\phi_n(z)$ depends on the population size z at time n, and is the number of individuals that can be parents of generation $n + 1$. Here, $\phi_n(z) \sim \text{Bin}(\xi(z), \gamma)$ with $\xi(z) = z + \lfloor \log z \rfloor$. González et al. (2022) modeled $X_{n,j}$ as i.i.d. with respect to n and j, with probabilities $\boldsymbol{p}(\kappa) = \{p_k = \mathbb{P}(X_{n,j} = k), k = 0, 1, \ldots, \kappa\}$ where κ is the maximum offspring capacity per individual. The data is simulated with $\gamma = 0.8$, $\kappa = 4$ and $\boldsymbol{p}(\kappa) \sim \text{Bin}(4, 0.9)$. From data statistics $\{Z_1, \ldots, Z_{10}\}$ and $\phi_9(Z_9)$ (e.g., the number of parents of the last generation),
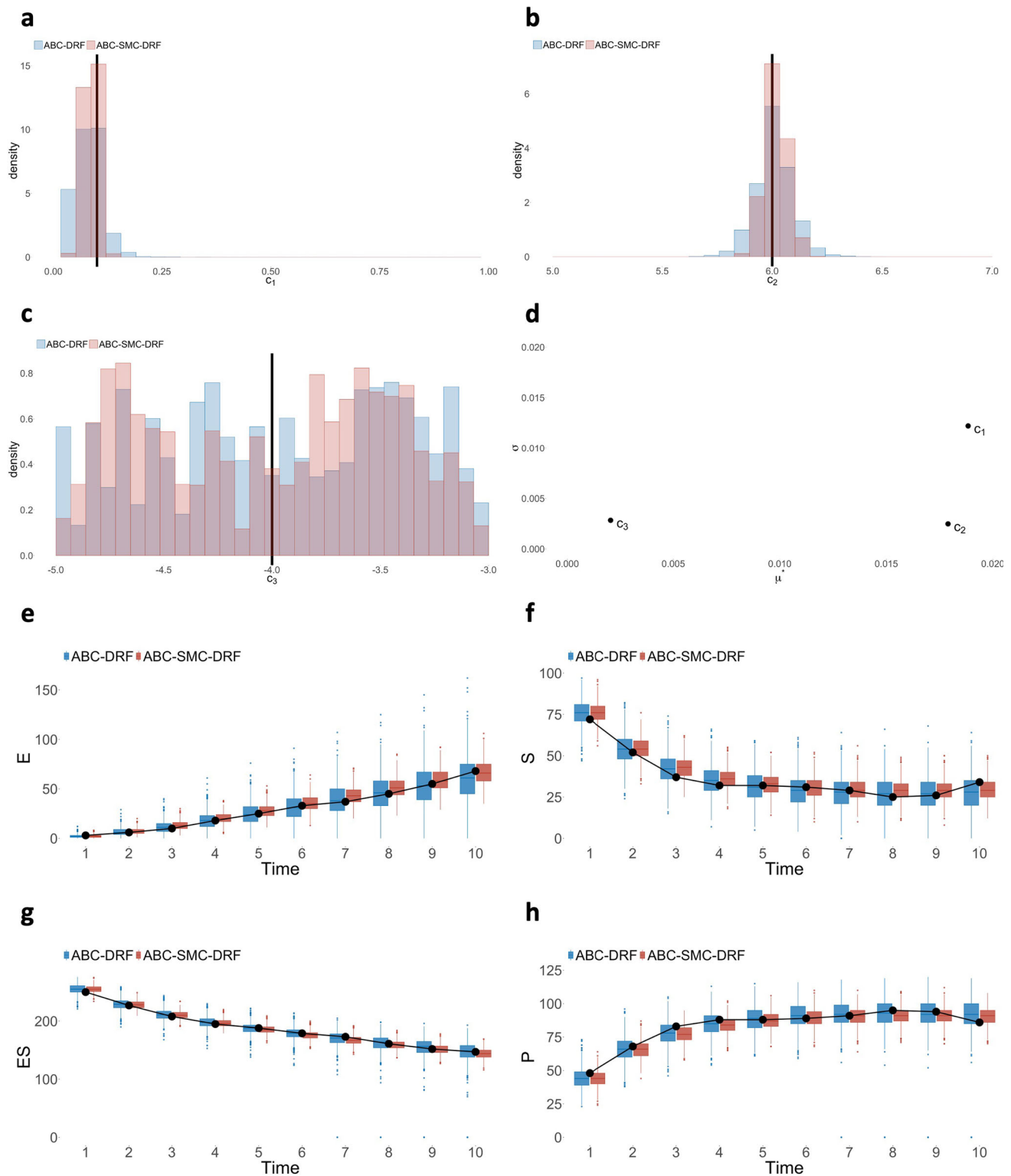
Fig. 3 Parameter inference for the Michaelis-Menten reaction system. **a, b, c**: Marginal distributions for reaction rates c_1 (**a**), c_2 (**b**) and c_3 (**c**), inferred with ABC-DRF and ABC-SMC-DRF, compared against true values (black lines). **d**: Absolute mean (μ^*) and standard deviation (σ) of elementary effects of c_1, c_2 and c_3, using Morris's global sensitivity analysis. **e, f, g, h**: Range of molecule numbers for E (**e**), S (**f**), ES (**g**) and P (**h**) across time, simulated with $\theta \sim \pi_{ABC-DRF}(\theta|s^{\mathrm{obs}})$ and $\pi_{ABC-SMC-DRF}(\theta|s^{\mathrm{obs}})$ (box plots), compared against observed data s^{obs} (black dots and lines)

