

# Multifidelity Approximate Bayesian Computation with Sequential Monte Carlo Parameter Sampling\*

Thomas P. Prescott<sup>†</sup> and Ruth E. Baker<sup>†</sup>

**Abstract.** Multifidelity approximate Bayesian computation (MF-ABC) is a likelihood-free technique for parameter inference that exploits model approximations to significantly increase the speed of ABC algorithms (Prescott and Baker, JUQ, 2020). Previous work has considered MF-ABC only in the context of rejection sampling, which does not explore parameter space particularly efficiently. In this work, we integrate the multifidelity approach with the ABC sequential Monte Carlo (ABC-SMC) algorithm into a new MF-ABC-SMC algorithm. We show that the improvements generated by each of ABC-SMC and MF-ABC to the efficiency of generating Monte Carlo samples and estimates from the ABC posterior are amplified when the two techniques are used together.

**Key words.** Bayesian inference, likelihood-free methods, stochastic simulation, multifidelity methods, sequential Monte Carlo

**AMS subject classifications.** 62F15, 65C20, 65C60, 93B30

**1. Introduction.** An important goal of the mathematical modelling of a physical system is to be able to make quantitative predictions about its behaviour. In order to make accurate predictions, the parameters of the mathematical model need to be calibrated against experimental data. Bayesian inference is a widely-used approach to model calibration that seeks to unify the information from observations with prior knowledge about the parameters to form a posterior distribution on parameter space [2, 15, 29]. This approach is based on Bayes' Theorem, whereby the posterior distribution is proportional to the product of the prior distribution and the likelihood of the data under the model.

However, in many practical settings, the model is often too complicated for the likelihood of the data to be calculated, making the posterior distribution unavailable. In this case, likelihood-free methods for Bayesian parameter inference become a useful option. Specifically, approximate Bayesian computation (ABC) is a class of such likelihood-free methods [30, 32]. Rather than calculating the likelihood of the observed data for a given parameter value, it is estimated. The model is simulated under a particular parameter value, and the simulated output is then compared with the observed data. If the simulation and observations are suitably close (according to a predetermined metric and threshold [11, 14]) then, in the classical rejection sampling approach, the likelihood is approximated as 1. Otherwise, the likelihood is approximated as 0. This binary approximation is usually interpreted as an acceptance or a rejection of the parameter value input into the simulation. Using this approach, a weighted Monte Carlo sample from an approximated posterior can be built by repeatedly

---

\*Originally submitted to the editors January 2020; resubmitted October 2020.

**Funding:** This work is supported by BBSRC through grant BB/R000816/1.

<sup>†</sup>Mathematical Institute, University of Oxford, Woodstock Road, Oxford, OX2 6GG, UK (prescott@maths.ox.ac.uk, ruth.baker@maths.ox.ac.uk)

**Funding:** REB is supported by a Royal Society Wolfson Research Merit Award and a Leverhulme Research Fellowship.

proposing parameters from the prior distribution, simulating the model with each parameter, and calculating the binary weight.

One widely-acknowledged weakness of the ABC approach is its heavy reliance on repeated simulation. There has been a significant amount of work dedicated to overcoming this reliance by exploiting methods for intelligent exploration of parameter space in order to reduce the number of simulations in areas of low likelihood [30]. For example, the parameters to be input into the simulation can instead be proposed from a importance distribution, rather than the prior, which we aim to construct in order to improve the algorithm’s performance. One successful approach to importance sampling is known as Sequential Monte Carlo ABC (ABC-SMC), which aims to build consecutive samples using parameter proposals taken from progressively closer approximations to the posterior, parameterised by decreasing ABC thresholds [33]. Research in this area has considered how to choose the sequence of decreasing thresholds and distance metrics [8, 27], and how best to evolve the parameters from one approximation to produce parameter proposals for the next [1, 3, 12].

Another proposed strategy for overcoming the simulation bottleneck is the multifidelity ABC (MF-ABC) approach [28]. This approach assumes that, in addition to the model under investigation (termed the *high-fidelity* model), there also exists a *low-fidelity* model, depending on the same parameters, that is significantly faster to simulate, usually at the cost of being less accurate [24]. The multifidelity approach to parameter inference uses simulations from both the low-fidelity model and high-fidelity model to approximate the likelihood. The high-fidelity model is simulated as little as possible, to reduce the simulation bottleneck, but just enough to ensure that the resulting posterior estimate is suitably accurate. This technique is related to multilevel approaches to ABC [17, 34, 35], which use a decreasing sequence of ABC-SMC thresholds to produce coupled estimates at different levels. Other approaches that can be interpreted in the multifidelity framework include lazy ABC [26], delayed acceptance ABC [10] and early rejection [25]. In each of these, low-fidelity simulations are sometimes used to reject parameters before completing a high-fidelity simulation. Importantly, the more general MF-ABC framework in [28] allows for early acceptance as well as early rejection.

A key observation that can be made about these two techniques for improving ABC performance is that they are orthogonal, in the sense that they improve different aspects of the ABC approach. ABC-SMC considers only improving the method for proposing parameters to use in simulations and does not directly affect the binary estimate of the likelihood. In contrast, MF-ABC makes no change to the parameter proposals, but instead directly alters the method used to estimate the likelihood by using a combination of both low-fidelity and high-fidelity model simulations. The complementarity of these two approaches has previously been shown in the specific context of combining delayed acceptance with SMC [10]. Thus, combining the general multifidelity framework of [28] with SMC should therefore yield significant speed-up over existing methods.

**1.1. Outline.** In this paper we bring together these two orthogonal approaches to speeding up ABC algorithms. We will introduce a combined multifidelity sequential Monte Carlo ABC algorithm (MF-ABC-SMC). Section 2 formulates the existing ABC algorithms briefly described above, and the techniques we can use to quantify their performance. We then show how these ABC approaches can be combined in Section 3, by incorporating the multifidelity

technique with the sequential importance sampling approach to ABC-SMC. In [Section 4](#) we then fully exploit the SMC framework to optimise the multifidelity approach, producing the MF-ABC-SMC algorithm in [Algorithm 4.1](#). This new algorithm is applied in [Section 5](#) to a heterogeneous network of Kuramoto oscillators in a hierarchical Bayes parameter estimation task, to produce low-variance ABC posterior estimates significantly faster than the classical ABC-SMC approach. Finally, in [Section 6](#), we discuss some important open questions for further optimising the MF-ABC-SMC algorithm.

**2. Theoretical background.** Assume that the model we are seeking to calibrate is a map (usually stochastic) from parameters  $\theta$ , taking values in a parameter space  $\Theta$ , to an output  $y$ , taking values in data space  $\mathcal{Y}$ . We denote this map as a conditional density  $f(\cdot | \theta)$  on  $\mathcal{Y}$ , and term the drawing of  $y \sim f(\cdot | \theta)$  as simulating the model, with  $y$  termed a simulation. For Bayesian inference, we furthermore assume the existence of a prior distribution  $\pi(\cdot)$  on  $\Theta$ , and of the observed data  $y_{\text{obs}} \in \mathcal{Y}$  that will be used to calibrate the model. The model induces the likelihood of the observed data, written  $L(\theta) = f(y_{\text{obs}} | \theta)$ , which is a function of  $\theta$ . As described previously, the goal of Bayesian inference is to infer the posterior distribution  $p(\theta | y_{\text{obs}})$  on  $\Theta$ , given  $y_{\text{obs}}$  and  $\pi(\cdot)$ . Bayes' Theorem equates the posterior to the product of likelihood and prior,

$$p(\theta | y_{\text{obs}}) = \frac{1}{\zeta} L(\theta) \pi(\theta),$$

where the normalisation constant,  $\zeta$ , ensures that  $p(\theta | y_{\text{obs}})$  is a probability density that integrates to unity.

**2.1. Approximate Bayesian computation.** Often, the model under consideration is sufficiently complicated that the likelihood cannot be calculated, necessitating a likelihood-free approach. We assume that, while the value of any  $f(y | \theta)$  is not available, we are still able to simulate  $y \sim f(\cdot | \theta)$ . Let  $d(y, y_{\text{obs}})$  denote a metric that quantifies how close any simulation,  $y$ , is to the observed data,  $y_{\text{obs}}$ . For a positive threshold value  $\epsilon > 0$ , we can then define a neighbourhood  $\Omega_\epsilon(d, y_{\text{obs}}) = \{y \in \mathcal{Y} | d(y, y_{\text{obs}}) < \epsilon\}$  of model simulations that are ‘close’ to the data. Typically the dataset,  $y_{\text{obs}}$ , is constant for the parameter estimation task, and the distance metric,  $d$ , is pre-determined. Hence we will often drop the  $(d, y_{\text{obs}})$  dependence from our notation and simply write  $\Omega_\epsilon$  for the  $\epsilon$ -neighbourhood of  $y_{\text{obs}}$  under the distance function  $d$ .

For a given positive distance threshold  $\epsilon > 0$ , ABC replaces the exact likelihood,  $L(\theta)$ , with the ABC approximation to the likelihood,

$$(2.1a) \quad L_\epsilon(\theta) = \mathbb{P}(y \in \Omega_\epsilon | \theta) = \int \mathbb{I}(y \in \Omega_\epsilon) f(y | \theta) dy,$$

which is, to leading order for small  $\epsilon$ , approximately proportional to  $L(\theta)$ . The ABC approximation to the likelihood then induces the ABC posterior,

$$(2.1b) \quad p_\epsilon(\theta | y_{\text{obs}}) = \frac{1}{Z} L_\epsilon(\theta) \pi(\theta),$$

where (similarly to  $\zeta$  above) the constant  $Z$  ensures that the ABC posterior is a probability distribution with unit integral.

**Algorithm 2.1** Importance sampling ABC (ABC-IS)

**Input:** Data  $y_{\text{obs}}$  and neighbourhood  $\Omega_\epsilon$ ; model  $f(\cdot \mid \theta)$ ; prior  $\pi$ ; importance distribution  $\hat{q}$  proportional to  $q$ ; sample index  $n = 0$ ; stopping criterion  $S$ .

**Output:** Weighted sample  $\{\theta_n, w_n\}_{n=1}^N$ .

- 1: **repeat**
- 2:   Increment  $n \leftarrow n + 1$ .
- 3:   Generate  $\theta_n \sim \hat{q}(\cdot)$ .
- 4:   Simulate  $y_n \sim f(\cdot \mid \theta_n)$ .
- 5:   Set  $w_n = [\pi(\theta_n)/q(\theta_n)] \cdot \mathbb{I}(y_n \in \Omega_\epsilon)$ .
- 6: **until**  $S = \text{true}$ .

The importance sampling ABC algorithm [23, 30] presented in [Algorithm 2.1](#) (ABC-IS) presents a simple method for drawing samples from the ABC posterior,  $p_\epsilon(\theta \mid y_{\text{obs}})$ . In addition to the data,  $y_{\text{obs}}$ , and the prior,  $\pi(\theta)$ , we assume an importance probability distribution  $\hat{q}(\theta) = q(\theta)/Z_q$  defined by the function  $q(\theta)$ , where  $q(\theta) > 0$  for all  $\theta$  in the support of  $\pi$ . Note that we do not assume that the normalisation constant  $Z_q$  is known. For each parameter proposal  $\theta_n \sim \hat{q}(\cdot)$  from the importance distribution, the model  $y_n \sim f(\cdot \mid \theta_n)$  is simulated and a weight  $w_n = w(\theta_n, y_n)$  is assigned, using the weighting function,

$$(2.2) \quad w(\theta, y) = [\pi(\theta)/q(\theta)] \cdot \mathbb{I}(y \in \Omega_\epsilon) \geq 0,$$

to produce a weighted sample  $\{w_n, \theta_n\}$ .

Note that the general stopping criterion used in [Algorithm 2.1](#) allows for the algorithm to terminate after (for example) a fixed number of non-zero weights, a fixed number of parameter proposals,  $N$ , when a fixed budget of total computational time or memory is reached, or any other more complicated combination of such conditions. Furthermore we note that ABC-IS is easily parallelised, although care must be taken to ensure that the chosen stopping condition is correctly applied in this case [16]. If we choose the importance distribution  $\hat{q} = \pi$  equal to the prior, then [Algorithm 2.1](#) (ABC-IS) is known as rejection sampling, which we refer to as ABC-RS.

For any arbitrary function  $F(\cdot)$  defined on the parameter space  $\Theta$ , we can estimate the expected value of  $F(\theta)$  under  $p_\epsilon(\theta \mid y_{\text{obs}})$ , such that

$$(2.3) \quad \mathbb{E}_{p_\epsilon}(F(\theta)) \approx \bar{F} = \frac{\sum_{n=1}^N w_n F(\theta_n)}{\sum_{n=1}^N w_n}.$$

Although this estimate is biased (except in the ABC-RS case), it is consistent, such that the mean squared error (MSE) of  $\bar{F}$  is dominated by the variance and decays to 0 on the order  $1/N$ .

**2.2. Sequential Monte Carlo (SMC).** Sequential Monte Carlo (SMC) is commonly used to efficiently explore parameter space in ABC. The goal is to propagate a sample from the prior through a sequence of intermediate distributions towards the target distribution,  $p_\epsilon(\theta \mid y_{\text{obs}})$ . The intermediate distributions are typically the sequence of ABC approximations  $p_{\epsilon_t}$ , for  $t = 1, \dots, T$ , defined by a sequence of decreasing thresholds,  $\epsilon_1 > \dots > \epsilon_T = \epsilon$ . [Algorithm 2.2](#)

presents the sequential importance sampling approach to ABC-SMC [3, 7, 31, 33], also known as population Monte Carlo [5]. Each Monte Carlo sample  $\{\theta_n^{(t)}, w_n^{(t)}\}$  built at generation  $t$  is used to construct an importance distribution  $\hat{q}_{t+1}$ , defined in Equation (2.4), that is used to generate the next generation's Monte Carlo sample. The final sample,  $\{\theta_n^{(T)}, w_n^{(T)}\}$ , is produced by Algorithm 2.1 using importance distribution  $\hat{q}_T$  and threshold  $\epsilon_T = \epsilon$ . Note that Algorithm 2.2 requires the specification of a threshold sequence  $\epsilon_t$ , the stopping conditions  $S_t$ , and the perturbation kernels  $K_t(\cdot | \theta)$  for  $t = 1, \dots, T$  [30]. We will not re-examine these aspects of ABC-SMC in detail in this paper, and will implement ABC-SMC using established techniques to select  $\epsilon_t$ ,  $S_t$  and  $K_t$  [3, 8, 12, 33].

---

**Algorithm 2.2** Sequential Monte Carlo ABC (ABC-SMC)

---

**Input:** Data  $y_{\text{obs}}$ ; sequence of nested neighbourhoods  $\Omega_{\epsilon_T} \subseteq \Omega_{\epsilon_{T-1}} \subseteq \dots \subseteq \Omega_{\epsilon_1}$  for  $0 < \epsilon = \epsilon_T < \epsilon_{T-1} < \dots < \epsilon_1$ ; prior  $\pi$ ; perturbation kernels  $K_t(\cdot | \theta)$ ; initial importance distribution  $\hat{q}_1$  (often set to  $\pi$ ); model  $f(\cdot | \theta)$ ; stopping conditions  $S_1, S_2, \dots, S_T$ .

