Approximate Bayesian computation using indirect inference

Christopher C. Drovandi, Anthony N. Pettitt and Malcolm J. Faddy

Queensland University of Technology, Brisbane, Australia

[Received October 2009. Revised October 2010]

Summary. We present a novel approach for developing summary statistics for use in approximate Bayesian computation (ABC) algorithms by using indirect inference. ABC methods are useful for posterior inference in the presence of an intractable likelihood function. In the indirect inference approach to ABC the parameters of an auxiliary model fitted to the data become the summary statistics. Although applicable to any ABC technique, we embed this approach within a sequential Monte Carlo algorithm that is completely adaptive and requires very little tuning. This methodological development was motivated by an application involving data on macroparasite population evolution modelled by a trivariate stochastic process for which there is no tractable likelihood function. The auxiliary model here is based on a beta–binomial distribution. The main objective of the analysis is to determine which parameters of the stochastic model are estimable from the observed data on mature parasite worms.

Keywords: Approximate Bayesian computation; Beta–binomial model; Indirect inference; Macroparasite; Markov process; Sequential Monte Carlo methods

1. Introduction

In approximate Bayesian computation (ABC), we seek to make inferences about the parameters of the posterior distribution when the likelihood function is computationally intractable. This usually occurs when the model that is used to describe the data is complex, e.g. in Markov process models (Tanaka *et al.*, 2006) and population genetic models (Marjoram *et al.*, 2003). Therefore the ABC methodology allows for valid inferences to be carried out on model parameters for increasingly realistic models.

Although the likelihood function itself cannot be computed easily, it is assumed that simulation from the model is relatively straightforward. Data are simulated from the model, $\mathbf{x}$, on the basis of proposed parameter values, which are accepted if the simulated data are sufficiently close to the true data $\mathbf{y}$. There are varying measures of 'closeness'. A popular approach, which was originally proposed by Weiss and von Haeseler (1998), is to compare carefully chosen summary statistics. Say that there are p summary statistics, $\mathbf{S}(\cdot) = (S_1(\cdot), \ldots, S_p(\cdot))^{\mathrm{T}}$; the distance between the true and simulated data, $\rho(\mathbf{y}, \mathbf{x})$, can be obtained by using

$$\rho(\mathbf{y}, \mathbf{x}) = \|\mathbf{S}(\mathbf{y}) - \mathbf{S}(\mathbf{x})\|,$$

for a suitably chosen norm. This method may be expected to be effective when the statistics are sufficient for the parameters of the model, in which case the true posterior is achieved as the discrepancy tends towards 0. However, in many applications sufficient statistics are not available and the practitioner must resort to a selection of carefully chosen data summaries.

Address for correspondence: Christopher C. Drovandi, Mathematical Sciences, Queensland University of Technology, GPO Box 2434, Brisbane 4001, Australia.
E-mail: c.drovandi@qut.edu.au

In this paper we investigate an alternative approach to obtain summary statistics that is based on indirect inference (Heggland and Frigessi, 2004). In indirect inference an auxiliary model is proposed whose likelihood function is tractable and provides a good description of the data. Alternatively, a set of estimating equations can be used (Heggland and Frigessi, 2004). The objective is to search for parameter values of the model of interest that produce simulated data that lead to auxiliary parameter estimates close to those based on maximum likelihood of the original data. Therefore a comparison of summary statistics involves a similarity measure involving estimates of such auxiliary parameters. We use an efficient searching strategy based on an adaptive sequential Monte Carlo (SMC) algorithm yielding a posterior distribution of the parameters that are involved in the model of interest.

We consider a stochastic process model that was developed by Riley *et al.* (2003) for a *Brugia pahangi* population within a host. The *Brugia pahangi* parasite is known to infect domestic cats and wild animals whereas *Brugia malayi* lives almost exclusively in humans. These are examples of microfilarial nematodes that affect lymph vessels, causing what is known as lymphatic filariasis. External symptoms of lymphatic filariasis include enlargement of the leg, arms, genitals, vulva and breasts, which can have significant psychological effects on its sufferers (Garcia, 2007). Furthermore, it is common for lymphatic filariasis to cause internal damage to the kidneys and lymphatic system (Langhammer *et al.*, 2003). The spread of infective larvae (L3) is vector borne via mosquitoes that bite infected individuals. Several moult stages occur within the host, producing L4 and L5 stage larvae, before the parasite matures into an adult (Garcia, 2007).