Table 2 Empirical variances in ABC-SMC-RF and ABC-SMC-DRF across different iterations

Variance	ABC-SMC-RF iterations			ABC-SMC-DRF iterations		
	$t = 2$	$t = 3$	$t = 4$	$t = 2$	$t = 3$	$t = 4$
$2 \cdot \sigma_{t-1}^{\lambda-\mu}$	0.1207	0.1166	0.1514	0.2349	0.2785	0.2187
$2 \cdot \sigma_{t-1}^{\mu/\lambda}$	0.0273	0.0323	0.0343	0.0231	0.0234	0.0209
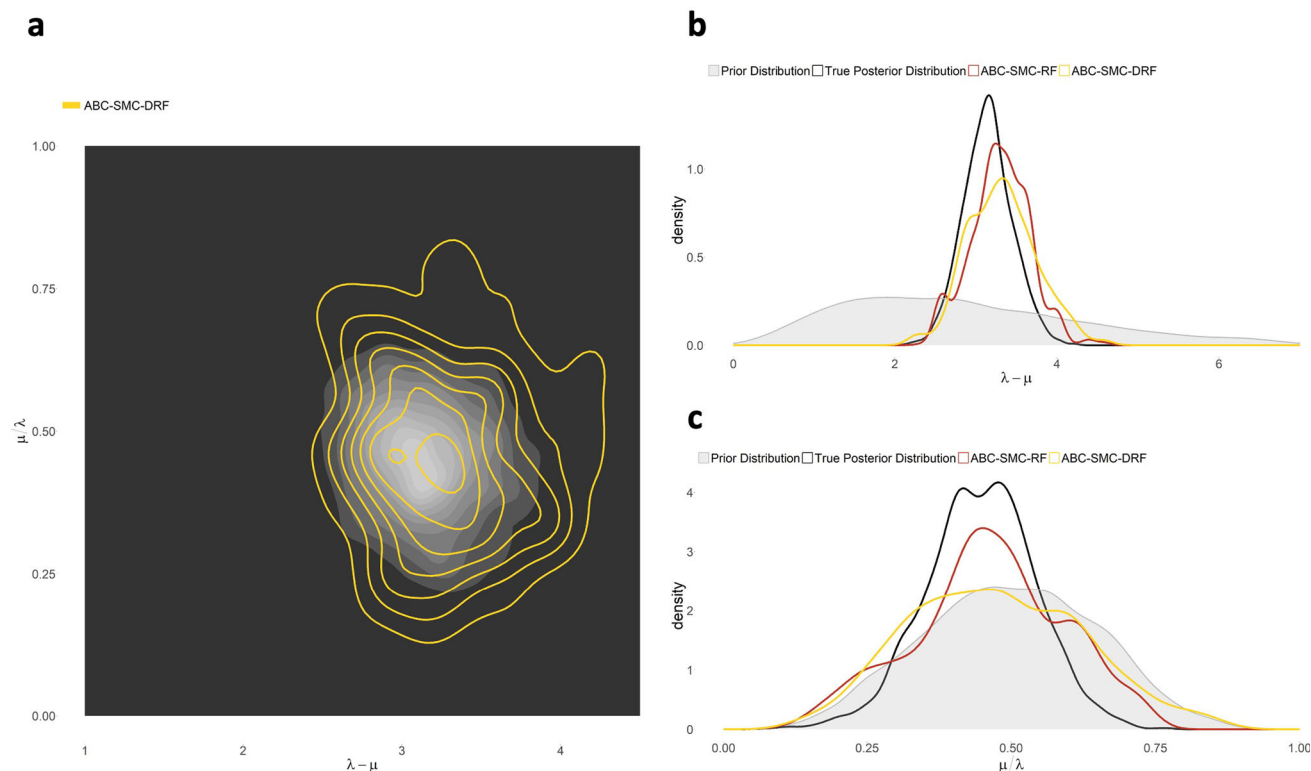
Fig. 4 Parameter inference for the linear birth-death branching process. **a**: Joint posterior distributions for the Malthusian parameter $\lambda - \mu$ and the extinction parameter μ/λ from ABC-SMC-DRF (yellow contours) against ground truth (Eq. 12, density heatmap in gray-scale). **b, c**: Marginal distributions for $\lambda - \mu$ (**b**) and μ/λ (**c**), inferred with ABC-SMC-RF and ABC-SMC-DRF