**Output:** Weighted sample  $\left\{ \theta_n^{(T)}, w_n^{(T)} \right\}_{n=1}^{N_T}$ .

- 1: **for**  $t = 1, \dots, T - 1$  **do**
- 2:   Produce  $\{\theta_n^{(t)}, w_n^{(t)}\}_{n=1}^{N_t}$  using Algorithm 2.1 (ABC-IS) with importance distribution  $\hat{q}_t$ , neighbourhood  $\Omega_{\epsilon_t}$ , and stopping condition  $S_t$ .
- 3:   Define the next importance distribution,  $\hat{q}_{t+1}$ , proportional to

$$(2.4) \quad q_{t+1}(\theta) = \begin{cases} \frac{\sum_{n=1}^{N_t} w_n^{(t)} K_t(\theta | \theta_n^{(t)})}{\sum_{m=1}^{N_t} w_m^{(t)}} & \pi(\theta) > 0, \\ 0 & \text{else.} \end{cases}$$

- 4: **end for**
  - 5: Produce  $\left\{ \theta_n^{(T)}, w_n^{(T)} \right\}_{n=1}^{N_T}$  using Algorithm 2.1 (ABC-IS) with importance distribution  $\hat{q}_T$ , neighbourhood  $\Omega_{\epsilon_T} = \Omega_\epsilon$ , and stopping condition  $S_T$ .
- 

One weakness of the population Monte Carlo approach taken in Algorithm 2.2 is the  $O(N^2)$  cost for each run of Algorithm 2.1 (ABC-IS), where  $N_t \sim N$  is the scale of each generation's sample size. To overcome this problem, the SMC sampler was adapted in [8] to reduce this cost to  $O(N)$ . When the calculation of each  $w_n^{(t)}$  is dominated by  $q_t(\theta_n^{(t)})$ , then a significant computational burden can be alleviated through using the  $O(N)$  sampling approach. However, in many practical problems, the calculation of each weight,  $w_n^{(t)}$ , is instead dominated by sampling  $y_n^{(t)} \sim f(\cdot | \theta_n^{(t)})$ . We will focus on the latter setting and aim to reduce the cost of ABC-SMC through reducing the cost of calculating each  $w_n^{(t)}$  using the multifidelity approach described in Subsection 2.3.

**2.3. Multifidelity ABC.** Subsection 2.2 describes how the SMC strategy provides parameter proposals such that simulation time is not wasted in regions of low likelihood. An orthogonal approach to improving ABC efficiency is to avoid computationally expensive simulations, where possible, by relying on the *multifidelity* framework [28]. We term the model of interest,  $f(\cdot | \theta)$ , which maps parameter space  $\Theta$  to an output space  $\mathcal{Y}$ , as the high-fidelity

model. We now assume that, in addition, there is a low-fidelity (i.e. approximate) model,  $\tilde{f}(\cdot | \theta)$ , of the same physical system with the same parameter space  $\Theta$ . The simulations of this model are denoted  $\tilde{y} \sim \tilde{f}(\cdot | \theta)$ , taking values in the output space  $\tilde{\mathcal{Y}}$ , which may differ from  $\mathcal{Y}$ . Importantly, we assume that the low-fidelity model is computationally cheaper, in the sense that simulations  $\tilde{y} \sim \tilde{f}(\cdot | \theta)$  of the low-fidelity model incur less computational burden than simulations  $y \sim f(\cdot | \theta)$  of the high-fidelity model.

We can also assume that the experimental observations  $y_{\text{obs}} \in \mathcal{Y}$  can be mapped to the new data space, giving  $\tilde{y}_{\text{obs}} \in \tilde{\mathcal{Y}}$ . Similarly, we define an associated region,

$$\tilde{\Omega}_\epsilon = \tilde{\Omega}_\epsilon(\tilde{y}_{\text{obs}}, \tilde{d}) = \{\tilde{y} \in \tilde{\mathcal{Y}} \mid \tilde{d}(\tilde{y}, \tilde{y}_{\text{obs}}) < \tilde{\epsilon}\},$$

that is the  $\tilde{\epsilon}$ -neighbourhood of the observed data,  $\tilde{y}_{\text{obs}}$ , as a subset of the output space  $\tilde{\mathcal{Y}}$ , defined under the distance metric  $\tilde{d}$ . However, in the interests of clarity, we will assume for the remainder of this article that the output spaces of each model fidelity are such that  $\tilde{\mathcal{Y}} = \mathcal{Y}$ . Similarly, we assume that the observed data are such that  $\tilde{y}_{\text{obs}} = y_{\text{obs}}$ , the distance metrics are such that  $\tilde{d} = d$ , and the ABC thresholds are such that  $\tilde{\epsilon} = \epsilon$ , so that  $\tilde{\Omega}_\epsilon = \Omega_\epsilon$ .

In general, the models  $f(\cdot | \theta)$  and  $\tilde{f}(\cdot | \theta)$  can be simulated independently for a given  $\theta$ . Then, if the low-fidelity model is a good approximation to the high-fidelity model, the outputs  $y$  and  $\tilde{y}$  will be near, in some sense, and the distances from data  $d(y, y_{\text{obs}})$  and  $d(\tilde{y}, y_{\text{obs}})$  will be correlated. However, to improve this correlation we will also allow for coupling between the two models, writing  $\tilde{f}(y, \tilde{y} | \theta)$  as the coupled density for  $(y, \tilde{y})$  with marginals that coincide with the independent models  $f(y | \theta)$  and  $\tilde{f}(\tilde{y} | \theta)$ . The benefit of this approach is that, with a judicious choice of coupling, we can produce a high-fidelity simulation  $y \sim f(\cdot | \tilde{y}, \theta)$  conditionally on a previously-simulated low-fidelity simulation,  $\tilde{y} \sim \tilde{f}(\cdot | \theta)$ , where the distances are more closely correlated. Furthermore, simulations of  $y \sim f(\cdot | \tilde{y}, \theta)$  may be less computationally burdensome than independent simulations of  $y \sim f(\cdot | \theta)$ .

For example, suppose the high-fidelity model is a Markovian stochastic dynamical system on the time horizon  $t \in [0, T]$ , and that the low-fidelity model is the same system on  $t \in [0, \tau]$  for  $\tau < T$  [26]. The low-fidelity and high-fidelity models can clearly be simulated independently on  $[0, \tau]$  and  $[0, T]$ , respectively, and the low-fidelity simulation will, on average, be less computationally expensive than the high-fidelity simulation. However, the natural coupling of the models fixes  $y(t) = \tilde{y}(t)$  over  $t \in [0, \tau]$ , allowing the high-fidelity model to be simulated conditional on a low-fidelity simulation. Many other possible couplings exist for different multifidelity models, often involving shared random noise processes, and methods for coupling are currently an area of active research [6, 18, 28].

In order to apply the multifidelity framework to ABC parameter inference, recall that each weight,  $w_n$ , generated by Algorithm 2.1 requires a simulation  $y_n \sim f(\cdot | \theta_n)$  from the high-fidelity model. The multifidelity approach in [28] calculates the weight  $w_n = w(\theta_n, \tilde{y}_n, u_n, y_n)$  by replacing the weighting function in Equation (2.2) with

$$(2.5) \quad w(\theta, \tilde{y}, u, y) = \frac{\pi(\theta)}{q(\theta)} \left( \mathbb{I}(\tilde{y} \in \Omega_\epsilon) + \frac{\mathbb{I}(u < \alpha(\theta, \tilde{y}))}{\alpha(\theta, \tilde{y})} [\mathbb{I}(y \in \Omega_\epsilon) - \mathbb{I}(\tilde{y} \in \Omega_\epsilon)] \right),$$

where  $(\tilde{y}, y) \sim \tilde{f}(\cdot, \cdot | \theta)$  are coupled multifidelity simulations, where  $u$  is a unit uniform random variable, and where  $\alpha(\theta, \tilde{y}) \in (0, 1]$  is a positive *continuation probability*. Note that,



in general, we could allow the continuation probability,  $\alpha(\theta, \tilde{y}, u, y)$ , to depend on all of the stochastic variables, but we will assume  $\alpha(\theta, \tilde{y}) \in (0, 1]$  to be independent of  $u$  and  $y$ .

The important consequence of the multifidelity weight in Equation (2.5) is that, by the specific order in which we simulate the variables, the weight  $w(\theta, \tilde{y}, u, y)$  may be calculated without the computational cost of simulating  $y$ . Given  $\theta$ , we first simulate  $\tilde{y} \sim \tilde{f}(\cdot \mid \theta)$  from the low-fidelity model. This defines the continuation probability  $\alpha(\theta, \tilde{y}) \in (0, 1]$ . Second, we generate the unit uniform random variable,  $u$ . If  $u \geq \alpha(\theta, \tilde{y}) \in (0, 1]$  then we can return  $w(\theta, \tilde{y}, u, y)$  without simulating  $y \sim f(\cdot \mid \theta, \tilde{y})$  from the coupling, thus incurring the lower computational expense of only simulating from  $\tilde{f}$ . Algorithm 2.3 is an adaptation of Algorithm 2.1 that returns a weighted sample  $\{\theta_n, w_n\}_{n=1}^N$  from the ABC posterior  $p_\epsilon(y_{\text{obs}} \mid \theta)$ , with weights calculated using  $w$  in Equation (2.5). As with Algorithm 2.1, in the rejection sampling case where  $\hat{q}(\theta) = \pi(\theta)$ , then we refer to this algorithm as MF-ABC-RS.

---

**Algorithm 2.3** Multifidelity ABC importance sampling (MF-ABC-IS)

---

**Input:** Data  $y_{\text{obs}}$  and neighbourhood  $\Omega_\epsilon$ ; prior  $\pi$ ; coupling  $\tilde{f}(\cdot, \cdot \mid \theta)$  of models  $f(\cdot \mid \theta)$  and  $\tilde{f}(\cdot \mid \theta)$ ; continuation probability function  $\alpha = \alpha(\theta, \tilde{y})$ ; sample index  $n = 0$ ; importance distribution  $\hat{q}$  proportional to  $q(\theta)$ ; stopping condition  $S$ .

**Output:** Weighted sample  $\{\theta_n, w_n\}_{n=1}^N$ .

```

1: repeat
2:   Increment  $n \leftarrow n + 1$ .
3:   Generate  $\theta_n \sim \hat{q}(\cdot)$ .
4:   Simulate  $\tilde{y}_n \sim \tilde{f}(\cdot \mid \theta_n)$ .
5:   Set  $w_n = \mathbb{I}(\tilde{y}_n \in \Omega_\epsilon)$ .
6:   Generate  $u_n \sim \text{Uniform}(0, 1)$ .
7:   if  $u_n < \alpha(\theta_n, \tilde{y}_n)$  then
8:     Simulate  $y_n \sim f(\cdot \mid \tilde{y}_n, \theta_n)$ .
9:     Update  $w_n \leftarrow w_n + [\mathbb{I}(y_n \in \Omega_\epsilon) - w_n] / \alpha(\theta_n, \tilde{y}_n)$ .
10:  end if
11:  Update  $w_n \leftarrow [\pi(\theta_n) / q(\theta_n)] w_n$ .
12: until  $S = \text{true}$ .
```

---

**Proposition 2.1.** *The weighted Monte Carlo sample  $\{\theta_n, w_n\}$  returned by Algorithm 2.3 is from the ABC posterior  $p_\epsilon(\theta \mid y_{\text{obs}})$ .*

*Proof.* We note that the density of each  $z = (\theta, \tilde{y}, u, y)$  sampled by Algorithm 2.3 is

$$g(z) = \tilde{f}(\tilde{y}, y \mid \theta) \hat{q}(\theta),$$

on  $\mathcal{Z} = \Theta \times \mathcal{Y} \times [0, 1] \times \mathcal{Y}$ . Furthermore, the multifidelity weight in Equation (2.5) integrates such that

$$\int_0^1 w(z) \, du = \frac{\pi(\theta)}{q(\theta)} \mathbb{I}(y \in \Omega_\epsilon),$$

which is independent of  $\tilde{y}$ . Therefore, for any integrable  $F : \Theta \rightarrow \mathbb{R}$ , we have the identity

$$\int_{\Theta} F(\theta) \pi_\epsilon(\theta \mid y_{\text{obs}}) \, d\theta = \frac{Z_q}{Z} \int_{\Theta \times \mathcal{Y}} F(\theta) \frac{\pi(\theta)}{q(\theta)} \mathbb{I}(y \in \Omega_\epsilon) f(y \mid \theta) \hat{q}(\theta) \, d\theta \, dy$$

$$\begin{aligned}
&= \frac{Z_q}{Z} \int_{\Theta \times \mathcal{Y}^2} F(\theta) \frac{\pi(\theta)}{q(\theta)} \mathbb{I}(y \in \Omega_\epsilon) \check{f}(\tilde{y}, y \mid \theta) \hat{q}(\theta) \, d\theta \, dy \, d\tilde{y} \\
&= \frac{Z_q}{Z} \int_{\mathcal{Z}} F(\theta) w(z) g(z) \, dz.
\end{aligned}$$

Thus the normalised Monte Carlo estimate in Equation (2.3) is consistent for a weighted sample from Algorithm 2.3.  $\blacksquare$

The key issue for Algorithm 2.3 (MF-ABC-IS) is the choice of continuation probability,  $\alpha(\theta, \tilde{y}) \in (0, 1]$ . Smaller values for  $\alpha$  lead to greater computational savings, since high-fidelity simulations are generated less often. However, it is possible to show that this comes at the cost of an increase in the mean squared error (MSE) of the estimators,  $\bar{F}$ , of integrable functions  $F : \Theta \rightarrow R$ . In Subsection 2.4, we quantify this tradeoff through the definition of the algorithm's efficiency.

**Note 2.2.** The multifidelity weight in Equation (2.5) compares a uniform  $u \in [0, 1]$  with  $\alpha(\theta, \tilde{y})$ . In more generality, we may consider any multifidelity weight  $w(\theta, \tilde{y}, u, y)$  that depends on  $u \in \mathcal{U}$  with any distribution,  $p(u \mid \theta, \tilde{y})$ , that does not rely on the expensive high-fidelity model. If  $w$  is designed such that

$$\int_{u \in \mathcal{U}} w(\theta, \tilde{y}, u, y) p(u \mid \theta, \tilde{y}) \, du = \frac{\pi(\theta)}{q(\theta)} \mathbb{I}(y \in \Omega_\epsilon),$$

then, by an extension of Proposition 2.1, the multifidelity weight remains valid.

**2.4. ABC Performance.** In Section 2 so far, we have summarised previous developments of ABC algorithms that aim to improve performance, but have not yet defined how to quantify this improvement. For any weighted sample  $\{\theta_n, w_n\}$  built using Algorithm 2.1 (ABC-IS) or Algorithm 2.3 (MF-ABC-IS), each sampled pair  $(\theta_n, w_n)$  also incurs a computational cost, which we will denote  $T_n$ , producing a total computational cost of  $T_{\text{total}} = \sum_n T_n$  for the sample. We define the observed efficiency of such a sample as follows.

**Definition 2.3.** The effective sample size [9, 19] of a weighted Monte Carlo sample  $\{\theta_n, w_n\}$  is

$$\text{ESS} = \frac{(\sum_n w_n)^2}{\sum_n w_n^2}.$$

Given a computational cost of  $T_n$  for each sampled pair  $(w_n, \theta_n)$ , the observed efficiency of a weighted Monte Carlo sample is

$$\frac{\text{ESS}}{T_{\text{total}}} = \frac{(\sum_n w_n)^2}{(\sum_n w_n^2) (\sum_n T_n)},$$

which is expressed in units of effective samples per time unit.

Larger values of ESS typically correspond to smaller values of MSE in estimates of the form in Equation (2.3). Since ESS and  $T_{\text{total}}$  both scale linearly with  $N$ , taking limits as  $N \rightarrow \infty$  in the observed efficiency motivates the following definition of the theoretical efficiency of an ABC algorithm.