In the experiment of Denham *et al.* (1972), each host is injected with many L3 larvae which either mature into an adult parasite or die. An individual data point consists of the mature parasite count at the time of necropsy for a particular host. The data, in the form of proportions (mature parasite count at autopsy time divided by the initial larvae injection), exhibit more variability than can be explained by the binomial distribution. A beta–binomial model is employed here to capture the mean and variability and to provide a description of the data, whereas the stochastic model encapsulates the biological system which drives the observed data.

Riley *et al.* (2003) showed that analytic approximations to the likelihood are not adequate in the region of the parameter space of interest and so we adopt a likelihood-free approach based on simulation from the model.

The indirect inference approach to obtaining summary statistics in this application arose as the data are not identically distributed. In particular, each observation has its own mean and variance owing to the presence of the autopsy time and initial juvenile infection covariates. Therefore it is difficult to derive simple summary statistics to capture important aspects of the data. Using the indirect inference approach, we can incorporate these regression effects through an appropriate data analytic model (beta–binomial in this case) and use the corresponding parameter estimates as summary statistics. If the autopsy time and initial injection were fixed for each host then fitting a beta–binomial model would be equivalent to using the sample mean and variance summaries. Drovandi and Pettitt (2010) used a goodness-of-fit statistic but fixed the number of juveniles and marginalized the observed and expected counts over the time variable. In this approach we can consider the full set of data by including the initial infection as a covariate.

Our approach has additional advantages. In most instances of ABC, the summary statistics are simple functions of the data. However, the parameter estimates of a well fitting auxiliary model can be more complex functions of the data and can carry most of the information that is contained in the full data likelihood. Furthermore, the ABC approach that is used here could also be useful in investigating a range of plausible distributions to use in moment closure approximations for Markov-process-type models. For example, Krishnarajah *et al.* (2005) developed a moment closure approximation for stochastic models which is similar to that used here based on the beta–binomial distribution. Our approach using the beta–binomial distribution as an auxiliary model suggests that for these data a moment closure approximation based on the beta–binomial distribution could be of value.

Using this method, a reliable determination can be made of posterior distributions of parameters describing the original stochastic model indicating what is estimable from the observed data. We investigate a range of priors and two different models and find that posterior inferences for ratios of some rate parameters are informative when inferences for the individual rate parameters are quite imprecise. We also informally use the summary statistics to argue that one model is preferable to the other simpler model.

This paper is structured as follows. In Section 2, the macroparasite data are described in more detail. The stochastic model that was proposed by Riley *et al.* (2003) is presented in Section 3. Additionally in this section, the beta–binomial model is developed and fitted to the data of Section 2. Section 4 introduces the concept of using indirect inference within the ABC framework. Section 5 shows the results and a concluding discussion is presented in Section 6.

2. Data

As mentioned previously, the data available for analysis consist of mature parasite counts at particular autopsy times for 212 hosts (Denham *et al.*, 1972). Each host was injected with approximately 100 or 200 larvae and necropsy time ranged between 24 and 1193 days after the initial infection. The proportions of mature parasites alive at autopsy time are shown in Fig. 1. There is clear evidence of overdispersion, which a binomial distribution alone cannot describe.

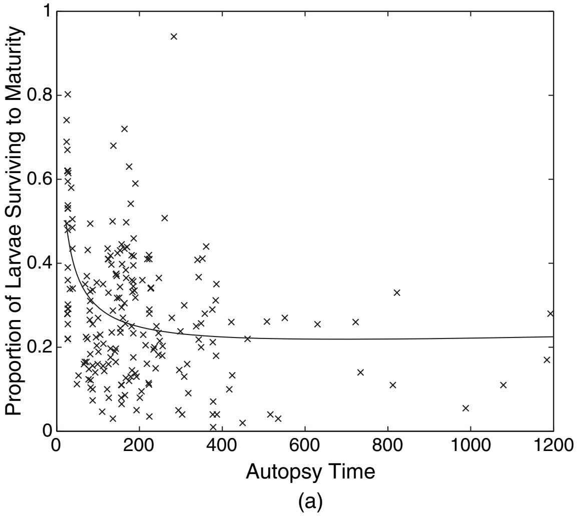
3. Modelling

In this section the stochastic process model of Riley *et al.* (2003) is presented and described. We also outline the fitting of the beta–binomial auxiliary model, which entails an iterative process of model estimation, comparison and evaluation.