we seek to find parameters $\theta = \{\gamma, \kappa, \boldsymbol{p}(\kappa)\}$ from prior distributions

$$\gamma \sim \beta(1,1)$$
$$\kappa \sim \mathcal{U}(2, 15)$$
$$\boldsymbol{p}(\kappa) \sim \mathcal{D}(\kappa + 1, \alpha_\kappa = (1, \ldots, 1)).$$

As in González et al. (2022), we analyze the results by examining the mean offspring count $m = \sum_{k=1}^{\kappa} k \cdot p_k(\kappa)$ instead of $\boldsymbol{p}(\kappa)$.

Similar to the linear birth-death process above, we perform ABC-SMC-RF and ABC-SMC-DRF with $T = 4$ iterations, $N_t = 5,000$ simulations per iteration and $B = 2,500$ trees in each random forest. We implement the uniform perturbations $K_t(\kappa|\kappa^*) = \mathcal{U}(\kappa^* - 3, \kappa^* + 3)$, $K_t(\gamma|\gamma^*) = \mathcal{U}(\gamma^* - 0.05, \gamma^* + 0.05)$ and $K_t(p_j|p_j^*) = \mathcal{U}(p_j^* - 0.05, p_j^* + 0.05)$. Figure 5 shows the comparisons between ABC-SMC-(D)RF

posterior distributions and the true parameter values $\kappa = 4$, $m = 3.6$ and $\gamma = 0.8$.

Figure 5 shows the inferred marginal posterior distributions of κ (Figure 5b), m (Figure 5c), γ (Figure 5d), as well as the joint posterior distribution of (γ, m) from ABC-SMC-DRF (Figure 5a). The marginal distributions for m and γ from both ABC-SMC-RF and ABC-SMC-DRF are centered around the true values. However, the ABC-SMC-RF distribution for m is more concentrated compared to ABC-SMC-DRF, and the reverse holds for γ. Table 3 also shows that the mean and median of each parameter distribution converge to the true values after successive ABC-SMC-DRF iterations, with decreasing variances. This results in good approximations for m and γ, with relative errors $< 10\%$ for both parameters in the final iteration.