**Definition 2.4.** *The theoretical efficiency of an ABC algorithm generating a weighted Monte Carlo sample  $\{\theta_n, w_n\}$  with simulation times  $\{T_n\}$  is*

$$\psi = \frac{\mathbb{E}(w)^2}{\mathbb{E}(w^2)\mathbb{E}(T)},$$

*which is expressed in units of effective samples per time unit.*

Note that the theoretical efficiency is not just a characteristic of the ABC algorithm and of the models, but also of the numerical implementation and hardware of the computers generating the simulation. For example, the theoretical efficiency of [Algorithm 2.1](#) (ABC-IS) is

$$\psi_{\text{ABC-IS}} = \frac{Z}{\mathbb{E}_{\pi_\epsilon}(\pi/\hat{q})\mathbb{E}_{\hat{q}}(T)},$$

where  $Z = p(y \in \Omega_\epsilon)$  is the normalisation constant in [Equation \(2.1b\)](#), and where  $\mathbb{E}_\nu$  denotes expectations with respect to probability density  $\nu$ . Setting  $\hat{q} = \pi$ , the theoretical efficiency of ABC-RS is

$$\psi_{\text{ABC-RS}} = \frac{Z}{\mathbb{E}_\pi(T)}.$$

Hence, the theoretical efficiency of ABC-IS is improved over ABC-RS by choosing an importance distribution,  $\hat{q}$ , that is more likely than  $\pi$  to propose  $\theta$  incurring smaller simulation times and high posterior likelihoods. In the case of [Algorithm 2.3](#) with  $\hat{q}(\theta) = \pi(\theta)$  (MF-ABC-RS), this performance measure has been used to determine a good choice of continuation probability,  $\alpha(\theta, \tilde{y})$  [28].

By using the theoretical efficiency,  $\psi$ , as a performance metric in the remainder of this paper, we will quantify the improvement in performance over [Algorithms 2.1 to 2.3](#) that can be achieved by combining multifidelity and SMC techniques.

**3. Multifidelity ABC-SMC.** There are two distinct approaches to improving the performance of ABC parameter inference specified in [Section 2](#). ABC-SMC proposes parameters  $\theta_n$  from a sequence of importance distributions that progressively approximate the target ABC approximation to the posterior. In contrast, MF-ABC enables sampled parameters to be weighted without necessarily having to produce a simulation,  $y_n$ , from the high-fidelity model. In this section we present the main contribution of this paper, which is to combine these orthogonal approaches into an MF-ABC-SMC algorithm.

We will replicate the procedure of extending [Algorithm 2.1](#) (ABC-IS) into [Algorithm 2.2](#) (ABC-SMC) in the multifidelity context, focusing first on the  $O(N^2)$  SMC sampler based on sequential importance sampling. Suppose that, as in [Subsection 2.2](#), we have a decreasing sequence of thresholds,  $\epsilon_1 > \dots > \epsilon_T = \epsilon$ , inducing the neighbourhoods,  $\Omega_{\epsilon_1} \supseteq \dots \supseteq \Omega_{\epsilon_T}$ , and a sequence of ABC posteriors,  $p_{\epsilon_t}(\theta \mid y_{\text{obs}})$ . In principle, we can replace the call of [Algorithm 2.1](#) (ABC-IS) in step 2 of [Algorithm 2.2](#) with a call of [Algorithm 2.3](#) (MF-ABC-IS) instead, and the SMC algorithm would proceed in much the same way. However, the key difficulty with implementing multifidelity ABC-SMC lies in the definition of the importance distribution,  $\hat{q}_{t+1}$ , for generation  $t+1$ , given the Monte Carlo sample,  $\{\theta_n^{(t)}, w_n^{(t)}\}_{n=1}^{N_t}$ , returned in generation  $t$ .

Since the weights,  $w_n$ , are now calculated using the multifidelity weight in Equation (2.5), there is a positive probability that there exist  $w_n < 0$ . Negative values of  $w_n$  are generated whenever  $\tilde{y}_n \in \Omega_\epsilon$ , but where we also simulate the high-fidelity model such that  $y_n \notin \Omega_\epsilon$ . This leads to two problems with the existing definition of the importance distribution in Equation (2.4). Primarily, there may exist  $\theta$  in the prior support with  $q_{t+1}(\theta) < 0$ , unless the perturbation kernels,  $K_t$ , are carefully designed to avoid this. Secondly, even if the  $K_t$  could be chosen to guarantee  $q_{t+1}(\theta) > 0$  on the prior support, it is not clear that we can easily sample from the importance distribution  $\hat{q}_{t+1} \propto q_{t+1}$  when some  $\theta_n^{(t)}$  have negative weights.

---

**Algorithm 3.1** Multifidelity ABC-SMC with pre-determined  $\alpha_t$  (MF-ABC-SMC- $\alpha$ )

---

**Input:** Data  $y_{\text{obs}}$ ; sequence of nested neighbourhoods  $\Omega_{\epsilon_T} \subseteq \Omega_{\epsilon_{T-1}} \subseteq \dots \subseteq \Omega_{\epsilon_1}$  for  $0 < \epsilon = \epsilon_T < \epsilon_{T-1} < \dots < \epsilon_1$ ; prior  $\pi$ ; coupling  $\check{f}(\cdot, \cdot \mid \theta)$  of models  $f(\cdot \mid \theta)$  and  $\tilde{f}(\cdot \mid \theta)$ ; initial importance distribution  $\hat{r}_1$  (often set to  $\pi$ ); perturbation kernels  $K_t(\cdot \mid \theta)$ ; continuation probabilities  $\alpha_t(\theta, \tilde{y})$ ; stopping conditions  $S_t$ ; where  $t = 1, \dots, T$ .

**Output:** Weighted sample  $\{\theta_n^{(T)}, w_n^{(T)}\}_{n=1}^{N_T}$ .

- 1: **for**  $t = 1, \dots, T - 1$  **do**
  - 2:   Produce  $\{\theta_n^{(t)}, w_n^{(t)}\}_{n=1}^{N_t}$  from Algorithm 2.3 (MF-ABC-IS), using the neighbourhood  $\Omega_{\epsilon_t}$ , continuation probability  $\alpha_t$ , importance distribution  $\hat{r}_t$ , and stopping condition  $S_t$ .
  - 3:   Set  $\hat{w}_n^{(t)} = |w_n^{(t)}|$  for  $n = 1, \dots, N_t$ .
  - 4:   Define  $\hat{r}_{t+1}$  proportional to  $r_{t+1}$  given in Equation (3.1)
  - 5: **end for**
  - 6: Produce  $\{\theta_n^{(T)}, w_n^{(T)}\}_{n=1}^{N_T}$  from Algorithm 2.3, using neighbourhood  $\Omega_{\epsilon}$ , continuation probability  $\alpha_T$ , importance distribution  $\hat{r}_T$  and stopping condition  $S_T$ .
- 

In Algorithm 3.1 (MF-ABC-SMC- $\alpha$ ) we have adapted the SMC approach to counter the possibility of negative weights, by considering a sampling algorithm that produces non-negative weights  $\hat{w}_n^{(t)} \geq 0$  in parallel with  $w_n^{(t)}$ . This method closely parallels the approach to the sign problem taken in [20]. At each generation we produce two weighted samples at once,  $\{\hat{w}_n^{(t)}, \theta_n^{(t)}\}$  and  $\{w_n^{(t)}, \theta_n^{(t)}\}$ . We replace the importance distribution  $\hat{q}_{t+1}$  defined by  $q_{t+1}$  in Equation (2.4) with the importance distribution  $\hat{r}_{t+1}$  proportional to the non-negative function

$$(3.1) \quad r_{t+1}(\theta) = \begin{cases} \sum_{n=1}^N \hat{w}_n^{(t)} K_t(\theta \mid \theta_n^{(t)}) / \sum_{m=1}^N \hat{w}_m^{(t)} & \pi(\theta) > 0, \\ 0 & \text{else,} \end{cases}$$

which is defined by the weights  $\hat{w}_n^{(t)} = |w_n^{(t)}|$ . With this choice of importance distribution, the weighted samples  $\{\hat{w}_n^{(t)}, \theta_n^{(t)}\}$  can be shown to be drawn from the alternative target distribution,

$$(3.2) \quad \rho_t(\theta) \propto p_{\epsilon_t}(\theta \mid y_{\text{obs}}) + \delta_t(\theta),$$

where the difference between the new target distribution,  $\rho_t$ , and the ABC posterior,  $p_{\epsilon_t}$ , is

given by

$$\delta_t(\theta) = 2\pi(\theta) \int_{\mathcal{Y}^2} (1 - \alpha_t(\theta, \tilde{y})) \mathbb{I}(\tilde{y} \in \Omega_{\epsilon_t}) \mathbb{I}(y \notin \Omega_{\epsilon_t}) \check{f}(\tilde{y}, y \mid \theta) \, d\tilde{y} \, dy,$$

for  $t = 1, \dots, T$ . The function  $\delta_t$  implies that the new target distribution (compared to  $\pi_{\epsilon_t}$ ) contains additional density in regions of parameter space where it is more likely that  $\tilde{y} \in \Omega_{\epsilon_t}$  but  $y \notin \Omega_{\epsilon_t}$ ; in other words, where  $w_n^{(t)} < 0$  is more likely.

The importance distribution in Equation (3.1) effectively makes the SMC algorithm target  $\rho_t$  at each generation instead of  $p_{\epsilon_t}$ . However, the weighted samples  $\{w_n^{(t)}, \theta_n^{(t)}\}$  based on the multifidelity weights  $w_n^{(t)}$  from each generation's weighting function,

$$(3.3) \quad w_t(\theta, \tilde{y}, u, y) = \frac{\pi(\theta)}{r_t(\theta)} \left( \mathbb{I}(\tilde{y} \in \Omega_{\epsilon_t}) + \frac{\mathbb{I}(u < \alpha_t(\theta, \tilde{y}))}{\alpha_t(\theta, \tilde{y})} [\mathbb{I}(y \in \Omega_{\epsilon_t}) - \mathbb{I}(\tilde{y} \in \Omega_{\epsilon_t})] \right),$$

remain from the ABC posteriors  $p_{\epsilon_t}$ . Hence, at any generation (and in particular at  $t = T$ ), we can produce an estimate of a  $p_{\epsilon_t}$ -integrable function  $F : \Theta \rightarrow \mathbb{R}$ , such that

$$\mathbb{E}_{p_{\epsilon_t}}(F) \approx \frac{\sum_n w_n^{(t)} F(\theta_n^{(t)})}{\sum_m w_m^{(t)}},$$

is a consistent Monte Carlo estimate of  $F$  under the ABC posterior.

**Note 3.1.** Each calculation of the weight in Equation (3.3) relies on the  $O(N)$  calculation of the importance weight in Equation (3.1), making the SMC sampler  $O(N^2)$ . An alternative sampling method that is linear in  $N$  is proposed in [8], which replaces the sequential importance sampling approach described above. However, as noted in [8], both the  $O(N)$  and  $O(N^2)$  sampling algorithms require  $O(N)$  simulations of  $(\tilde{y}, y) \sim \check{f}(\cdot, \cdot \mid \theta)$  or  $\tilde{y} \sim \tilde{f}(\cdot \mid \theta)$ . In this multifidelity setting, we are assuming that simulation time comprises the vast majority of the computational burden of each calculation of Equation (3.3). The benefit of the multifidelity weight is that it reduces the computational burden of generating  $(\tilde{y}, y)$  by sometimes requiring  $\tilde{y}$  alone. In Section SM1 we show that the  $O(N)$  SMC sampling algorithm dilutes this benefit. Therefore, in this work we will focus on the sequential importance sampling SMC sampler described in Algorithm 3.1.

**4. Adaptive MF-ABC-SMC.** In Algorithm 3.1 (MF-ABC-SMC- $\alpha$ ), in addition to an assumed sequence of ABC thresholds,  $\epsilon_t$ , perturbation kernels,  $K_t$ , and stopping conditions  $S_t$ , for  $t = 1, \dots, T$ , we also assume a given sequence of continuation probabilities,  $\alpha_t$ . In this algorithm, each importance distribution,  $\hat{r}_{t+1}$ , is determined by the output at generation  $t$ . Methodologies for adaptively choosing the perturbation kernels,  $K_{t+1}$ , and the ABC thresholds,  $\epsilon_{t+1}$ , based on the preceding generations' samples, have been explored in previous work [8, 12]. In this section, we will consider the adaptive approach to choosing each generation's continuation probability,  $\alpha_{t+1}$ , based on the simulation output of generation  $t$ .

**4.1. Optimal continuation probabilities.** For each generation,  $t$ , the continuation probability,  $\alpha_t$ , is an input into that generation's call of Algorithm 2.3. Dropping the generational indexing  $t$  temporarily, in this subsection we first consider how to choose a continuation probability function,  $\alpha(\theta, \tilde{y})$ , to maximise the theoretical efficiency,  $\psi$ , of any run of Algorithm 2.3 (MF-ABC-IS), as specified in Definition 2.4.

For simplicity, we will constrain the search for optimal  $\alpha(\theta, \tilde{y})$  to the piecewise constant function

$$(4.1) \quad \alpha(\theta, \tilde{y}) = \eta_1 \mathbb{I}(\tilde{y} \in \Omega_\epsilon) + \eta_2 \mathbb{I}(\tilde{y} \notin \Omega_\epsilon),$$

for the constants  $\eta_1, \eta_2 \in (0, 1]$ . Here,  $\eta_1$  is the probability of generating  $y$  after a ‘positive’ low-fidelity simulation (where  $\tilde{y} \in \Omega_\epsilon$ ) and  $\eta_2$  is the probability of generating  $y$  after a ‘negative’ low-fidelity simulation (where  $\tilde{y} \notin \Omega_\epsilon$ ). The goal of this section is to specify the values of the two parameters,  $\eta_1$  and  $\eta_2$ , that will give the largest theoretical efficiency,  $\psi$ . In previous work, we have derived the optimal values of  $\eta_1$  and  $\eta_2$  to use in the special case of [Algorithm 2.3](#) (MF-ABC-RS) corresponding to rejection sampling, where the importance distribution  $\hat{q} = \pi$  equal to the prior distribution [\[28\]](#). We can now extend this analysis by finding optimal values of  $\eta_1$  and  $\eta_2$  to use in the more general case of [Algorithm 2.3](#) (MF-ABC-IS). The key to this optimisation is the following lemma, which describes how the efficiency of [Algorithm 2.3](#) varies with the continuation probabilities used. The lemma assumes a given importance distribution,  $\hat{q}(\theta)$ , defined as the normalisation of the known non-negative function  $q(\theta)$ .