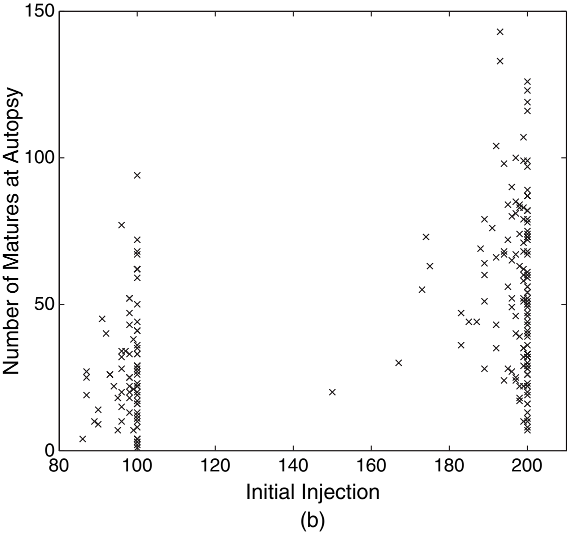
3.1. Stochastic process model

The following stochastic model was developed by Riley *et al.* (2003) to help to explain the population dynamics of *Brugia pahangi*. At time t any host is described by three random variables $\{M(t), L(t), I(t)\}$, where $M(t)$ is the number of mature parasites, $L(t)$ is the number of larvae and $I(t)$ is a discrete version of the host's experience to infection, referred to as the number of units of immunity attained by the host. Initially cats are infected with L_{I} larvae and after a certain time (in days) the hosts are autopsied and the number of mature parasites is counted and recorded. Hence the initial conditions are $M(0)=0$, $L(0)=L_{\mathrm{I}}$ and $I(0)=0$.

It is assumed that each larva matures at a constant rate of γ per day. Larvae die at a rate $\mu_{\mathrm{L}}+\beta I(t)$ per larva where μ_{L} represents the rate at which natural death of larvae occurs and β is a rate parameter that describes additional death of larvae due to the immune response of the host. The acquisition of immunity is assumed to be dependent only on the number of larvae and occurs at rate $\nu L(t)$, and a host loses immunity at a rate μ_{I} per unit of immunity. Mature parasites die at a rate of μ_{M} adults per day. Parameters γ and μ_{M} have been previously estimated at 0.04 (Suswillo *et al.*, 1982) and 0.0015 (Michael *et al.*, 1998) respectively. We consider two models; the first assumes that these two parameters are fixed and the second allows γ to be estimated from the data. Two assumptions that Riley *et al.* (2003) applied in this application



(a)



(b)

Fig. 1. (a) Proportion of mature parasite count to initial larvae infection for each host at their time of autopsy (also shown is the mean of a beta–binomial fit to the data (see Section 3.2)) and (b) initial larvae injected against the number of recorded mature parasites at autopsy: source of data, Denham *et al.* (1972)

are that all cats are independent of each other and that the parameters of the model do not vary between hosts.

The analogous deterministic model can be represented by the following system of interacting differential equations:

$$\left.\begin{aligned}\frac{\mathrm{dL}}{\mathrm{dt}}&=-\mu_{\mathrm{L}}L-\beta IL-\gamma L,\\ \frac{\mathrm{dM}}{\mathrm{dt}}&=\gamma L-\mu_{\mathrm{M}}M,\\ \frac{\mathrm{dI}}{\mathrm{dt}}&=\nu L-\mu_{\mathrm{I}}I,\end{aligned}\right\}\qquad(1)$$

with the initial conditions as before. This approach is valid if there is relatively little variability in the system, which can occur in large populations of parasites. However, as is clear from the data in Fig. 1, there is a substantial amount of variability in the data. Therefore, Riley *et al.* (2003) modelled the data stochastically via a continuous time discrete trivariate Markov process. Given current values of the states at time t, $M(t)=i$, $L(t)=j$ and $I(t)=k$, and a small time interval Δ_t so that at most one event can occur, the transition probabilities at time $t+\Delta_t$ are given by

$$\left.\begin{aligned} P(i+1,j-1,k)&=\gamma j\Delta_t+o(\Delta_t),\\ P(i,j-1,k)&=(\mu_{\mathrm{L}}+\beta k)j\Delta_t+o(\Delta_t),\\ P(i-1,j,k)&=\mu_{\mathrm{M}}i\Delta_t+o(\Delta_t),\\ P(i,j,k+1)&=\nu j\Delta_t+o(\Delta_t),\\ P(i,j,k-1)&=\mu_{\mathrm{I}}k\Delta_t+o(\Delta_t), \end{aligned}\right\} \tag{2}$$