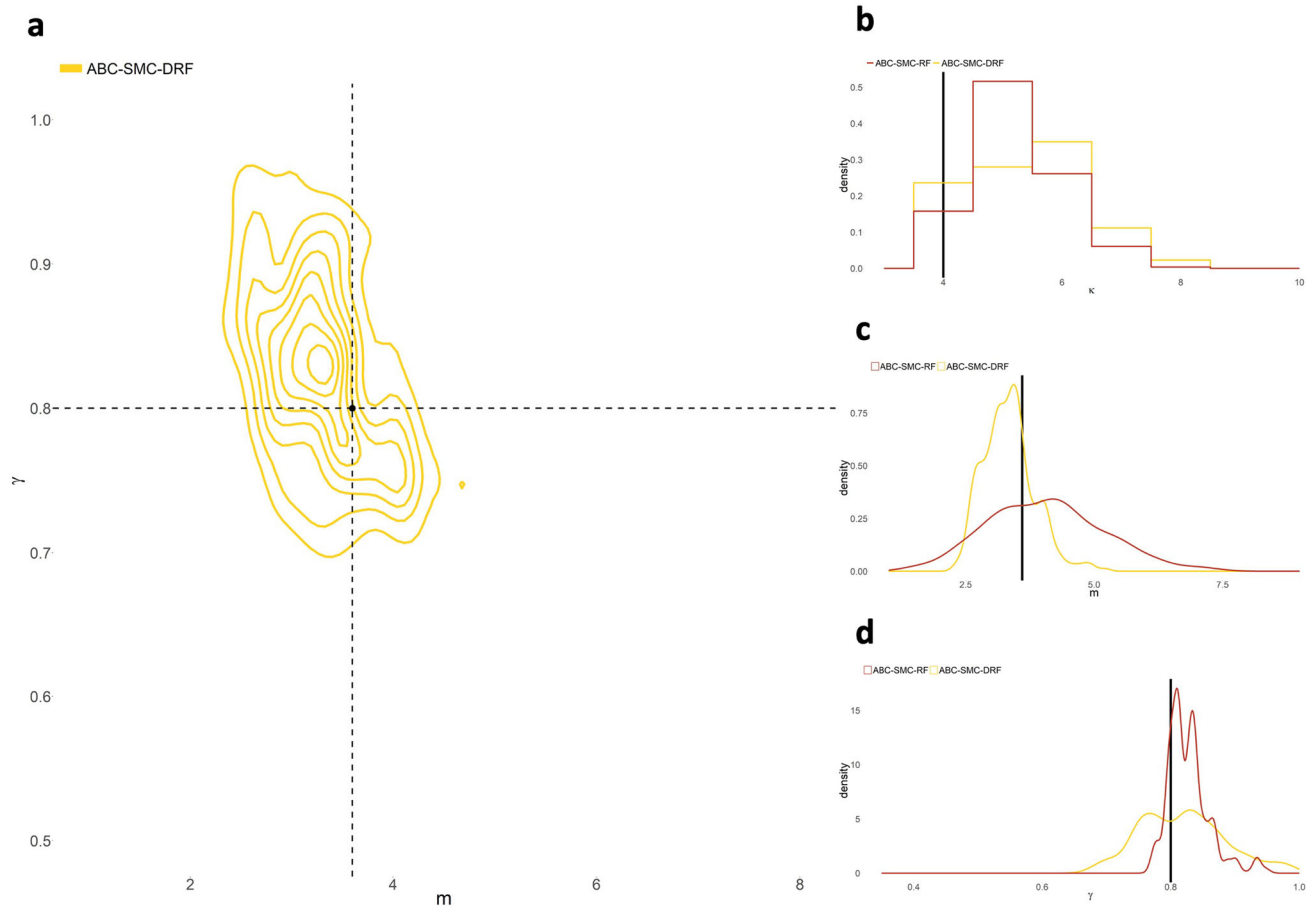Both ABC-SMC-RF and ABC-SMC-DRF overestimate κ (Figure 5b). However, these distributions are similar to