**Lemma 4.1.** *The theoretical efficiency, given in [Definition 2.4](#), of [Algorithm 2.3](#) (MF-ABC-IS) varies with the continuation probabilities  $\eta_1$  and  $\eta_2$  according to*

$$\psi(\eta_1, \eta_2) = \frac{\mathbb{E}(w)^2}{\mathbb{E}(w^2)\mathbb{E}(T)} = \frac{Z^2}{\phi(\eta_1, \eta_2)},$$

where the denominator is expressed as a function of  $(\eta_1, \eta_2)$  such that

$$(4.2) \quad \phi(\eta_1, \eta_2) = \left( W + \left( \frac{1}{\eta_1} - 1 \right) W_{\text{fp}} + \left( \frac{1}{\eta_2} - 1 \right) W_{\text{fn}} \right) (\bar{T}_{\text{lo}} + \eta_1 \bar{T}_{\text{hi,p}} + \eta_2 \bar{T}_{\text{hi,n}}).$$

The coefficients in  $\psi$  are given by the integrals

$$(4.3a) \quad Z = \int L_\epsilon(\theta) \pi(\theta) \, d\theta, \quad L_\epsilon(\theta) = \int_{\mathcal{Y}^2} \mathbb{I}(y \in \Omega_\epsilon) \check{f}(\tilde{y}, y \mid \theta) \, d\tilde{y} \, dy,$$

$$(4.3b) \quad W = \int \frac{\pi(\theta)}{q(\theta)} L_\epsilon(\theta) \pi(\theta) \, d\theta,$$

$$(4.3c) \quad W_{\text{fp}} = \int \frac{\pi(\theta)^2}{q(\theta)} p_{\text{fp}}(\theta) \, d\theta, \quad p_{\text{fp}}(\theta) = \int_{\mathcal{Y}^2} \mathbb{I}(\tilde{y} \in \Omega_\epsilon) \mathbb{I}(y \notin \Omega_\epsilon) \check{f}(\tilde{y}, y \mid \theta) \, d\tilde{y} \, dy,$$

$$(4.3d) \quad W_{\text{fn}} = \int \frac{\pi(\theta)^2}{q(\theta)} p_{\text{fn}}(\theta) \, d\theta, \quad p_{\text{fn}}(\theta) = \int_{\mathcal{Y}^2} \mathbb{I}(\tilde{y} \notin \Omega_\epsilon) \mathbb{I}(y \in \Omega_\epsilon) \check{f}(\tilde{y}, y \mid \theta) \, d\tilde{y} \, dy,$$

$$(4.3e) \quad \bar{T}_{\text{lo}} = \int T_{\text{lo}}(\theta) q(\theta) \, d\theta, \quad T_{\text{lo}}(\theta) = \int_{\mathcal{Y}^2} T(\tilde{y}) \check{f}(\tilde{y}, y \mid \theta) \, d\tilde{y} \, dy,$$

$$(4.3f) \quad \bar{T}_{\text{hi,p}} = \int T_{\text{hi,p}}(\theta) q(\theta) \, d\theta, \quad T_{\text{hi,p}}(\theta) = \int_{\mathcal{Y}^2} T(y) \mathbb{I}(\tilde{y} \in \Omega_\epsilon) \check{f}(\tilde{y}, y \mid \theta) \, d\tilde{y} \, dy,$$

$$(4.3g) \quad \bar{T}_{\text{hi,n}} = \int T_{\text{hi,n}}(\theta) q(\theta) \, d\theta, \quad T_{\text{hi,n}}(\theta) = \int_{\mathcal{Y}^2} T(y) \mathbb{I}(\tilde{y} \notin \Omega_\epsilon) \check{f}(\tilde{y}, y \mid \theta) \, d\tilde{y} \, dy,$$

where  $T(\tilde{y})$  is the computational cost of simulating  $\tilde{y} \sim \check{f}(\cdot \mid \theta)$  and where  $T(y)$  is the cost of simulating  $y \sim f(\cdot \mid \theta, \tilde{y})$ , such that  $T(\tilde{y}) + T(y)$  is the cost of simulating  $(\tilde{y}, y) \sim \check{f}(\cdot, \cdot \mid \theta)$ .

**4.1.1. Optimising efficiency.** We can conclude from [Lemma 4.1](#) that the optimal continuation probabilities  $(\eta_1^*, \eta_2^*)$  in any closed, bounded domain  $\mathcal{H} \subseteq (0, 1]^2$  are those that minimise the function  $\phi(\eta_1, \eta_2)$  given in [Equation \(4.2\)](#). [Lemmas 4.2](#) and [4.3](#) below explicitly find the global minimiser of  $\phi$  over  $[0, \infty)^2$  (if it exists), and then over the boundary  $\partial\mathcal{H}$  of a rectangular domain  $\mathcal{H} = [\rho_1, 1] \times [\rho_2, 1]$ , where the user-specified lower bounds  $\rho_1$  and  $\rho_2$  are chosen to ensure  $\mathcal{H}$  is closed. These results combine in [Proposition 4.4](#) to give the minimiser of  $\phi$  over  $\mathcal{H}$ , and hence the optimal continuation probabilities for use in [Algorithm 2.3](#).

**Lemma 4.2.** *We first consider all non-negative values of  $\eta_1, \eta_2 \geq 0$ . If  $W > W_{\text{fp}} + W_{\text{fn}}$ , then the minimum value of  $\phi(\eta_1, \eta_2)$  in [Equation \(4.2\)](#), and the optimal value of  $(\eta_1, \eta_2)$  in the entire positive quadrant, are given by*

$$(4.4a) \quad \bar{\phi} = \left( \sqrt{(W - W_{\text{fp}} - W_{\text{fn}})\bar{T}_{\text{lo}}} + \sqrt{W_{\text{fp}}\bar{T}_{\text{hi,p}}} + \sqrt{W_{\text{fn}}\bar{T}_{\text{hi,n}}} \right)^2,$$

$$(4.4b) \quad (\bar{\eta}_1, \bar{\eta}_2) = \left( \sqrt{\frac{\bar{T}_{\text{lo}}}{W - W_{\text{fp}} - W_{\text{fn}}} \cdot \frac{W_{\text{fp}}}{\bar{T}_{\text{hi,p}}}}, \sqrt{\frac{\bar{T}_{\text{lo}}}{W - W_{\text{fp}} - W_{\text{fn}}} \cdot \frac{W_{\text{fn}}}{\bar{T}_{\text{hi,n}}}} \right),$$

respectively. If  $W \leq W_{\text{fp}} + W_{\text{fn}}$ , then there is no minimum of  $\phi(\eta_1, \eta_2)$  in  $\eta_1, \eta_2 \geq 0$ .

**Lemma 4.3.** *Under the same conditions as [Lemma 4.2](#), fix the closed region  $\mathcal{H} = [\rho_1, 1] \times [\rho_2, 1]$  of positive continuation probabilities with user-defined lower bounds  $\rho_1, \rho_2 \in (0, 1)$ . Define the two functions, for  $x > 0$ ,*

$$(4.5a) \quad \eta_1(x) = \max \left\{ \rho_1, \min \left[ 1, \sqrt{\frac{\bar{T}_{\text{lo}} + \bar{T}_{\text{hi,n}}x}{W - W_{\text{fp}} - (1 - x^{-1})W_{\text{fn}}} \cdot \frac{W_{\text{fp}}}{\bar{T}_{\text{hi,p}}}} \right] \right\},$$

$$(4.5b) \quad \eta_2(x) = \max \left\{ \rho_2, \min \left[ 1, \sqrt{\frac{\bar{T}_{\text{lo}} + \bar{T}_{\text{hi,p}}x}{W - (1 - x^{-1})W_{\text{fp}} - W_{\text{fn}}} \cdot \frac{W_{\text{fn}}}{\bar{T}_{\text{hi,n}}}} \right] \right\}.$$

Then the minimum value of  $\phi$  on the boundary,  $\partial\mathcal{H}$ , of  $\mathcal{H}$  is attained at the minimum of  $\phi(1, \eta_2(1))$ ,  $\phi(\eta_1(1), 1)$ ,  $\phi(\rho_1, \eta_2(\rho_1))$  or  $\phi(\eta_1(\rho_2), \rho_2)$ .

**Proposition 4.4.** *Assume the same conditions as [Lemmas 4.2](#) and [4.3](#). Compute the minimiser,  $(\bar{\eta}_1, \bar{\eta}_2)$ , and minimal value,  $\bar{\phi}$ , of  $\phi$  in  $(0, \infty)^2$  using [Lemma 4.2](#), if they exist. If  $(\bar{\eta}_1, \bar{\eta}_2) \in \mathcal{H}$  then set  $(\eta_1^*, \eta_2^*) = (\bar{\eta}_1, \bar{\eta}_2)$  and  $\phi^* = \bar{\phi}$ . Otherwise, set  $\phi^*$  equal to the minimum of the four values of  $\phi$  listed in [Lemma 4.3](#), and  $(\eta_1^*, \eta_2^*)$  to the associated argument. Then  $\phi^*$  is the minimum value of  $\phi$  over  $(\eta_1, \eta_2) \in \mathcal{H}$ , and  $(\eta_1^*, \eta_2^*)$  are the minimising continuation probabilities.*

**4.1.2. Interpreting efficiency.** The optimised efficiency of [Algorithm 2.3](#) is determined by the values of the various coefficients defined in [Equation \(4.3\)](#). The normalisation constant,  $Z$ , and the ABC approximation to the likelihood,  $L_\epsilon$ , are properties of the ABC approach as defined in [Equation \(2.1\)](#). The coefficient  $W$  can be written  $W = Z\mathbb{E}_\pi(p_\epsilon/q)$ , as a scaling of the prior expectation of the ratio  $p_\epsilon/q$ . Larger values of  $W$  require  $q$  to be more concentrated (relative to  $\pi$ ) in regions of high posterior density, which is a well-known characteristic of importance sampling [\[23\]](#). Thus, neither  $Z$  nor  $W$  relate specifically to the multifidelity approach and are properties of ABC importance sampling.

The coefficient  $\bar{T}_{lo}$  represents the average time taken to simulate  $\tilde{y}$  from the low-fidelity model. The other two time-based coefficients,  $\bar{T}_{hi,p}$  and  $\bar{T}_{hi,n}$ , represent the average time taken to complete the simulation of  $(\tilde{y}, y) \sim \tilde{f}$  from the coupling, conditional on whether or not  $\tilde{y} \in \Omega_\epsilon$ . These coefficients therefore determine how much time can be saved by avoiding expensive simulations and stopping after generating  $\tilde{y} \sim \tilde{f}(\cdot \mid \theta)$ .

The key determinants of the success of the multifidelity technique are  $W_{fp}$  and  $W_{fn}$  and their tradeoff between the high-fidelity simulation costs,  $\bar{T}_{hi,p}$  and  $\bar{T}_{hi,n}$ . Equation (4.2) implies that the marginal cost to the efficiency of decreasing  $\eta_1$  and  $\eta_2$  is smaller for smaller values of  $W_{fp}$  and  $W_{fn}$ . These coefficients can be written as the two prior expectations,

$$W_{fp} = \mathbb{E}_\pi \left( \frac{\pi}{q} p_{fp} \right), \quad W_{fn} = \mathbb{E}_\pi \left( \frac{\pi}{q} p_{fn} \right).$$

Thus, they are scalings of the probabilities of a false positive (where  $\tilde{y} \in \Omega_\epsilon$  but  $y \notin \Omega_\epsilon$ ) and a false negative (where  $\tilde{y} \notin \Omega_\epsilon$  but  $y \in \Omega_\epsilon$ ), respectively. Hence, small values of  $W_{fp}$  and  $W_{fn}$  correspond to at least one of the following cases. First, if the low-fidelity and high-fidelity models are closely correlated, then the probability of a false positive or false negative is small. This demonstrates the value of coupling the models, as described in Subsection 2.3. Second, for small values of  $W_{fp}$  and  $W_{fn}$  we require  $q$  to be larger than  $\pi$  in regions of parameter space where false positives or false negatives are relatively likely. We can interpret this as a requirement that (in addition to  $q$  being concentrated in regions of high posterior density) the region of parameter space where simulations of the low-fidelity and high-fidelity model are less often in agreement (in terms of membership of  $\Omega_\epsilon$ ) should be explored more thoroughly by  $q$  than by  $\pi$ .

**4.2. Constructing continuation probabilities.** There is an important barrier to implementing Proposition 4.4 as a method for choosing optimal continuation probabilities. Before running any ABC iterations, and in the absence of extensive analysis of the models being simulated, the quantities in Equation (4.3) are unknown. We therefore cannot directly construct the optimisers in Equations (4.4) and (4.5). However, recall that we are aiming to use this method in the context of sequential Monte Carlo to adaptively produce continuation probabilities. In Algorithm 3.1, at each generation  $t \geq 1$ , we have a sample  $\{\theta_n^{(t)}, w_n^{(t)}\}$  that is used to produce  $\hat{r}_{t+1}$ . The proposed approach to constructing continuation probabilities is similar: we will use the same Monte Carlo sample to also produce approximately optimal values of  $\eta_1$  and  $\eta_2$  defining the continuation probability  $\alpha_{t+1}$  to use at generation  $t + 1$ . The following definition specifies how to calculate approximations of the quantities in Equation (4.3) using an existing Monte Carlo sample. These approximations can then be substituted into Equations (4.4) and (4.5). Hence, we can estimate the optimal continuation probabilities as given by Proposition 4.4.

**Definition 4.5.** Consider a Monte Carlo sample  $\{\theta_n, w_n\}_{n=1}^N$  constructed from a run of Algorithm 2.3, which used the importance distribution  $\hat{q}(\theta)$  proportional to  $q(\theta)$  and continuation probability  $\alpha(\theta, \tilde{y})$ . For the set of low-fidelity simulations  $\{\tilde{y}_n\}_{n=1}^N$  and high-fidelity simulations  $\{y_n\}_{n \in M}$ , where  $M = \{n : y_n \text{ exists}\}$ , store: the simulation times,  $\tilde{t}_n = T(\tilde{y}_n)$  and  $t_n = T(y_n)$ ; the distances from data,  $\tilde{d}_n = d(\tilde{y}_n, y_{\text{obs}})$  and  $d_n = d(y_n, y_{\text{obs}})$ ; the importance densities  $q_n = q(\theta_n)$ ; and the continuation probabilities  $\alpha_n = \alpha(\theta_n, \tilde{y}_n)$ .

We now consider finding the optimal continuation probability,  $\alpha^*$ , to be used in a new run of [Algorithm 2.3](#), with the new importance distribution,  $q^*(\theta)$ , and the new ABC threshold,  $\epsilon$ . We define the Monte Carlo estimates,

$$(4.6a) \quad \hat{Z} = \frac{1}{N} \left[ \sum_{n=1}^N \frac{\pi(\theta_n)}{q_n} \mathbb{I}(\tilde{d}_n < \epsilon) + \sum_{n \in M} \frac{\pi(\theta_n)}{q_n \alpha_n} \left( \mathbb{I}(d_n < \epsilon) - \mathbb{I}(\tilde{d}_n < \epsilon) \right) \right],$$

$$(4.6b) \quad \hat{W} = \frac{1}{N} \left[ \sum_{n=1}^N \frac{\pi(\theta_n)^2}{q^*(\theta_n) q_n} \mathbb{I}(\tilde{d}_n < \epsilon) + \sum_{n \in M} \frac{\pi(\theta_n)^2}{q^*(\theta_n) q_n \alpha_n} \left( \mathbb{I}(d_n < \epsilon) - \mathbb{I}(\tilde{d}_n < \epsilon) \right) \right],$$

$$(4.6c) \quad \hat{W}_{\text{fp}} = \frac{1}{N} \sum_{n \in M} \frac{\pi(\theta_n)^2}{q^*(\theta_n) q_n \alpha_n} \mathbb{I}(\tilde{d}_n < \epsilon) \mathbb{I}(d_n \geq \epsilon),$$

$$(4.6d) \quad \hat{W}_{\text{fn}} = \frac{1}{N} \sum_{n \in M} \frac{\pi(\theta_n)^2}{q^*(\theta_n) q_n \alpha_n} \mathbb{I}(\tilde{d}_n \geq \epsilon) \mathbb{I}(d_n < \epsilon),$$

$$(4.6e) \quad \hat{T}_{\text{lo}} = \frac{1}{N} \sum_{n=1}^N \frac{q^*(\theta_n)}{q_n} \tilde{t}_n,$$

$$(4.6f) \quad \hat{T}_{\text{hi,p}} = \frac{1}{N} \sum_{n \in M} \frac{q^*(\theta_n)}{q_n \alpha_n} \mathbb{I}(\tilde{d}_n < \epsilon) t_n,$$

$$(4.6g) \quad \hat{T}_{\text{hi,n}} = \frac{1}{N} \sum_{n \in M} \frac{q^*(\theta_n)}{q_n \alpha_n} \mathbb{I}(\tilde{d}_n \geq \epsilon) t_n,$$

corresponding to the quantities in [Equation \(4.3\)](#).