and the probability of remaining in the same state is 1 minus the sum of the above probabilities. Only the final mature parasite count is observed whereas the immunity and larvae counts are unobserved throughout the process. Moreover, the immune response variable $I(t)$ is unbounded. One approach to determining the likelihood is to estimate it empirically by producing many realizations of the stochastic process. Another would consist of bounding the immunity appropriately, yielding a finite state process, and using the matrix exponential formulation of the transition probability matrix (Grimmett and Stirzaker (2001), page 258). Unfortunately both approaches are not computationally feasible because of the large number of states in the process due to the initial number of larvae, about 100 or 200 (see Fig. 1). Analytic approximations to the likelihood by using moment closure were developed in Riley *et al.* (2003) but their accuracy was restricted to a small range of parameter values which, when used in the simulations, could not reproduce the variability of the data. Given these unsuccessful methods for likelihood calculation, an alternative approach uses the likelihood-free and simulation method that is adopted here. The Markov process for each host is simulated independently by using the well-known algorithm of Gillespie (1977).

3.2. *Indirect model*

3.2.1. *Beta–binomial model*

Since the Markov process model assumes that no reproduction occurs at either stage of the life cycle, the number of mature parasites at necropsy times can be calculated by the number of juveniles that matured minus the number that die. Hence each observation represents a proportion. These proportions cannot be modelled appropriately by assuming a binomial distribution owing to the high variability in the data.

To overcome this, we propose a beta–binomial model, which contains an extra parameter to capture the overdispersion. More specifically, the ith observation has the probability distribution

$$P(M_i=m_i|\alpha_i,\beta_i)=\binom{l_i}{m_i}\frac{B(m_i+\alpha_i,l_i-m_i+\beta_i)}{B(\alpha_i,\beta_i)},$$

where l_i is the larvae injection, m_i is the mature parasite count and $B(\cdot,\cdot)$ is the beta function. It is convenient to use a reparameterization in terms of the proportion, $p_i=\alpha_i/(\alpha_i+\beta_i)$, and overdispersion, $\theta_i=1/(\alpha_i+\beta_i)$, parameters, with $\theta=0$ corresponding to the binomial distribution. These parameters can then be related to the necropsy time t_i and initial larvae burden l_i through

$$\begin{aligned} \operatorname{logit}(p_i) &= f_p(t_i, l_i), \\ \log(\theta_i) &= f_\theta(t_i, l_i). \end{aligned} \tag{3}$$

The choice of the link functions logit and log is to some extent arbitrary but appears to work well here.

3.2.2. *Model selection and comparison*

The two functions $f_p(t_i, l_i)$ and $f_\theta(t_i, l_i)$ are chosen appropriately to fit the data. The log-likelihood function can be calculated straightforwardly in this instance and we compute the maximum likelihood estimates of the parameters by using the simplex algorithm (Nelder and Mead, 1965). Model comparison can then proceed by using, for example, the Akaike information criterion AIC.

Initially we considered several choices of the functions $f_p(\cdot)$ and $f_\theta(\cdot)$ of only the autopsy time covariate t_i. Table 1 reveals a subset of these. It can be seen from Table 1 that a linear trend in $f_\theta(t_i)$ provides only slight improvements in fit and is thus neglected. We also tried the complementary log–log-link function as opposed to the logit in model (3) and again there was a minimal change in fit.

The model corresponding to the smallest AIC in Table 1 is that involving linear and quadratic terms in $\log(t_i)$. To improve the model fit further we introduced a second covariate l_i into the model. We found that adding this covariate into the proportion function $f_p(t_i, l_i)$ did not provide much improvement to the fit. However, including this additional covariate into $f_\theta(t_i, l_i)$ did produce an enhanced model fit. We introduced two θ-parameters, θ_{100} and θ_{200}, to account for the overdispersion in the proportion of mature parasite count and initial larvae infection for $l_i \approx 100$ and $l_i \approx 200$ respectively. Therefore

$$\log(\theta_i) = \begin{cases} \eta_{100}, & \text{if } l_i \leqslant 100, \\ \eta_{200}, & \text{if } l_i > 100, \end{cases} \tag{4}$$

where $\eta_{100} = \log(\theta_{100})$ and $\eta_{200} = \log(\theta_{200})$. For this model (with a quadratic mean trend in log-time) the AIC-value was 1897, demonstrating a substantial improvement in model fit. Therefore the indirect model that we choose consists of five parameters, β_0, β_1, β_2, η_{100} and η_{200} (see Fig. 1 for a mean based on these parameters fitted to the experimental data). Residual analysis indicated that this beta–binomial model provides a good fit to the observed data.