Fig. 5 Parameter inference for the controlled branching process. **a**: Joint posterior distributions for m and γ from ABC-SMC-DRF (yellow contours) against true values (broken lines). **b, c, d**: Marginal distri-butions for κ (**b**), m (**c**) and γ (**d**), inferred with ABC-SMC-RF and ABC-SMC-DRF, compared to true values (black lines)

the results reported by González et al. (2022) (Figure 1 in González et al. 2022). As noted by the authors, the model with maximum offspring count κ behaves identically to one with $\kappa' > \kappa$ where $p_{\kappa+1} = \cdots = p_{\kappa'} = 0$. Therefore, κ can be inferred to be greater than the true value $\bar{\kappa}$, as long as the probabilities p_j with $j > \bar{\kappa}$ are negligible (e.g., $< 10^{-3}$). We note that González et al. (2022) utilizes a multi-step inference framework, where κ is estimated first with ABC-SMC and (γ, m) are then inferred with ABC-rejection, conditioned on the selected κ. In comparison, our framework of using ABC-SMC-(D)RF to infer κ, m and γ simultaneously may be more practical for implementation.

There are multiple sources of variability in the ABC-SMC-(D)RF results, including both the stochastic simulations that form the reference table in each iteration and the construction of the random forests. Therefore, we perform the ABC-SMC-(D)RF inference for the controlled branching process 500 times and analyze the changes in the inferred distributions (Table 4). Overall, there is some evidence that ABC-SMC-DRF underestimates m and overestimates γ; the true values

of both $m = 3.6$ and $\gamma = 0.8$ fall outside of their respective 95% interquantile ranges, $[q_{2.5\%}, q_{97.5\%}]$. In comparison, the interquantile ranges from ABC-SMC-RF successfully contain the true values, albeit with higher variance for m. This agrees with the results from the single inference run shown in Figure 5.

5.3.3 A logistic branching process

Our final branching example concerns the logistic branching process discussed by Lambert (2005). This has state space $\{0, 1, 2, \ldots\}$ and non-zero transition rates given by

$$n \to n + 1 \text{ at rate } n\lambda;$$
$$n \to n - 1 \text{ at rate } n(\mu + (n-1)c), \tag{13}$$

for non-negative parameters λ, μ, c. The case $c = 0$ corresponds to the model discussed in Section 5.3.1, and we assume henceforth that $c > 0$. Lambert showed that there are two dichotomous behaviors in this model, depending on

Table 3 Means, medians and variances of the marginal posterior distributions for the controlled branching process from ABC-SMC-DRF

Statistics	ABC-SMC-DRF iterations			
	1	2	3	4 (Final posterior)
$\mathbb{E}(\kappa)$	6.326	5.478	5.239	5.406
$q_{50\%}(\kappa)$	6	5	5	5
$\text{Var}(\kappa)$	3.297	1.900	1.332	1.076
$\mathbb{E}(\gamma)$	0.833	0.856	0.854	0.823
$q_{50\%}(\gamma)$	0.873	0.872	0.869	0.825
$\text{Var}(\gamma)$	0.013	0.007	0.006	0.005
$\mathbb{E}(m)$	3.368	3.225	3.258	3.353
$q_{50\%}(m)$	3.106	3.137	3.165	3.332
$\text{Var}(m)$	0.500	0.243	0.234	0.259

whether $\mu = 0$ or $\mu > 0$. When $\mu = 0$ the population size $Z(\cdot)$ remains positive and converges in distribution to a Poisson random variable P with mean λ/c, conditional on $P > 0$. When $\mu > 0$ the population goes extinct with probability 1. If we observe the population size at time points $t_1 < t_2 < \ldots < t_n$ and it happens that $Z(t_n) > 0$, we might ask whether there is evidence that $\mu = 0$ or $\mu > 0$.