The estimates in [Equation \(4.6\)](#) are scaled Monte Carlo estimates of the quantities in [Equation \(4.3\)](#), such that the approximation

$$\psi(\eta_1, \eta_2) = \frac{Z^2}{\phi(\eta_1, \eta_2)} \approx \frac{\hat{Z}^2}{\hat{\phi}(\eta_1, \eta_2)},$$

holds, with

$$(4.7) \quad \hat{\phi}(\eta_1, \eta_2) = \left( \hat{W} + \left( \frac{1}{\eta_1} - 1 \right) \hat{W}_{\text{fp}} + \left( \frac{1}{\eta_2} - 1 \right) \hat{W}_{\text{fn}} \right) \left( \hat{T}_{\text{lo}} + \eta_1 \hat{T}_{\text{hi,p}} + \eta_2 \hat{T}_{\text{hi,n}} \right).$$

Thus, we can substitute the estimates in [Equation \(4.6\)](#) into [Equations \(4.4\) and \(4.5\)](#). Applying [Proposition 4.4](#) with these estimates thus provides near-optimal continuation probabilities for the new run of [Algorithm 2.3](#), constructed from the existing Monte Carlo sample.

**Note 4.6.** The Monte Carlo estimates in [Equation \(4.6\)](#) are not independent of each other, so there is a bias in the approximation  $\hat{Z}^2/\hat{\phi}$ . Hence, the continuation probabilities  $(\eta_1^*, \eta_2^*)$  returned by [Proposition 4.4](#), if using the estimates in [Equation \(4.6\)](#), can only be near-optimal.

**4.3. Adaptive MF-ABC-SMC algorithm.** [Algorithm 3.1](#) presents a multifidelity ABC-SMC algorithm that relies on a predetermined sequence of continuation probability functions,



$\alpha_t(\theta, \tilde{y})$ . In [Subsections 4.1](#) and [4.2](#) we have shown how to use an existing Monte Carlo sample to produce continuation probabilities of the form

$$\alpha_t(\theta, \tilde{y}) = \eta_1 \mathbb{I}(\tilde{y} \in \Omega_{\epsilon_t}) + \eta_2 \mathbb{I}(\tilde{y} \notin \Omega_{\epsilon_t}),$$

where the values of  $\eta_1$  and  $\eta_2$  are chosen according to [Proposition 4.4](#), in order to (approximately) optimise the efficiency of generating a sample from [Algorithm 2.3](#). This result allows us to write an adaptive MF-ABC-SMC algorithm that uses the Monte Carlo output of [Algorithm 2.3](#) at generation  $t$  to construct not just an importance distribution,  $\hat{r}_{t+1}(\theta)$ , but also a continuation probability,  $\alpha_{t+1}(\theta, \tilde{y})$ , for use in the next generation.

[Algorithm 4.1](#) (MF-ABC-SMC) is an adaptive multifidelity sequential Monte Carlo algorithm for ABC parameter inference. In place of the pre-defined continuation probabilities  $\alpha_t$  used in [Algorithm 3.1](#) (MF-ABC-SMC- $\alpha$ ), we instead only require an initial continuation probability, most sensibly set to  $\alpha_1 \equiv 1$ , and the two lower bounds,  $\rho_1$  and  $\rho_2$ , on the allowed values of the continuation probabilities. Since the final sample is generated by a run of [Algorithm 2.3](#), by [Proposition 2.1](#) it follows that the weighted sample is from the ABC posterior,  $p_\epsilon(\theta \mid y_{\text{obs}})$ .

---

**Algorithm 4.1** Multifidelity ABC-SMC (MF-ABC-SMC)

---

**Input:** Data  $y_{\text{obs}}$ ; sequence of nested neighbourhoods  $\Omega_{\epsilon_T} \subseteq \Omega_{\epsilon_{T-1}} \subseteq \dots \subseteq \Omega_{\epsilon_1}$  for  $0 < \epsilon = \epsilon_T < \epsilon_{T-1} < \dots < \epsilon_1$ ; prior  $\pi$ ; coupling  $\check{f}(\cdot, \cdot \mid \theta)$  of models  $f(\cdot \mid \theta)$  and  $\check{f}(\cdot \mid \theta)$ ; initial importance distribution  $\hat{r}_1$  (often set to  $\pi$ ); initial continuation probability  $\alpha_1 \equiv 1$ ; lower bounds on continuation probabilities,  $\rho_1, \rho_2 \in (0, 1)$ ; perturbation kernels  $K_t(\cdot \mid \theta)$ ; stopping conditions  $S_t$ ; where  $t = 1, \dots, T$ .

**Output:** Weighted sample  $\{\theta_n^{(T)}, w_n^{(T)}\}_{n=1}^{N_T}$ .

- 1: **for**  $t = 1, \dots, T - 1$  **do**
  - 2:   Produce  $\{\theta_n^{(t)}, w_n^{(t)}\}_{n=1}^{N_t}$  from [Algorithm 2.3](#) (MF-ABC-IS), using the neighbourhood  $\Omega_{\epsilon_t}$ , continuation probability  $\alpha_t$ , importance distribution  $\hat{r}_t$ , and stopping condition  $S_t$ . Store simulation times, distances, importance densities and continuation probabilities as specified in [Definition 4.5](#).
  - 3:   Set  $\hat{w}_n^{(t)} = |w_n^{(t)}|$  for  $n = 1, \dots, N_t$ .
  - 4:   Define  $\hat{r}_{t+1}$  proportional to  $r_{t+1}$  given in [Equation \(3.1\)](#)
  - 5:   Update the estimates in [Equation \(4.6\)](#) with the values stored at step 2, using importance distribution  $r_{t+1}(\theta)$  and ABC threshold  $\epsilon_{t+1}$ .
  - 6:   Calculate  $(\eta_1^*, \eta_2^*)$  using [Proposition 4.4](#) with lower bounds  $\rho_1, \rho_2$ .
  - 7:   Set  $\alpha_{t+1}(\tilde{y}, \theta) = \eta_1^* \mathbb{I}(\tilde{y} \in \Omega_{\epsilon_t}) + \eta_2^* \mathbb{I}(\tilde{y} \notin \Omega_{\epsilon_t})$ .
  - 8: **end for**
  - 9: Produce  $\{\theta_n^{(T)}, w_n^{(T)}\}_{n=1}^{N_T}$  from [Algorithm 2.3](#), using neighbourhood  $\Omega_{\epsilon_T}$ , continuation probability  $\alpha_T$ , importance distribution  $\hat{r}_T$  and stopping condition  $S_T$ .
- 

*Note 4.7.* In common with the non-adaptive [Algorithm 3.1](#) (MF-ABC-SMC- $\alpha$ ), the importance weights  $r_t(\theta_n^{(t)})$  are required to construct the weights  $w_n^{(t)}$  in step 2, each of which is an  $O(N)$  calculation. However, each calculation of  $r_{t+1}(\theta_n^{(t)})$  required in step 5 is also  $O(N)$ .

Thus, there is extra cost at each generation on the order  $O(N^2)$ . However, in common with [Note 3.1](#), we will assume that this cost is justified by our aim to reduce the large simulation burden that dominates the algorithm's run time.

**5. Example: Kuramoto Oscillator Network.** To demonstrate the multifidelity and SMC approaches to parameter inference, we will infer the parameters of a Kuramoto oscillator model on a complete network, with stochastic heterogeneity in each node's intrinsic frequency. In [Subsection 5.1](#) we consider the performance of the previously developed ABC algorithms introduced in [Section 2](#) (ABC-RS, ABC-SMC and MF-ABC-RS) and demonstrate the orthogonal ways in which the SMC and multifidelity techniques improve performance. In [Subsection 5.2](#) we apply the adaptive algorithm, [Algorithm 4.1](#) (MF-ABC-SMC) to demonstrate that the efficiency of parameter estimation is significantly improved by combining the multifidelity and SMC approaches. The algorithms have been implemented in Julia [\[4\]](#) and the source code can be found at [github.com/tpprescott/mf-abc-smc](https://github.com/tpprescott/mf-abc-smc).

The Kuramoto oscillator model is defined on a complete network of  $M$  nodes, where each node,  $i$ , has a dynamically evolving phase value,  $\phi_i$ , determined by the ordinary differential equation

$$(5.1) \quad \dot{\phi}_i = \omega_i + \frac{K}{M} \sum_{j=1}^M \sin(\phi_j - \phi_i),$$

for  $i = 1, \dots, M$ . Each constant  $\omega_i$ , the intrinsic angular velocity, is an independent draw from a Cauchy distribution with median  $\omega_0$  and dispersion parameter  $\gamma$ . In addition to these two parameters, we have an interconnection strength  $K$ . Simulations of the ODE system are run over a fixed time interval  $t \in [0, T]$ , and we will assume fixed initial conditions  $\phi_i(0) = 0$  for all  $i$ .

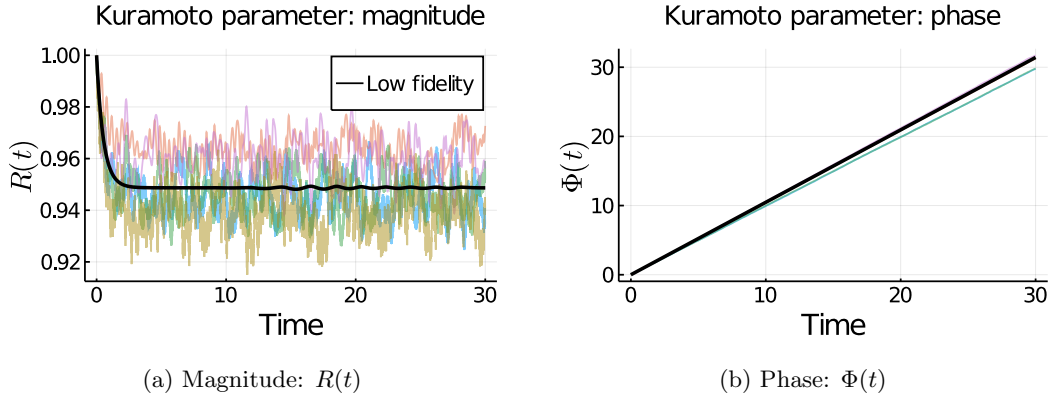
The multifidelity approach makes use of a low-dimensional approximation of the coupled oscillator dynamics, as described in [\[13, 21, 22\]](#). The approximation is based on tracking the Daido order parameters, which are a set of complex-valued representations of the high-dimensional vector  $(\phi_i)_{i=1}^M$ , defined as

$$Z_n(t) = \frac{1}{M} \sum_{j=1}^M \exp(in\phi_j),$$

for positive integers  $n$  and the imaginary unit  $i$ . A system of coupled ODEs can be generated for the set of  $Z_n$ . Under the assumption that  $Z_n(t) = Z_1(t)^n$ , known as the Ott-Antonsen ansatz [\[13, 21\]](#), the system can be reduced to a single ODE for  $Z_1$ , which is known as the Kuramoto parameter. This complex-valued trajectory is usually represented by two real trajectories, corresponding to its magnitude  $R(t) = \|Z_1(t)\|$  and phase  $\Phi(t) = \arg(Z_1(t))$ . The approximation of the  $M$ -dimensional ODE system in [Equation \(5.1\)](#) under the OA ansatz is thus given by the two-dimensional ODE system

$$(5.2a) \quad \dot{R} = \left( \frac{K}{2} - \gamma \right) \tilde{R} - \frac{K}{2} \tilde{R}^3,$$

$$(5.2b) \quad \dot{\Phi} = \omega_0,$$



**Figure 5.1.** In colour are five trajectories for  $R(t)$  and  $\Phi(t)$  for the high-fidelity model in Equation (5.1). Stochasticity arises from sampling  $\omega_i \sim \text{Cauchy}(\omega_0, \gamma)$  for  $i = 1, \dots, 256$ . In black is the deterministic trajectory from the low-fidelity model in Equation (5.2). All simulations were completed using parameters  $K = 2$ ,  $\omega_0 = \pi/3$  and  $\gamma = 0.1$ . Note that the phase plot “unwraps” the trajectories of  $\Phi(t) = \arg(Z_1(t))$  to avoid  $2\pi$ -discontinuities.

with initial conditions  $(\tilde{R}(0), \tilde{\Phi}(0)) = (1, 0)$ , which directly simulates the low-dimensional representation of the  $M$ -dimensional state vector.

The goal of this example is to infer the parameters  $(K, \omega_0, \gamma)$  based on synthetic data, generated by simulating a system of  $M = 256$  oscillators with random angular velocities  $\omega_i$  over  $t \in (0, 30]$ . We record the trajectories  $R_{\text{obs}}(t)$  and  $\Phi_{\text{obs}}(t)$  of the magnitude and phase of the Kuramoto parameter. The parameter values used to generate these data are  $(K = 2, \omega_0 = \pi/3, \gamma = 0.1)$ . The likelihood of the observed data under the model in Equation (5.1) with stochastic parameters  $\omega_i \sim \text{Cauchy}(\omega_0, \gamma)$  is unavailable, and we must therefore resort to ABC inference, requiring repeated simulation.

Example trajectories of the high-fidelity and low-fidelity models in Equations (5.1) and (5.2) are given in Figure 5.1. The trajectories  $R_{\text{obs}}(t)$ ,  $\Phi_{\text{obs}}(t)$ ,  $R(t)$ ,  $\Phi(t)$ ,  $\tilde{R}(t)$  and  $\tilde{\Phi}(t)$  on  $t \in [0, 30]$  are infinite dimensional. In order to easily compare trajectories, we will select a finite number of informative *summary statistics* from the trajectories, guided by the approximated system in Equation (5.2). We take

$$\begin{aligned} S_1(R, \Phi) &= \left( \frac{1}{30} \int_0^{30} R(t) \, dt \right)^2, \\ S_2(R, \Phi) &= \frac{1}{30} (\Phi(30) - \Phi(0)), \\ S_3(R, \Phi) &= R(T_{1/2}), \end{aligned}$$

where  $T_{1/2}$  is the first value of  $t \in [0, 30]$  for which  $R_{\text{obs}}(t)$  is halfway between  $R_{\text{obs}}(0) = 1$  and its average value  $S_1(R_{\text{obs}}, \Phi_{\text{obs}})^{1/2}$ . Justification for the choice of these summary statistics and distances is provided in Section SM2.

Simulation of the high-fidelity model produces  $y \sim f(\cdot \mid (K, \omega_0, \gamma))$  by: (a) generating  $\omega_i$ ,  $i = 1, \dots, 256$ , from  $\text{Cauchy}(\omega_0, \gamma)$ ; then (b) simulating the ODE system in Equation (5.1);

Algorithm	ESS	Sim. time (min)	Efficiency (ESS min <sup>-1</sup> )	
Algorithm 2.1 (ABC-RS)	148.0	43.6	3.39	
Algorithm 2.2 (ABC-SMC)	255.1	48.8	5.23	×1.54
Algorithm 2.3 (MF-ABC-RS)	126.4	22.9	5.52	×1.63

Table 5.1

Comparing existing ABC algorithms for sampling  $p_{0.5}((K, \omega_0, \gamma) \mid y_{\text{obs}})$  based on a uniform prior, using  $\epsilon = 0.5$ . The stopping condition for Algorithms 2.1 and 2.3 is  $N = 6000$ . The threshold schedule for Algorithm 2.2 is  $(2, 1.5, 1, 0.5)$  with stopping conditions  $S_t$  of  $N_t = 1500$  parameter proposals for  $t = 1, 2, 3, 4$ , leading to the same total number of parameter proposals as Algorithms 2.1 and 2.3. The perturbation kernels  $K_t$  in Algorithm 2.2 are Gaussian with diagonal covariance equal to twice the empirical variance of the sample at generation  $t$  [3]. The continuation probability used in Algorithm 2.3 is fixed at the constant  $\alpha \equiv 0.5$ . Percentages refer to the increase in efficiency over the base efficiency of ABC-RS.

then (c) computing  $y = (S_1(R, \Phi), S_2(R, \Phi), S_3(R, \Phi))$ . Simulation of the low-fidelity model produces  $\tilde{y} \sim \tilde{f}(\cdot \mid (K, \omega_0, \gamma))$  by: (a) simulating the ODE system in Equation (5.2); then (b) computing  $\tilde{y} = (S_1(\tilde{R}, \tilde{\Phi}), S_2(\tilde{R}, \tilde{\Phi}), S_3(\tilde{R}, \tilde{\Phi}))$ . The distances  $d(y, y_{\text{obs}})$  and  $d(\tilde{y}, y_{\text{obs}})$  are defined according to the weighted Euclidean norm,  $d(a, b)^2 = 4(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2$ . Note that the low-fidelity model is deterministic. Therefore there is no meaningful definition of a coupling between the models at each fidelity: any simulation  $y \sim f(\cdot \mid \tilde{y}, (K, \omega_0, \gamma)) = f(\cdot \mid (K, \omega_0, \gamma))$  from the high-fidelity model will be independent of  $\tilde{y}$ .