Table 1. Set of functions tested for the beta–binomial model together with their AIC-values

$f_p(t_i)$	$f_\theta(t_i)$	*AIC*
$\beta_0 + \beta_1 t_i$	η_0	1930
$\beta_0 + \beta_1 t_i$	$\eta_0 + \eta_1 t_i$	1931
$\beta_0 + \beta_1 t_i + \beta_2 t_i^2$	η_0	1921
$\beta_0 + \beta_1 t_i + \beta_2 t_i^2$	$\eta_0 + \eta_1 t_i$	1922
$\beta_0 + \beta_1 \log(t_i)$	η_0	1911
$\beta_0 + \beta_1 \log(t_i)$	$\eta_0 + \eta_1 t_i$	1912
$\beta_0 + \beta_1 \log(t_i) + \beta_2 \log(t_i)^2$	η_0	1909
$\beta_0 + \beta_1 \log(t_i) + \beta_2 \log(t_i)^2$	$\eta_0 + \eta_1 t_i$	1911

To help to ensure that the regression estimates of the parameters are not highly correlated, the log-time covariate was centred:

$$f_p(t_i, l_i) = \beta_0 + \beta_1\{\log(t_i) - \overline{\log(t)}\} + \beta_2\{\log(t_i) - \overline{\log(t)}\}^2. \tag{5}$$

The maximum likelihood estimates for the parameters with standard errors (based on an estimate of the Hessian) given in parentheses were $\hat{\beta}_0 = -1.002$ (0.062), $\hat{\beta}_1 = -0.341$ (0.057), $\hat{\beta}_2 = 0.108$ (0.048), $\hat{\eta}_{100} = -1.694$ (0.157) and $\hat{\eta}_{200} = -2.434$ (0.127). Much of the information in the data is contained in these parameter estimates, and we use these as summary statistics of the actual data in the ABC algorithm. An estimate of the Hessian matrix at the maximum likelihood estimate generated the correlation matrix

$$\widehat{\text{corr}}(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\eta}_0, \hat{\eta}_1) \approx \begin{pmatrix} 1.0 & -0.12 & -0.60 & 0.10 & 0.16 \\ -0.12 & 1.0 & 0.30 & 0.06 & 0.06 \\ -0.60 & 0.30 & 1.0 & 0.05 & -0.06 \\ 0.10 & 0.06 & 0.05 & 1.0 & 0.02 \\ 0.16 & 0.06 & -0.06 & 0.02 & 1.0 \end{pmatrix}.$$

4. Approximate Bayesian computation

In the usual approach to ABC a target distribution is defined similarly to the posterior distribution but uses the simulated data as an auxiliary variable. The most general form of the approximate posterior distribution is given by

$$\pi(\boldsymbol{\theta}, \mathbf{x}|\mathbf{y}) \propto g(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})\, f(\mathbf{x}|\boldsymbol{\theta})\, \pi(\boldsymbol{\theta}),$$

(Luciani *et al.*, 2009) where $f(\mathbf{x}|\boldsymbol{\theta})$ is the probability model evaluated at the simulated data, $\pi(\boldsymbol{\theta})$ is the prior distribution of the parameter and $g(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$ is some function that assesses the similarity between the observed and the simulated data, giving higher weight as they become closer. One special case is to put $g(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = g(\mathbf{y}|\mathbf{x})$, resulting in a hierarchical structure, $\mathbf{y} \to \mathbf{x} \to \boldsymbol{\theta}$ (Reeves and Pettitt, 2005). Some potentially useful choices for $g(\mathbf{y}|\mathbf{x})$ involve a Gaussian or Epanechnikov kernel weighting function. In many applications $g(\mathbf{y}|\mathbf{x})$ is taken to be proportional to the indicator function (a uniform weighting function), which is 1 if the distance between observed and simulated data (usually based on summary statistics), ρ, is within a certain predefined target threshold ε_T. If this choice is adopted the target distribution simplifies:

$$\pi\{\boldsymbol{\theta}, \mathbf{x}|\rho(\mathbf{y}, \mathbf{x}) \leqslant \varepsilon_T\} \propto f(\mathbf{x}|\boldsymbol{\theta})\, \pi(\boldsymbol{\theta}) \mathbf{1}_{\rho(\mathbf{y},\mathbf{x}) \leqslant \varepsilon_T}. \tag{6}$$

The marginal posterior distribution of the parameters can be approximated by simulating B data sets, $\mathbf{x}_1, \ldots, \mathbf{x}_B$ based on values from the prior and computing

$$\pi\{\boldsymbol{\theta}|\rho(\mathbf{y}, \mathbf{x}) \leqslant \varepsilon_T\} \approx \frac{1}{B} \sum_{i=1}^{B} \mathbf{1}_{\rho(\mathbf{y},\mathbf{x}_i) \leqslant \varepsilon_T}. \tag{7}$$

In the SMC setting it has been advocated that $B = 1$ is a reasonable choice (McKinley *et al.*, 2009) and we adopt such a preference.