This question can be addressed in principle using `abcrf` and `drf` approaches, by assuming priors for μ and c (and, w.l.o.g., $\lambda = 1$), generating a reference set and proceeding as earlier. One object of interest is the posterior probability that $\mu = 0$, which gives information about the likelihood the population will survive.

Rather than detail the results here, we note that this example seems to be a case where the SMC version of the random forest methods is not feasible. This is because the algorithms successively sample new particles from the previous iteration and perturb them with Markov kernels, assuming smooth transitions. The perturbed particles in iterations $t \geq 2$ therefore include $\mu > 0$ with probability one, hence ABC-SMC-(D)RF is incapable of accurately evaluating the probability that $\mu = 0$.

6 Discussion

In this paper, we introduce a new Bayesian inference method, Approximate Bayesian Computation sequential Monte Carlo

with random forests (ABC-SMC-(D)RF). It inherits random forest's non-parametric nature, resulting in less dependence on user-defined arguments compared to traditional ABC methods such as ABC-REJ, MCMC and ABC-SMC.

The random forest is embedded in a sequential Monte Carlo regime that progressively updates the parameter distributions to focus on regions in the parameter space with higher likelihood. We perform numerical experiments for deterministic and stochastic models in ecology, population genetics and systems biology, and observe that ABC-SMC-(D)RF typically results in better posterior approximations than previous RF methods. The Lotka-Volterra example illustrates that generic ABC-SMC-RF produces results comparable to traditional ABC implementations with optimized arguments.

However, there are several areas of improvement that can extend ABC-SMC-(D)RF's capabilities. First, a hyper-parameter in ABC-SMC-(D)RF that potentially has an impact on the results is the choice of perturbation kernels $K_t(\theta|\theta')$, which should balance between exploring the parameter space and targeting the regions already found to contain high likelihood from previous iterations. In ABC-MCMC and ABC-SMC, it is common to choose uniform or Gaussian kernels, but the optimal kernel form and parameterization for specific problems may be complicated zdrovandi2011estimation,lee2012choice. A typical approach is to employ kernels that are adjusted dynamically depending on the performance of the previous iteration (Beaumont et al. 2009; Del Moral et al. 2012; Liu et al. 2000; Filippi et al. 2013; Atchadé and Fort 2010). By default, ABC-SMC-(D)RF implements Gaussian kernels $K_t(\theta|\theta') = $ Normal$(\theta', 2 \cdot \text{var}(\theta_{t-1}))$, proposed by Beaumont et al. (2009) for ABC-SMC. Further work might evaluate other adaptive kernel frameworks.

Second, it is sometimes not necessary to continue iteration in ABC-SMC-(D)RF, if further computation is unlikely to yield significant improvements in the posterior distributions. There have been a number of papers determining the stopping criterion in ABC-SMC (for example, Prangle (2017)), and their incorporation in ABC-SMC-(D)RF promises to lower the computational cost while retaining accuracy.

Finally, in this paper we have focused on parameter inference. ABC-RF can perform model selection (Pudlo et al. 2016), and the same framework could potentially be implemented in ABC-DRF and ABC-SMC-(D)RF. This will be

Table 4 Mean ($\mathbb{E}$), 2.5^{th} and 97.5^{th} percentiles ($[q_{2.5\%}, q_{97.5\%}]$), and standard deviation (std) of ABC-SMC-(D)RF posterior marginal distribution medians for the controlled branching process across 500 runs. The data statistics are the same across different inference runs

Method	κ			γ			m		
	$\mathbb{E}$	$[q_{2.5\%}, q_{97.5\%}]$	std	$\mathbb{E}$	$[q_{2.5\%}, q_{97.5\%}]$	std	$\mathbb{E}$	$[q_{2.5\%}, q_{97.5\%}]$	std
ABC-SMC-RF	5.8	[5, 7]	0.6	0.817	[0.795, 0.841]	0.012	4.03	[3.05, 5.05]	0.53
ABC-SMC-DRF	5.3	[5, 6]	0.5	0.864	[0.832, 0.902]	0.020	3.34	[3.15, 3.56]	0.10

relevant for problems similar to the logistic branching process in Section 5.3.3, where there is interest in model selection. ABC-SMC-(D)RF is designed as a wrapper around `abcrf` (Marin et al. 2022) and `drf` (Michel and Ćevid 2021) and can therefore be updated together with the original libraries.