**5.1. Existing ABC algorithms.** We set independent uniform priors on  $[1, 3]$ ,  $[-2\pi, 2\pi]$  and  $[0, 1]$  for  $K$ ,  $\omega_0$  and  $\gamma$ , respectively. Using the uniform prior as importance distributions, samples from the ABC posterior,  $p_{0.5}((K, \omega_0, \gamma) \mid y_{\text{obs}})$ , are produced using Algorithm 2.1 (ABC-RS), Algorithm 2.2 (ABC-SMC), and Algorithm 2.3 (MF-ABC-RS) with  $\epsilon = 0.5$ . The continuation probability used in Algorithm 2.3 (MF-ABC-RS) is the constant  $\alpha \equiv 0.5$ . The resulting samples are depicted in Figures SM3.1 to SM3.3. Table 5.1 shows the observed values for the performance of each of these algorithms, quantified in terms of ESS, total simulation time, and observed efficiency (i.e. the ratio of the first two). Note that the ESS of the sample from Algorithm 2.2 (ABC-SMC) depends on only  $N_4 = 1500$  weights,  $w_n^{(4)}$ , corresponding to the final generation. However, we will measure the observed efficiency by using the total time to simulate, which includes the total simulation time of the preceding generations.

Even with minimal tuning of Algorithms 2.2 and 2.3, the samples built using these algorithms both show significant improvements in efficiency. We have chosen stopping conditions to ensure equal number of parameter proposals for each algorithm, in order to demonstrate the distinct effects of each. Algorithm 2.2 (ABC-SMC) produces a larger ESS for a similar simulation time. This is characteristic of ABC-SMC, whereby parameters with low likelihood are less likely to be proposed. However, Algorithm 2.3 (MF-ABC-RS) instead speeds up the simulation time of the fixed number of parameter proposals, albeit with some damage to the ESS. This result illustrates the orthogonal effects of the SMC and multifidelity ABC algorithms, and thus the potential for combining the techniques in Algorithm 4.1 to produce further gains in efficiency.

The key to the success of MF-ABC-RS and the multifidelity approach in general is the assumption that the low-fidelity model is cheaper to simulate than the high-fidelity model. In this case, the high-fidelity model in Equation (5.1) has a mean (respectively, standard deviation) simulation time of approximately 520 (638)  $\mu\text{s}$  to simulate, while the low-fidelity model in Equation (5.2) has a simulation time of approximately 10 (12)  $\mu\text{s}$ . Note that these averages and standard deviations are observed across the uniform distribution of parameter values on the intervals  $[1, 3]$ ,  $[-2\pi, 2\pi]$  and  $[0, 1]$  for  $K$ ,  $\omega_0$  and  $\gamma$ , respectively. The relatively large standard deviations imply that parameter proposals in this domain can produce very different simulation times. Different importance distributions will therefore vastly alter the relative simulation costs of the high-fidelity and low-fidelity models.

**5.2. Multifidelity ABC-SMC.** In order to demonstrate the increased efficiency of combining multifidelity approaches with SMC, we produced 100 samples from the ABC posterior,  $p_{0.1}((K, \omega_0, \gamma) \mid y_{\text{obs}})$ , consisting of 50 replicates from each of Algorithm 2.2 (ABC-SMC) and Algorithm 4.1 (MF-ABC-SMC). Common to both algorithms is the number of generations,  $T = 8$ , which corresponds to the nested sequence of ABC neighbourhoods  $\Omega_{\epsilon_t}$  with the sequence of thresholds 2.0, 1.5, 1.0, 0.8, 0.6, 0.4, 0.2 and 0.1. Each generation has a stopping condition of  $\text{ESS} \geq 400$ , evaluated after every 100 parameter proposals (to allow for parallelisation). This condition reflects a specification that we need each generation's sample to be, in some sense, 'good enough' to produce a reliable importance distribution that can be used in the next generation. Finally, we specified the perturbation kernels  $K_t(\cdot \mid (K, \omega_0, \gamma)_n^{(t)})$  at each generation to be Gaussians centred on the parameter value  $(K, \omega_0, \gamma)_n^{(t)}$ . The covariance matrices are diagonal matrices  $\text{diag}(\sigma_K^{(t)}, \sigma_{\omega_0}^{(t)}, \sigma_{\gamma}^{(t)})$ , where

$$(\sigma_K^{(t)})^2 = 2 \frac{\sum |w_n^{(t)}| (K_n^{(t)} - \mu_K^{(t)})^2}{\sum |w_n^{(t)}|},$$

$$\mu_K^{(t)} = \frac{\sum |w_n^{(t)}| K_n^{(t)}}{\sum |w_n^{(t)}|},$$

and similarly for  $\sigma_{\omega_0}^{(t)}$  and  $\sigma_{\gamma}^{(t)}$ . These perturbation kernels implement a typical choice for the covariance of using twice the empirical variance of the observed parameter values [3, 12]. Note that we use this definition for the multifidelity case also, where weights  $w_n^{(t)}$  may be negative, since we are using the  $\rho_t$  (as defined in Equation (3.2)) and not the  $p_{\epsilon_t}$  as the target distributions. Further to these common inputs, the parameters  $\rho_1$  and  $\rho_2$  are the only additional algorithm parameters we need to specify to implement Algorithm 4.1 (MF-ABC-SMC). We set lower bounds of  $\rho_1 = \rho_2 = 0.01$  on the allowed continuation probabilities, with the aim to limit the variability of  $w_n^{(t)}$  to prevent the collapse of the ESS.

**5.2.1. Multifidelity ABC-SMC increases observed efficiency.** Algorithms 2.2 and 4.1 were implemented and run using Julia 1.5.1 on a 36 core CPU ( $2 \times 18$  core with hyperthreading), 2.3/3.7 GHz, 768 GB RAM. Table 5.2 quantifies the average performance for each of Algorithms 2.2 and 4.1, separated out for each generation from  $t = 1, \dots, 8$ , and also across the entire SMC algorithm. The final row of this table demonstrates that Algorithm 4.1 (MF-ABC-SMC) results in a 60% saving in the total simulation time required to produce a final

$t$	a. Sim. time / proposal $\mu\text{s}$			b. Simulation time min			c. Efficiency ESS $\text{s}^{-1}$		
	ABC	MF-ABC		ABC	MF-ABC		ABC	MF-ABC	
1	537	530	−1%	13.6	13.4	−2%	0.51	0.52	$\times 1.01$
2	675	31	−95%	11.2	1.3	−88%	0.63	8.35	$\times 13.4$
3	665	33	−95%	15.2	2.0	−87%	0.46	4.41	$\times 9.59$
4	599	35	−94%	11.6	1.9	−84%	0.60	4.99	$\times 8.26$
5	525	35	−93%	11.3	2.2	−81%	0.62	3.99	$\times 6.47$
6	424	36	−92%	10.8	2.6	−76%	0.64	2.90	$\times 4.53$
7	330	46	−86%	13.6	6.0	−56%	0.50	1.16	$\times 2.31$
8	268	80	−70%	15.5	11.8	−24%	0.43	0.57	$\times 1.31$
SMC	449	65	−85%	102.9	41.2	−60%	0.066	0.163	$\times 2.48$

Table 5.2

Row  $t$  shows empirical mean values for 50 ABC-SMC samples and 50 MF-ABC-SMC samples of: (a) the simulation time divided by the number,  $N_t$ , of parameter proposals in generation  $t$ ; (b) the simulation time in generation  $t$ ; (c) ESS divided by simulation time in generation  $t$ . The final row is the empirical mean values for each sample, of (a) the total simulation time divided by the total number,  $\sum_t N_t$ , of parameter proposals; (b) the total simulation time; (c) ESS from generation 8 divided by total simulation time (all generations). The third columns for each of (a), (b) and (c) quantify the improvement in performance of MF-ABC-SMC over ABC-SMC.

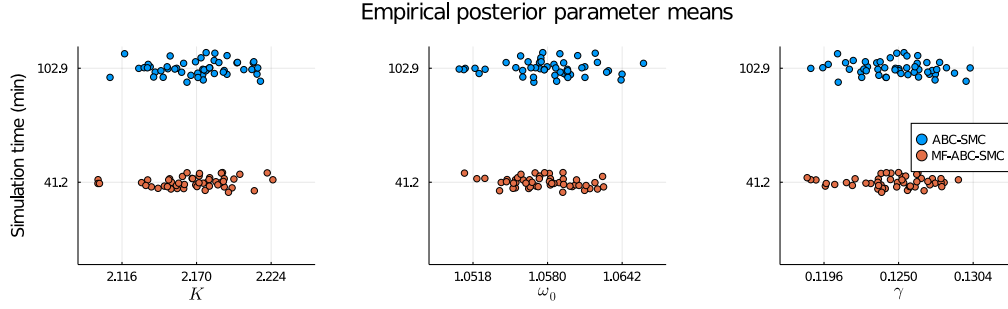
sample from the ABC posterior  $p_{0.1}$  with an ESS of approximately 400. This corresponds to an efficiency 2.48 times that of Algorithm 2.2 (ABC-SMC). This performance improvement is derived from an 85% saving in the simulation time required for each parameter proposal when averaged across the entire SMC algorithm.

Figure 5.2 depicts 100 estimates of the posterior mean of each of the parameters,  $\mathbb{E}_{p_{0.1}}(K)$ ,  $\mathbb{E}_{p_{0.1}}(\omega_0)$  and  $\mathbb{E}_{p_{0.1}}(\gamma)$ , constructed from the 50 samples generated by each of Algorithms 2.2 and 4.1. This figure demonstrates that the estimates generated by the multifidelity algorithm, MF-ABC-SMC, are of a similar quality to those produced by ABC-SMC.<sup>1</sup> Taking  $K$ ,  $\omega_0$  and  $\gamma$  in turn, the means of the 50 estimates produced by each algorithm are, to three significant figures, indistinguishable at 2.17, 1.06 and 0.125. Similarly, the variability of these estimates are also broadly similar: Algorithm 2.2 (respectively, Algorithm 4.1) produces estimates with standard deviations .0268 (.0277), .0034 (.0027), and .00271 (.00265). Importantly, the distribution of total simulation times for each of these 100 estimates demonstrates that MF-ABC-SMC is reliably significantly less computationally expensive than ABC-SMC to produce comparable samples, with the average simulation times reflecting the 60% saving identified in Table 5.2.

Recall our initial observation that the average simulation time of the low-fidelity model, 10  $\mu\text{s}$ , is approximately 2% of that of the high-fidelity model, 520  $\mu\text{s}$ , with expectations taken over the prior. Given this initial difference in each model’s average simulation time, the observed

<sup>1</sup>In addition to the observation of the posterior means in Figure 5.2, we have also depicted representative posterior samples from each algorithm in Figures SM3.4 and SM3.5.





**Figure 5.2.** The empirical posterior means of each of  $K$ ,  $\omega_0$  and  $\gamma$  for each of the 100 samples, plotted against the total simulation time required to generate each mean.

148% increase in efficiency from Algorithm 2.2 to Algorithm 4.1 is determined by a number of other factors specific to the characteristics of SMC sampling, which we now explore.

**5.2.2. MF-ABC-SMC is more effective in early generations.** In producing the samples summarised in Table 5.2 and Figure 5.2, we enforced the same decreasing schedule of  $\epsilon_t$  and the same stopping criteria ( $\text{ESS} \geq 400$ ) for all runs of both Algorithm 2.2 (ABC-SMC) and Algorithm 4.1 (MF-ABC-SMC). This allows a direct comparison between the two algorithms of the efficiencies at each generation, in addition to their overall performance. Figure 5.3 depicts the distributions of each of the measures for which the means are given in Table 5.2. As the generation index varies, there are significant differences between the performance improvement generated by MF-ABC-SMC over ABC-SMC. By all three of the measures, the benefit of MF-ABC-SMC appears to accrue most significantly in the earlier generations. There are a number of factors that explain these observed differences.

First, we observe that the average simulation time per proposal of Algorithm 2.2 (ABC-SMC) decreases as  $t$  increases. As the importance distribution evolves towards the posterior through the SMC algorithm, the most expensive high-fidelity simulations are required much less often. This means that the relative saving of using the low-fidelity model also evolves with the generation,  $t$ . The continuation probabilities used in Algorithm 4.1 (MF-ABC-SMC) aim to balance the saving in simulation cost against the probability of false positives and false negatives, according to the optima in Equation (4.4). Since there is less computational cost available to save in later generations, we therefore find larger continuation probabilities. Figure 5.4 shows how the values of  $\eta_1$  and  $\eta_2$  vary for each of the 50 instances of Algorithm 4.1 (MF-ABC-SMC) across generations  $t = 2, 4, 6, 8$ , reflecting the adaptation of the algorithm to changing distributions in simulation costs. Values in the bottom-left of the figure correspond to smaller continuation probabilities, and thus smaller total simulation times. At  $t = 8$ , the optimal continuation probabilities are no longer clustered to the bottom-left of the figure, but instead have begun to migrate to the (1, 1) corner, which corresponds to the classical ABC-SMC approach. This leads to the increase in simulation time per proposal for MF-ABC-SMC at generation  $t = 8$ , as more high-fidelity simulations are required.

Second, we observe that for generation  $t = 8$  we have an average of 70% improvement in simulation time per proposal but only a 24% average improvement in the generation's total



simulation time. It follows that MF-ABC-SMC requires approximately 2.5 times as many parameter proposals, on average, in the final generation to produce an ESS of at least 400 when compared with ABC-SMC, meaning that only a 30% improvement in efficiency can be found. Thus, the final generation incurs over a quarter of the total simulation cost of MF-ABC-SMC, on average. The improved efficiency at the final generation is relatively small because at smaller ABC thresholds the inaccuracy of the low-fidelity model becomes more significant. In this example, the underlying stochasticity of the high-fidelity model means that the probability  $\mathbb{P}(y \in \Omega_{0.1} \mid \tilde{y} \in \Omega_{0.1})$  is sufficiently small to produce a high false positive rate, corresponding to a relatively large value of  $W_{\text{fp}}$  in [Lemma 4.2](#). This reduces the optimal available efficiency towards that of classical ABC-SMC. It should be noted that a strength of the adaptive method of selecting continuation probabilities is that this case can be detected, so that the efficiency of [Algorithm 4.1](#) (MF-ABC-SMC) is bounded below by the efficiency of [Algorithm 2.2](#) (ABC-SMC).

Finally, note that the sequence of  $\epsilon_t$  was chosen to be equal across ABC-SMC and MF-ABC-SMC to allow for a direct comparison between the efficiencies at each generation. This constraint produced an overall 60% saving in simulation time, and a 148% increase in efficiency. However, in practice, we should also aim to adaptively choose the sequence  $\epsilon_t$  (for either algorithm) to optimise performance. In the following subsection, we describe how incorporating the adaptive selection of  $\epsilon_t$  based on the preceding generations' output allows for a better comparison between the two algorithms.

**5.2.3. MF-ABC-SMC reduces ABC approximation bias.** Implementing [Algorithm 4.1](#) (MF-ABC-SMC) requires a decreasing sequence of  $\epsilon_t$  values to be pre-specified. Often, appropriate values of  $\epsilon_t$  cannot be determined before any simulations have been generated. If the sequence of  $\epsilon_t$  decreases too slowly then the algorithm will take a long time to reach the target posterior; too quickly, and acceptance rates will be too low. As a result, rather than specifying a sequence of thresholds *a priori*, previous work in the SMC context [8] has explored choosing  $\epsilon_{t+1}$  by predicting its effect on the ESS of that generation. In the spirit of this approach, we adaptively choose the sequence  $\epsilon_t$  by predicting its effect on the efficiency of that generation, as defined in [Lemma 4.1](#).