4.1. Approximate Bayesian computation algorithms

Several ABC methods have been proposed in the literature. These include acceptance sampling (popularized by Beaumont *et al.* (2002)) and Markov chain Monte Carlo (MCMC) sampling (Marjoram *et al.*, 2003). See Marjoram and Tavaré (2006) for a review. However, we consider

SMC approaches, pioneered by Sisson *et al.* (2007), which have proven to be more efficient than previous methods. In particular, we use the SMC ABC replenishment algorithm of Drovandi and Pettitt (2010), where N weighted samples (particles) are traversed through a sequence of target distributions that have a sequence of non-increasing ABC tolerances.

In this algorithm we have a starting tolerance ε_1 and a target tolerance ε_T. The intermediate tolerances are adaptively determined by dropping a proportion α of the particles with the worst tolerance values at each iteration. The population is replenished by resampling from the 'alive' particles and moved according to an MCMC kernel. We repeat the MCMC step numerous times to ensure that each resampled particle is moved with a theoretical probability of $1-c$. The stopping rule for the algorithm is when the MCMC acceptance rate becomes unacceptably low, and this determines ε_T. See Drovandi and Pettitt (2010) for more details.

4.2. Approximate Bayesian computation with indirect inference

In ABC with indirect inference (see Heggland and Frigessi (2004) as a reference on indirect inference), an auxiliary model is developed with a different set of parameters $\boldsymbol{\theta}_{\mathrm{a}}$ (as opposed to the parameters of the model of interest, $\boldsymbol{\theta}$). It is paramount that the likelihood of this model is tractable and the model provides a good description of the data. Let $\hat{\boldsymbol{\theta}}_{\mathrm{a}}$ denote the parameter estimates that are in some sense optimal when applying the auxiliary model to the true data (e.g. those based on maximum likelihood or posterior modes). These estimates become the summary statistics of the data. It is also important that optimal $\boldsymbol{\theta}_{\mathrm{a}}$ are quickly computable for any simulated data set from the model of interest.

As before, data $\mathbf{x}$ are simulated from the original model by using parameter values of this model. The auxiliary model is fitted to the simulated data, producing parameter estimates $\boldsymbol{\theta}_{\mathrm{a}}^{\mathbf{x}}$. We then compare the summary statistics $\boldsymbol{\theta}_{\mathrm{a}}^{\mathbf{x}}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{a}}$. This can be done, for example, via the Mahalanobis distance

$$\rho(\mathbf{y},\mathbf{x}) = \rho(\hat{\boldsymbol{\theta}}_{\mathrm{a}}, \boldsymbol{\theta}_{\mathrm{a}}^{\mathbf{x}}) = \surd\{(\hat{\boldsymbol{\theta}}_{\mathrm{a}} - \boldsymbol{\theta}_{\mathrm{a}}^{\mathbf{x}})^{\mathrm{T}} \mathbf{S}^{-1} (\hat{\boldsymbol{\theta}}_{\mathrm{a}} - \boldsymbol{\theta}_{\mathrm{a}}^{\mathbf{x}})\},$$

where $\mathbf{S}$ is the covariance matrix of the sampling distribution of the parameter estimates (likelihood-based inference) or parameters (Bayesian inference). Hence the objective of such an algorithm is a search on the parameter space of the model of interest that leads to parameters of the auxiliary model near the auxiliary parameters that are favourable for the actual data.

4.3. Model deficiency in approximate Bayesian computation

Following the approach of Ratmann *et al.* (2009), any deficiencies in the model can be found in ABC by looking at the distribution of the difference between each summary statistic of the simulated and the observed data. This basically describes which summary statistics are contributing the most to the overall discrepancy. In this application the summary statistics are the auxiliary model parameter estimates. These auxiliary parameters can be retained for each particle. Then the distribution of the error between these simulated auxiliary parameters and the auxiliary parameters estimated from the actual data can be determined.

In the ideal situation, the optimal auxiliary parameters tuned to the observed data provide close to sufficient statistics. In this case, the preferred outcome is that the model of interest can generate data sets that reproduce the optimal auxiliary parameters with a modal error of 0. Such an outcome would suggest that these statistics are reproduced unbiasedly as the target tolerance tends towards 0. A discrepancy with a non-zero mode may indicate a model deficiency, or a lack of goodness of fit. It is to be hoped, however, that the discrepancy distribution has a mode close to 0.