Declarations

Competing interests The authors declare no competing interests.

References

Atchadé, Y., Fort, G.: Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. Bernoulli **16**, 116–154 (2010)

Beaumont, M., Zhang, W., Balding, D.: Approximate Bayesian computation in population genetics. Genetics **162**, 2025–2035 (2002)

Beaumont, M.A., Cornuet, J.-M., Marin, J.-M., Robert, C.P.: Adaptive approximate Bayesian computation. Biometrika **96**, 983–990 (2009)

Breiman, L.: Random forests. Mach. Learn. **45**, 5–32 (2001)

Ćevid, D., Michel, L., Näf, J., Bühlmann, P., Meinshausen, N.: Distributional random forests: Heterogeneity adjustment and multivariate distributional regression. J. Mach. Learn. Res. **23**, 1–79 (2022)

Degasperi, A., Gilmore, S.: Sensitivity analysis of stochastic models of bistable biochemical reactions. In: Bernardo, M., Degano, P., Zavattaro, G. (eds.) Formal Methods for Computational Systems

Biology. volume 5016, pp. 1–20. Springer-Verlag, Berlin, Heidelberg (2008)

Del Moral, P., Doucet, A., Jasra, A.: An adaptive sequential Monte Carlo method for approximate Bayesian computation. Stat. Comput. **22**, 1009–1020 (2012)

Desai, S., Ouarda, T.B.: Regional hydrological frequency analysis at ungauged sites with random forest regression. J. Hydrol. **594**, 125861 (2021)

Dinh, K.N., Tavaré, S., Zhang, Z.: Irving institute for cancer dynamics, (2024). URL https://cancerdynamics.columbia.edu/news/approximate-bayesian-computation-and-distributional-random-forests. Accessed on February 26, 2024

Drovandi, C.C., Pettitt, A.N.: Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. Biometrics **67**, 225–233 (2011)

Filippi, S., Barnes, C.P., Cornebise, J., Stumpf, M.P.: On optimality of kernels for approximate Bayesian computation using sequential Monte Carlo. Statistical Applications in Genetics and Molecular Biology **12**, 87–107 (2013)

Fu, Y.-X., Li, W.-H.: Estimating the age of the common ancestor of a sample of DNA sequences. Mol. Biol. Evol. **14**, 195–199 (1997)

Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. **81**, 2340–2361 (1977)

González, M., Minuesa, C., del Puerto, I.: Approximate Bayesian computation approach on the maximal offspring and parameters in controlled branching processes. Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas **116**(1), 147 (2022). https://doi.org/10.1007/s13398-022-01290-w

Gretton, A., Borgwardt, K.M., Rasch, M., Schölkopf, B., Smola, A.J.: A kernel method for the two-sample problem. In Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference. The MIT Press, (2007)

Iooss, B., Veiga, S.D., Janon, A., Pujol, G.: with contributions from Baptiste Broto, Boumhaout, K., Clouvel, L., Delage, T., Amri, R.E., Fruth, J., Gilquin, L., Guillaume, J., Herin, M., Idrissi, M.I., Le Gratiet, L., Lemaitre, P., Marrel, A., Meynaoui, A., Nelson, B.L., Monari, F., Oomen, R., Rakovec, O., Ramos, B., Rochet, P., Roustant, O., Sarazin, G., Song, E., Staum, J., Sueur, R., Touati, T., Verges, V., Weber, F.: sensitivity: Global Sensitivity Analysis of Model Outputs and Importance Measures, (2024). URL https://CRAN.R-project.org/package=sensitivity. R package version 1.30.0

Jabot, F., Faure, T., Dumoulin, N., Albert, C.: EasyABC: Efficient Approximate Bayesian Computation Sampling Schemes, (2023). URL https://CRAN.R-project.org/package=EasyABC. R package version 1.5.2