[Figure 5.5](#) depicts a possible strategy for choosing  $\epsilon_{t+1}$  conditionally on the output from generation  $t$  in ABC-SMC and MF-ABC-SMC. The key to this strategy is the observation that the efficiency of [Algorithm 2.3](#), defined by Equations (4.2) and (4.3), depends on the ABC threshold,  $\epsilon$ , and the importance distribution,  $\hat{q}$ , used in [Algorithm 2.3](#). By writing the efficiency as  $\psi(\eta_1, \eta_2; \epsilon, \hat{q})$  and fixing  $\eta_1$ ,  $\eta_2$  and  $\hat{q}$ , we can consider the efficiency of [Algorithm 2.3](#) as a function of  $\epsilon$ . In particular, by setting  $\eta_1 = \eta_2 = 1$ , the efficiency of [Algorithm 2.1](#),  $\psi(1, 1; \epsilon, \hat{q})$ , also varies with  $\epsilon$ . We assume that at each generation we have a target efficiency,  $\psi_t^*$ , that enables a given ESS to be generated with a known computational budget.

In generation  $t$  of [Algorithm 4.1](#) (MF-ABC-SMC), we produce the sample  $\{\theta_n^{(t)}, w_n^{(t)}\}$  from [Algorithm 2.3](#) (MF-ABC-IS), which can be used to define an importance distribution,  $\hat{r}_{t+1}$ . Steps 5 and 6 of [Algorithm 4.1](#) then use the next ABC threshold,  $\epsilon_{t+1}$ , to calculate optimal continuation probabilities  $(\eta_1^*, \eta_2^*)$  by maximising the efficiency function,  $\psi(\eta_1, \eta_2; \epsilon_{t+1}, \hat{r}_{t+1})$ . In the case where  $\epsilon_{t+1}$  is unknown, an adaptive approach to finding an appropriate value is

to replace steps 5 and 6 of [Algorithm 4.1](#) with the following subroutine:

- a. find  $(\eta_1^*, \eta_2^*)$  to maximise the efficiency,  $\psi(\eta_1, \eta_2; \epsilon_t, \hat{r}_{t+1})$ , using [Proposition 4.4](#);
- b. set  $\epsilon^* \leftarrow \max \{0 < \epsilon \leq \epsilon_t : \psi(\eta_1^*, \eta_2^*; \epsilon, \hat{r}_{t+1}) \leq \psi_{t+1}^*\}$ , or  $\epsilon^* \leftarrow \epsilon_t$  if this set is empty;
- c. set  $\epsilon_{t+1} \leftarrow \epsilon^*$  and continue to step 7 of [Algorithm 4.1](#).

This procedure produces a sequence of ABC thresholds designed such that each generation's efficiency is maintained at a target level. It alternates between finding continuation probabilities that maximise the efficiency at the preceding threshold,  $\epsilon_t$ , and then finding a value of  $\epsilon_{t+1} < \epsilon_t$  such that the predicted efficiency matches the target. If  $\epsilon_{t+1} = \epsilon_t$  then the target efficiency,  $\psi_{t+1}^*$ , needs to be reviewed as it is not achievable. Note that for the case of ABC-SMC, which is equivalent to MF-ABC-SMC with fixed continuation probabilities  $\eta_1 = \eta_2 = 1$ , we skip the optimisation in step a and use  $\eta_1^* = \eta_2^* = 1$  in step b.

The strength of the multifidelity approach in this context is that the additional degrees of freedom afforded by the continuation probabilities allows for the ABC thresholds to decrease more quickly, while maintaining a target efficiency at each generation. This benefit is depicted in [Figure 5.5](#). Each curve in the left-hand plot is the predicted efficiency  $\psi(1, 1; \epsilon, \hat{q}_{t+1})$  as a function of  $\epsilon$ . Each curve in the right-hand plot is the predicted efficiency  $\psi(\eta_1^*, \eta_2^*; \epsilon, \hat{r}_{t+1})$ , again as a function of  $\epsilon$ , where we have found optimal continuation probabilities,  $(\eta_1^*, \eta_2^*)$ . For each algorithm, we choose the target efficiency at each generation  $t > 1$  to equal the efficiency observed in generation 1. [Figure 5.5a](#) demonstrates the adaptive threshold selection implemented for four generations of [Algorithm 2.2](#) (ABC-SMC), producing a decreasing sequence for  $\epsilon_t$  of  $2 > 1.21 > 0.83 > 0.51$ . In comparison, [Figure 5.5b](#) shows how the adaptive selection of thresholds in [Algorithm 4.1](#) (MF-ABC-SMC) produces a sequence for  $\epsilon_t$  of  $2 > 0.62 > 0.23 > 0.11$ . Clearly, the adaptive sequence of  $\epsilon_t$  enabled by MF-ABC-SMC decreases much more quickly than the equivalent sequence for ABC-SMC, while the efficiency of each generation remains broadly constant and predictable.

In our example, each generation has a stopping condition of  $\text{ESS} \geq 400$ . As a result, by choosing to specify a constant target efficiency equal to the observed efficiency of generation  $t = 1$ , we effectively impose a constant target simulation budget for each generation. Since (for the example in [Figure 5.5](#)) we have specified four generations for each run of the adaptive versions of [Algorithms 2.2](#) and [4.1](#), we have thus specified a fixed, equal total simulation time for each algorithm. In this setting, the results in [Figure 5.5](#) show that the bias incurred by using an ABC approximation to the posterior with threshold  $\epsilon > 0$  is vastly reduced by implementing the multifidelity approach to SMC. Using MF-ABC-SMC, the sample produced in generation 4 is from  $p_{0.11}(\theta \mid y_{\text{obs}})$ , while ABC-SMC can only produce a sample from  $p_{0.51}(\theta \mid y_{\text{obs}})$  for a similar computational cost. Thus, by incorporating the multifidelity approach into a method for adaptively selecting ABC thresholds, we can afford to allow the sequence  $\epsilon_t$  of MF-ABC-SMC thresholds to decrease much more rapidly, at no cost to the efficiency of the algorithm.

**6. Discussion and conclusions.** In this work we have examined how to integrate two approaches to overcoming the computational bottleneck of repeated simulation within ABC: the SMC technique for producing parameter proposals, and the multifidelity technique for using low-fidelity models to reduce overall simulation time. By combining these approaches, we have produced the MF-ABC-SMC algorithm in [Algorithm 4.1](#).

The results in [Section 5](#) demonstrate that the efficiency of sampling from the ABC posterior (measured as the ratio of the ESS to simulation time) can be significantly improved by using [Algorithm 4.1](#) (MF-ABC-SMC) in place of [Algorithm 2.2](#) (ABC-SMC). This improvement was demonstrated by using a common schedule of decreasing ABC thresholds for both algorithms. In this case, the increase in efficiency was most significant during the early SMC generations, where both the average simulation time of the high fidelity model and overall acceptance rates are relatively large. By also implementing an adaptive ABC thresholding scheme into both algorithms and thus allowing different sequences of ABC threshold, MF-ABC-SMC is shown to greatly reduce the bias incurred by using an ABC approximation to the posterior, in comparison to ABC-SMC.

Having introduced the MF-ABC-SMC algorithm, a number of open questions emerge. Some of these questions are specific to the implementation of multifidelity approaches. However, others arise from a re-evaluation of SMC implementation strategies in this new context. Below, we consider these two classes of question in turn.

**6.1. Multifidelity implementation.** The key to maximising the benefit of MF-ABC-SMC is the ability to set a continuation probability based on the simulations generated during preceding generations. The estimates in [Equation \(4.6\)](#) of the quantities in [Equation \(4.3\)](#) are natural Monte Carlo approximations to the required integrals. In the SMC context, when generating the continuation probability for generation  $t + 1$ , each of these estimates could actually be constructed using the parameter proposals, importance weights, simulations, distances, and continuation probabilities of any (or all) generations  $1 \leq s \leq t$ , not just generation  $t$ . Future work should clarify how best to combine many generations' samples into estimates of [Equation \(4.3\)](#), and the potential for improvement that might arise from this.

Another question arises when we break the assumption made at the start of [Subsection 2.3](#), and no longer assume that the output spaces of each model fidelity are such that  $\tilde{\mathcal{Y}} = \mathcal{Y}$ . In general, the observed data,  $\tilde{y}_{\text{obs}} \neq y_{\text{obs}}$ , the distance metrics,  $\tilde{d}(\tilde{y}, \tilde{y}_{\text{obs}}) \neq d(y, y_{\text{obs}})$ , and the thresholds,  $\tilde{\epsilon} \neq \epsilon$ , may all be distinct. In this case, any estimate,  $\tilde{w}$ , of  $\mathbb{I}(y \in \Omega_\epsilon)$  can be used in place of  $\mathbb{I}(\tilde{y} \in \Omega_\epsilon)$  to give a multifidelity acceptance weight of the form

$$w(\theta, \tilde{y}, u, y) = \tilde{w}(\theta, \tilde{y}) + \frac{\mathbb{I}(u < \alpha(\theta, \tilde{y}))}{\alpha(\theta, \tilde{y})} (\mathbb{I}(y \in \Omega_\epsilon) - \tilde{w}(\theta, \tilde{y})).$$

Note that, in the case of equal output spaces, we might consider using  $\tilde{w}(y_n) = \mathbb{I}(\tilde{y} \in \Omega_{\tilde{\epsilon}})$  for distinct thresholds,  $\tilde{\epsilon} \neq \epsilon$ . However, in general,  $\tilde{w}(\theta, \tilde{y})$  may also encompass completely different output spaces based on distinct modelling frameworks for the same system (albeit with the same parameter space). In a similar way to evolving acceptance probabilities across generations, we could also evolve the estimate  $\tilde{w}$  across generations, using the information gathered from repeated simulation of both high-fidelity and low-fidelity models to better approximate  $\mathbb{I}(y \in \Omega_\epsilon)$  and thus reduce our reliance on the high-fidelity model.

The form of continuation probability,  $\alpha(\theta, \tilde{y})$ , defined in [Equation \(4.1\)](#) implements a multifidelity algorithm that provides a single continuation probability for  $\tilde{y} \in \Omega_\epsilon$  and another for  $\tilde{y} \notin \Omega_\epsilon$ , independently of the parameter value. There may be significant improvements to the multifidelity approach available through making  $\alpha(\theta, \tilde{y})$  depend more generally on  $\theta$  and  $\tilde{y}$ . For example, it is probable that there should be less need to simulate  $y$  after generating  $\tilde{y}$  such

that  $d(\tilde{y}, y_{\text{obs}}) \gg \epsilon$ , than if  $d(\tilde{y}, y_{\text{obs}}) = 1.01\epsilon$ . However, with  $\alpha$  as defined in Equation (4.1), these two cases are treated equally. There is likely to be significant potential for improved performance from exploring less simplistic forms for the continuation probability.

Finally, as has been noted in previous work on multifidelity ABC [28], there is much potential in being able to use multiple low-fidelity models, beyond a single low-fidelity approximation. If there exist multiple low-fidelity models, different generations of MF-ABC-SMC may allow us to progressively focus on using the most efficient model, and identify the specific regions of parameter space for which one model or another may bring most benefit for parameter estimation.

**6.2. SMC implementation.** Previous research into the implementation of ABC-SMC has ensured that the importance distribution formed from the preceding generation, as given in Equation (2.4), is optimal, by choice of the perturbation kernels,  $K_t$  [12]. This has typically been treated as a requirement to trade-off a wide exploration of parameter space against a high acceptance rate. We have replaced the acceptance rate by the theoretical efficiency as the quantification of an ABC algorithm’s performance. Therefore, since we now explicitly include the simulation time in the definition of the algorithm’s performance, the perturbation kernels that optimise the tradeoff between efficiency and exploration may be reformulated and hence different optima used.

In Subsection 5.2 we applied a widely-used strategy for determining the perturbation kernels [3]. This strategy has been justified only in the context of positive weights and an importance distribution that approximates the ABC posterior. However, the importance distribution used in MF-ABC-SMC potentially makes this choice of perturbation kernels suboptimal. It remains to extend existing results on the optimality of perturbation kernels, such as those in [12], to apply to importance distributions of the form in Equation (3.1) that approximate the alternative target distribution in Equation (3.2). There is therefore justification for reopening the question of specifying optimal perturbation kernels for SMC, in the context of both including simulation time in the performance tradeoff and for dealing with multifidelity samples with negative weights.

We have restricted the choice of SMC sampler to the  $O(N^2)$  PMC sampler, due to the effect of the  $O(N)$  sampler in diluting the benefits of the multifidelity approach, as discussed in Sections SM1 and 3. Although this choice can be justified when simulation times dominate the algorithm, future work in this area should seek to optimise a multifidelity approach in the context of the more efficient sampling technique. Given that the observed effect of multifidelity ABC is to significantly reduce the simulation time per parameter proposal, and thereby make the simulation cost at each iteration much *less* dominant, this extension will be necessary to ensure optimal performance of MF-ABC-SMC in larger sample sizes.

In Subsection 5.2.3 we described how to adapt Algorithms 2.2 and 4.1 to implement an adaptive sequence of  $\epsilon_t$  as an approach to minimising bias for a fixed computational budget. The strategy we used was to choose  $\epsilon_t$  to maintain an efficiency as close as possible to a target, set to equal the observed efficiency of the first generation. Further work in this area should investigate the use of more sophisticated strategies for choosing each  $\epsilon_t$ . This question relates closely to the sequence of stopping criteria. In Subsection 5.2, we constrained the effective sample size at each generation to be at least 400 to ensure a relatively low variance in each gen-

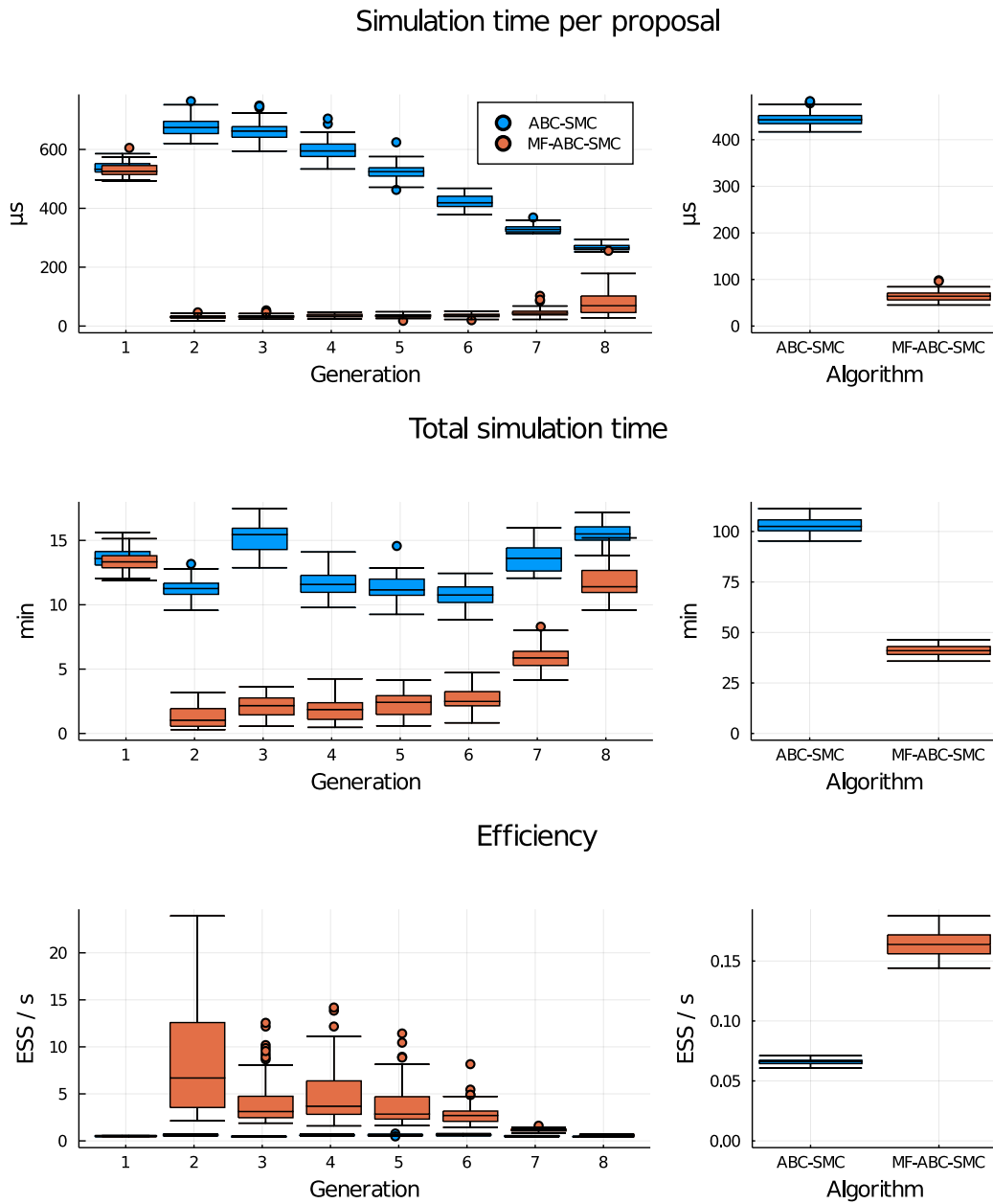
eration's sample, but this choice was made arbitrarily. Future work should therefore consider how to best achieve the ultimate goal of the SMC algorithm: a sample from a final generation with minimal bias relative to the true posterior, with a small variance, and constructed quickly. This goal should be achieved by optimising the complex, interdependent choices of stopping criteria, ABC thresholds, continuation probabilities and perturbation kernels.