5. Results

5.1. Results for the tractable model

Here we consider an artificial example consisting of only 10 initial juveniles. In this way the populations are small so likelihood-based inference is tractable. We set the maximum allowable immunity value at 2 to ensure a finite state process. The transition probability matrix can be computed by using the matrix exponential formulation and the likelihood evaluated by marginalizing over the juvenile and immunity variables. We assume that all parameters are fixed apart from ν, which is given a $U(0, 1)$ prior. We estimate the normalizing constant of the posterior by using the trapezoidal rule to obtain a good approximation to the true posterior.

5.1.1. Data and auxiliary model

The data are simulated on the basis of optimal parameter estimates in Riley *et al.* (2003), i.e. $\nu = 0.000\,84$, $\mu_{\mathrm{I}} = 0.31$, $\mu_{\mathrm{L}} = 0.0011$, $\beta = 1.1$, $\gamma = 0.04$ and $\mu_{\mathrm{M}} = 0.0015$. The autopsy times were evenly spaced between 50 and 545 days with a time increment of 5 days, creating 100 observations. All hosts are infected with 10 juveniles.

The auxiliary model is a beta–binomial model with $\mathrm{logit}(p_i) = \beta_0 + \beta_1 (t_i - \bar{t})$ and $\log(\theta_i) = \eta$. The data together with the mean fit are shown in Fig. 2.

5.1.2. Results

The distributions from the likelihood-based and likelihood-free inference are shown in Fig. 3. The median (and 95% credible interval in parentheses) for the true and ABC posteriors were estimated as 0.000 88 (0.000 39, 0.0016) and 0.000 82 (0.000 35, 0.0016) respectively. It is evident here that the ABC inference compares favourably, demonstrating the utility of the method.

5.2. Results for the observed data

5.2.1. Models and priors

We considered two variants of the stochastic model above. Initially we chose to keep γ and μ_{M}

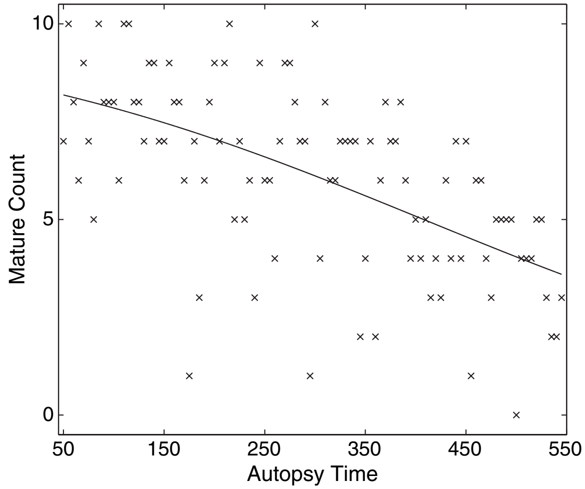


Fig. 2. Simulated data with 10 juveniles to create a tractable likelihood: also shown is the mean of a beta–binomial fit to the data

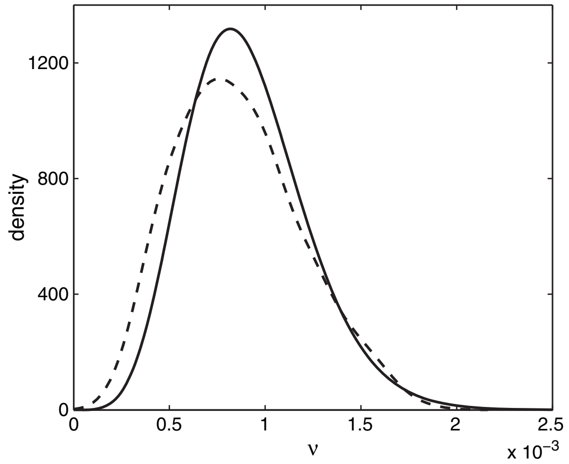


Fig. 3. Posterior distribution for ν when considering the simulated data set that allows likelihood-based inference: shown are the true posterior (———) and that obtained via the indirect inference ABC approach (– – –)

constant at their previously estimated values (referred to as model 1). For the second model, γ was allowed to vary for reasons that will become apparent later (model 2 hereafter). We also tried to estimate μ_M from the data but found an estimate close to the fixed value. Uniform and independent priors were placed on all the parameters. We proposed three sets of prior distributions for each model. In models 1 and 2 the prior distributions for ν and μ_L were all $U(0,1)$. The three prior distributions for β and μ_I were $U(0,2)$, $U(0,5)$ and $U(0,10)$ (i.e. β and μ_I had the same prior). Finally, the prior distributions for γ were $U(0,1)$, $U(0,2.5)$ and $U(0,5)$ (for model 2). This results in six inference cases (two models and three prior configurations).