Jung, H., Marjoram, P.: Choice of summary statistic weights in Approximate Bayesian Computation. Statistical Applications in Genetics and Molecular Biology, 10:art. 45, (2011)

Keiding, N.: Maximum likelihood estimation in the birth-and-death process. Ann. Stat. **3**, 363–372 (1975)

Kendall, D.G.: On the generalized "birth-and-death" process. Ann. Math. Stat. **19**, 1–15 (1948)

Lambert, A.: The branching process with logistic growth. Ann. Appl. Probab. **15**, 1506–1535 (2005)

Lee, A.: On the choice of MCMC kernels for approximate Bayesian computation with SMC samplers. In Proceedings of the 2012 Winter simulation conference (WSC), pages 1–12. IEEE, (2012)

Liu, J.S., Liang, F., Wong, W.H.: The multiple-try method and local optimization in Metropolis sampling. J. Am. Stat. Assoc. **95**, 121–134 (2000)

Lotka, A.J.: Elements of Physical Biology. Williams and Wilkins Co., London (1925)

Marin, J.-M., Pudlo, P., Estoup, A., Robert, C.: Likelihood-free model choice. In Sisson, S.A., Fan, Y., Beaumont, M.: editors, Handbook

of Approximate Bayesian Computation, pages 153–178. Chapman and Hall/CRC, (2018)

Marin, J.-M., Raynal, L., Pudlo, P., Robert, C. P., Estoup, A.: abcrf: Approximate Bayesian Computation via Random Forests, (2022). URL https://CRAN.R-project.org/package=abcrf. R package version 1.9

Marjoram, P., Molitor, J., Plagnol, V., Tavaré, S.: Markov chain Monte Carlo without likelihoods. Proc. Natl. Acad. Sci. **100**, 15324–15328 (2003)

Michel, L., Ćevid, D.: drf: Distributional Random Forests, (2021). URL https://CRAN.R-project.org/package=drf. R package version 1.1.0

Monari, F., Strachan, P.: Characterization of an airflow network model by sensitivity analysis: parameter screening, fixing, prioritizing and mapping. J. Build. Perform. Simul. **10**, 17–36 (2017)

Morris, M.D.: Factorial sampling plans for preliminary computational experiments. Technometrics **33**, 161–174 (1991)

Prangle, D.: Adapting the ABC distance function. Bayesian Anal. **12**, 289–309 (2017)

Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A., Feldman, M.W.: Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol. Biol. Evol. **16**, 1791–1798 (1999)

Pudlo, P., Marin, J.-M., Estoup, A., Cornuet, J.-M., Gautier, M., Robert, C.P.: Reliable ABC model choice via random forests. Bioinformatics **32**, 859–866 (2016)

Raynal, L., Marin, J.-M., Pudlo, P., Ribatet, M., Robert, C.P., Estoup, A.: ABC random forests for Bayesian parameter inference. Bioinformatics **35**, 1720–1728 (2019)

Rigatti, S.J.: Random forest. J. Insur. Med. **47**, 31–39 (2017)

Segal, M.R.: Machine learning benchmarks and random forest regression. Technical report, UCSF: Center for Bioinformatics and Molecular Biostatistics, (2004). URL: Retrieved fromhttps://escholarship.org/uc/item/35x3v9t4

Sisson, S.A., Fan, Y., Beaumont, M.: editors. Handbook of Approximate Bayesian Computation. CRC Press, (2018)

Tavaré, S.: The linear birth-death process: an inferential retrospective. Advances in Applied Probability **50**(A), 253–269 (2018)

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

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., Stumpf, M.P.: Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J. R. Soc. Interface **6**, 187–202 (2008)

Volterra, V.: Variations and fluctuations of the number of individuals in animal species living together. ICES J. Mar. Sci. **3**, 3–51 (1928)

Wilkinson, D.J.: Stochastic Modelling for Systems Biology. Chapman and Hall/CRC, (2018)