## REFERENCES

- [1] J. ALSING, B. D. WANDEL, AND S. M. FEENEY, *Optimal proposals for approximate Bayesian computation*, (2018), <https://arxiv.org/abs/1808.06040v1>.
- [2] M. A. BEAUMONT, *Approximate Bayesian computation in evolution and ecology*, Annual Review of Ecology, Evolution, and Systematics, 41 (2010), pp. 379–406, <https://doi.org/10.1146/annurev-ecolsys-102209-144621>.
- [3] M. A. BEAUMONT, J.-M. CORNUET, J.-M. MARIN, AND C. P. ROBERT, *Adaptive approximate Bayesian computation*, Biometrika, 96 (2009), pp. 983–990, <https://doi.org/10.1093/biomet/asq052>.
- [4] J. BEZANSON, A. EDELMAN, S. KARPINSKI, AND V. B. SHAH, *Julia: A fresh approach to numerical computing*, SIAM Review, 59 (2017), pp. 65–98, <https://doi.org/10.1137/141000671>.
- [5] O. CAPPÈ, A. GUILLIN, J. M. MARIN, AND C. P. ROBERT, *Population Monte Carlo*, Journal of Computational and Graphical Statistics, 13 (2004), pp. 907–929, <https://doi.org/10.1198/106186004X12803>, <https://doi.org/10.1198/106186004X12803>, <https://arxiv.org/abs/https://doi.org/10.1198/106186004X12803>.
- [6] M. CROCI, M. B. GILES, M. E. ROGNES, AND P. E. FARRELL, *Efficient white noise sampling and coupling for multilevel Monte Carlo with non-nested meshes*, (2018), <https://arxiv.org/abs/1803.04857v2>.
- [7] P. DEL MORAL, A. DOUCET, AND A. JASRA, *Sequential Monte Carlo samplers*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68 (2006), pp. 411–436, <https://doi.org/10.1111/j.1467-9868.2006.00553.x>.
- [8] P. DEL MORAL, A. DOUCET, AND A. JASRA, *An adaptive sequential Monte Carlo method for approximate Bayesian computation*, Statistics and Computing, 22 (2012), pp. 1009–1020, <https://doi.org/10.1007/s11222-011-9271-y>.
- [9] V. ELVIRA, L. MARTINO, AND C. P. ROBERT, *Rethinking the effective sample size*, (2018), <https://arxiv.org/abs/1809.04129v1>.
- [10] R. G. EVERITT AND P. A. ROWIŃSKA, *Delayed acceptance ABC-SMC*, (2017), <https://arxiv.org/abs/1708.02230v1>.
- [11] P. FEARHEAD AND D. PRANGLE, *Constructing summary statistics for approximate Bayesian computation: Semi-automatic approximate Bayesian computation*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 74 (2012), pp. 419–474, <https://doi.org/10.1111/j.1467-9868.2011.01010.x>.
- [12] S. FILIPPI, C. P. BARNES, J. CORNEBISE, AND M. P. H. STUMPF, *On optimality of kernels for approximate Bayesian computation using sequential Monte Carlo*, Statistical Applications in Genetics and Molecular Biology, 12 (2013), <https://doi.org/10.1515/sagmb-2012-0069>.
- [13] K. M. HANNAY, D. B. FORGER, AND V. BOOTH, *Macroscopic models for networks of coupled biological oscillators*, Science Advances, 4 (2018), p. e1701047, <https://doi.org/10.1126/sciadv.1701047>.
- [14] J. U. HARRISON AND R. E. BAKER, *An automatic adaptive method to combine summary statistics in approximate Bayesian computation*. 2017, <https://arxiv.org/abs/1703.02341v1>.
- [15] K. E. HINES, *A primer on Bayesian inference for biophysical systems*, Biophysical Journal, 108 (2015), pp. 2103–2113, <https://doi.org/10.1016/j.bpj.2015.03.042>.
- [16] N. JAGIELLA, D. RICKERT, F. J. THEIS, AND J. HASENAUER, *Parallelization and high-performance computing enables automated statistical inference of multi-scale models*, Cell Systems, 4 (2017), pp. 194–206, <https://doi.org/10.1016/j.cels.2016.12.002>.
- [17] A. JASRA, S. JO, D. NOTT, C. SHOEMAKER, AND R. TEMPONE, *Multilevel Monte Carlo in Approximate Bayesian Computation*. arXiv:1702.03628v1, 2017.
- [18] C. LESTER, C. A. YATES, AND R. E. BAKER, *Robustly simulating biochemical reaction kinetics using*

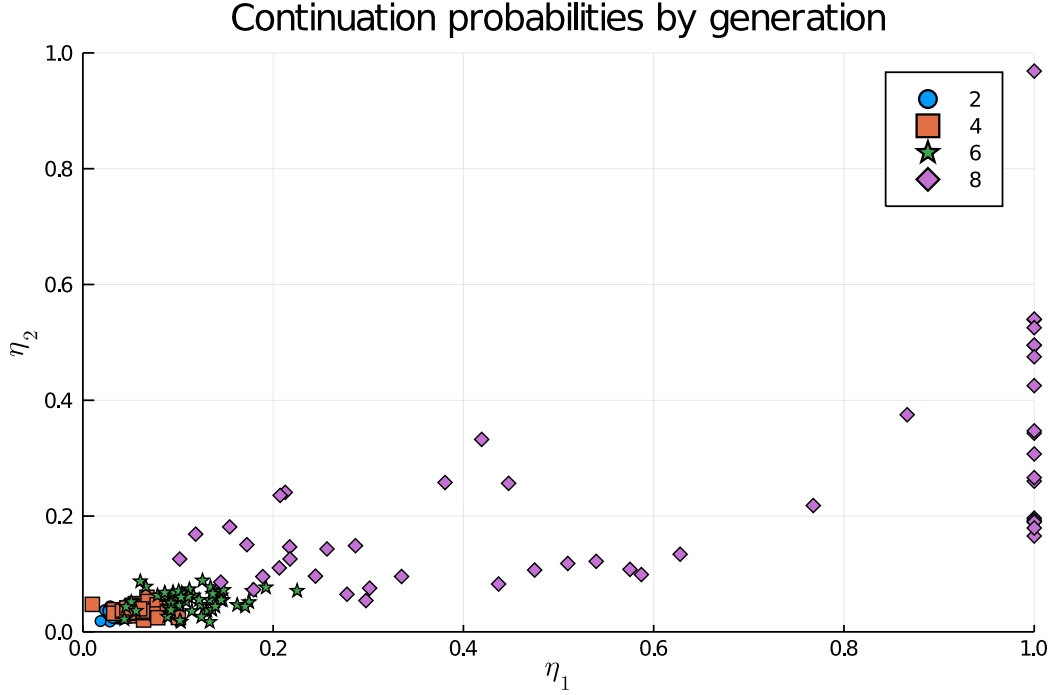


- multi-level Monte Carlo approaches*, Journal of Computational Physics, 375 (2018), pp. 1401–1423, <https://doi.org/10.1016/j.jcp.2018.06.045>.
- [19] J. S. LIU, *Monte Carlo Strategies in Scientific Computing*, Springer New York, 2004, <https://doi.org/10.1007/978-0-387-76371-2>.
  - [20] A.-M. LYNE, M. GIROLAMI, Y. ATCHADÉ, H. STRATHMANN, AND D. SIMPSON, *On Russian Roulette estimates for Bayesian inference with doubly-intractable likelihoods*, Statistical Science, 30 (2015), pp. 443–467, <https://doi.org/10.1214/15-sts523>.
  - [21] E. OTT AND T. M. ANTONSEN, *Low dimensional behavior of large systems of globally coupled oscillators*, Chaos: An Interdisciplinary Journal of Nonlinear Science, 18 (2008), p. 037113, <https://doi.org/10.1063/1.2930766>.
  - [22] E. OTT AND T. M. ANTONSEN, *Long time evolution of phase oscillator systems*, Chaos: An Interdisciplinary Journal of Nonlinear Science, 19 (2009), p. 023117, <https://doi.org/10.1063/1.3136851>.
  - [23] A. B. OWEN, *Monte Carlo theory, methods and examples*, 2013.
  - [24] B. PEHERSTORFER, K. WILLCOX, AND M. GUNZBURGER, *Survey of multifidelity methods in uncertainty propagation, inference, and optimization*, SIAM Review, 60 (2018), pp. 550–591, <https://doi.org/10.1137/16M1082469>.
  - [25] U. PICCHINI AND J. L. FORMAN, *Accelerating inference for diffusions observed with measurement error and large sample sizes using approximate Bayesian computation*, Journal of Statistical Computation and Simulation, 86 (2016), pp. 195–213, <https://doi.org/10.1080/00949655.2014.1002101>.
  - [26] D. PRANGLE, *Lazy ABC*, Statistics and Computing, 26 (2016), pp. 171–185, <https://doi.org/10.1007/s11222-014-9544-3>.
  - [27] D. PRANGLE, *Adapting the ABC distance function*, Bayesian Analysis, 12 (2017), pp. 289–309, <https://doi.org/10.1214/16-BA1002>.
  - [28] T. P. PRESCOTT AND R. E. BAKER, *Multifidelity approximate Bayesian computation*, SIAM/ASA Journal on Uncertainty Quantification, 8 (2020), pp. 114–138, <https://doi.org/10.1137/18M1229742>.
  - [29] D. SCHNOERR, G. SANGUINETTI, AND R. GRIMA, *Approximation and inference methods for stochastic biochemical kinetics—a tutorial review*, Journal of Physics A: Mathematical and Theoretical, 50 (2017), p. 093001, <https://doi.org/10.1088/1751-8121/aa54d9>.
  - [30] S. A. SISSON, Y. FAN, AND M. A. BEAUMONT, *Handbook of Approximate Bayesian Computation*, Chapman & Hall / CRC, 2018.
  - [31] S. A. SISSON, Y. FAN, AND M. M. TANAKA, *Sequential Monte Carlo without likelihoods*, Proceedings of the National Academy of Sciences, 104 (2007), pp. 1760–1765, <https://doi.org/10.1073/pnas.0607208104>.
  - [32] M. SUNNÅKER, A. G. Busetto, E. NUMMINEN, J. CORANDER, M. FOLL, AND C. DESSIMOZ, *Approximate Bayesian computation*, PLOS Computational Biology, 9 (2013), pp. 1–10, <https://doi.org/10.1371/journal.pcbi.1002803>.
  - [33] T. TONI, D. WELCH, N. STRELKOWA, A. IPSEN, AND M. P. H. STUMPF, *Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems*, Journal of the Royal Society Interface, 6 (2009), pp. 187–202, <https://doi.org/10.1098/rsif.2008.0172>.
  - [34] M. N. TRAN AND R. KOHN, *Exact ABC using importance sampling*, (2015), <https://arxiv.org/abs/1509.08076v1>.
  - [35] D. J. WARNE, R. E. BAKER, AND M. J. SIMPSON, *Multilevel rejection sampling for approximate Bayesian computation*, Computational Statistics and Data Analysis, 124 (2018), pp. 71–86, <https://doi.org/10.1016/j.csda.2018.02.009>, <https://arxiv.org/abs/1702.03126v3>.

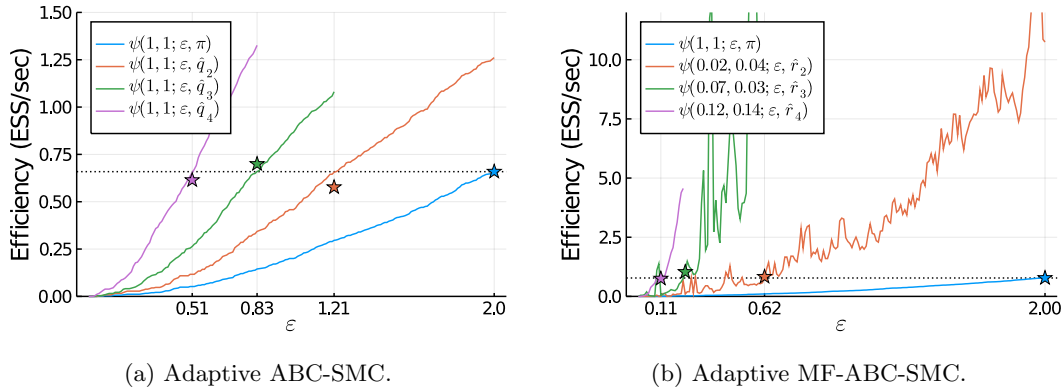


**Figure 5.3.** Observed distributions of simulation time per proposal, total simulation time, and efficiency for 50 samples from [Algorithm 2.2](#) (ABC-SMC, blue) and 50 samples from [Algorithm 4.1](#) (MF-ABC-SMC, orange). Left: Summarised at each SMC generation. Right: Aggregated across all eight generations. Means of each observed distribution are given in [Table 5.2](#).





**Figure 5.4.** Estimated values of the optimal continuation probabilities  $\alpha = \eta_1 \mathbb{I}(\tilde{y} \in \Omega_{\epsilon_t}) + \eta_2 \mathbb{I}(\tilde{y} \notin \Omega_{\epsilon_t})$  used to produce the multifidelity sample for each of the 50 replicates of generations 2, 4, 6 and 8.



**Figure 5.5.** Adaptively selecting  $\epsilon_t$  to achieve a predicted target efficiency equal to the efficiency of the first generation. Curves plot the estimated efficiency as a function of  $\epsilon$  for (a) importance distributions  $\hat{q}_t$ , and (b) importance distributions  $\hat{r}_t$  and associated optimised continuation probabilities. Stars plot observed efficiencies,  $\text{ESS}/T_{\text{tot}}$ , against the selected value of  $\epsilon_t$ .