5.2.2. Algorithm implementation and details

In all cases we used $N=1000$ particles. The MCMC move step was based on a multivariate normal random walk with the elements of the covariance matrix changing adaptively on the basis of sample moments from the particles satisfying the next target distribution. However, the variances were scaled by using a single factor, $0<b<1$, and b was selected so that the eigenvalues of the covariance matrix remained positive. The value of b changed adaptively with the targets as well (we observed values of b roughly in (0.6,1)) and this helped to improve acceptance rates. Additionally, we found that the MCMC move step was more efficient by applying a log-transformation to ν, $\nu^* = -\log(\nu)$. This transformation was appropriate here since the approximate marginal posteriors for ν were skewed for larger tolerances. The MCMC rejection rate tuning parameter c was set at $c=0.01$ (in general, we observed higher rejection rates than this) and we set $\alpha=0.5$.

The first tolerance ε_1 for model 1 was chosen to be 113. This resulted in initial acceptance sampling rates of roughly 50%, 35% and 30% for the three prior configurations. For model 2, an initial tolerance of $\varepsilon_1=50$, $\varepsilon_1=45$ and $\varepsilon_1=45$ resulted in initial acceptance rates of around 50%, 65% and 80% respectively for the three prior configurations. It is quite clear that allowing γ to vary initially generated better fitting simulated data. We note that posterior inference based on the tolerance ε_T is not sensitive to the choice of ε_1 but experience suggests that it is useful to eliminate at least some of the parameter space early. The final tolerance was set to $\varepsilon_T=4.9$ and $\varepsilon_T=3.4$ for models 1 and 2 respectively. This tolerance was determined as the MCMC acceptance rate fell to about 2–3%. A smaller final discrepancy for model 2 would appear to give an indication of a preference for this model over model 1; however, we provide more details of the model comparison later in this section.

5.2.3. Posterior inferences

We now turn to the posterior inferences of the parameters of model 1: ν and μ_{L} are precisely estimated whereas β and μ_{I} are poorly determined (providing motivation for testing sensitivity to the prior). The posterior summaries for each prior configuration can be found in Table 2. Fortunately, inferences for ν were insensitive to the prior. There was a slight shift to the right for the posterior of μ_{L} as the upper limit of β and μ_{I} is increased (with a mode shifting slightly to the left, indicating an increasingly skewed distribution), but it still remains relatively unaffected

Table 2. Posterior summaries for models 1 and 2†

Model	*Prior*	*Parameter*	*Mode*	*Mean*	*Standard deviation*	*(2.5%,50%,97.5%) quantiles*
1	1	ν‡	0.13	0.13	0.03	(0.07,0.13,0.20)
1	1	μ_{I}	1.08	1.03	0.47	(0.15,1.02,1.88)
1	1	μ_L‡	0.55	0.85	0.60	(0.04,0.73,2.35)
1	1	β	1.34	1.20	0.44	(0.34,1.22,1.96)
1	1	μ_{I}/β	0.83	0.85	0.27	(0.33,0.85,1.43)
1	2	ν‡	0.13	0.14	0.03	(0.07,0.14,0.21)
1	2	μ_{I}	3.08	2.76	1.12	(0.58,2.78,4.76)
1	2	μ_L‡	0.43	0.88	0.62	(0.05,0.81,2.47)
1	2	β	3.25	2.96	1.10	(0.73,3.04,4.83)
1	2	μ_{I}/β	0.92	0.95	0.26	(0.44,0.94,1.49)
1	3	ν‡	0.14	0.14	0.03	(0.07,0.14,0.21)
1	3	μ_{I}	6.38	5.70	2.22	(1.43,5.79,9.58)
1	3	μ_L‡	0.50	0.89	0.65	(0.06,0.74,2.50)
1	3	β	5.91	6.01	2.22	(1.56,6.05,9.69)
1	3	μ_{I}/β	1.02	0.98	0.27	(0.36,0.99,1.53)
2	1	ν‡	0.99	1.09	0.41	(0.44,1.04,2.04)
2	1	μ_{I}	0.54	0.73	0.46	(0.04,0.67,1.74)
2	1	μ_L‡	1.50	3.76	2.85	(0.19,3.13,10.5)
2	1	β	1.85	1.60	0.30	(0.89,1.65,1.99)
2	1	γ	0.32	0.34	0.11	(0.15,0.34,0.57)
2	1	ν/γ‡	3.05	3.21	0.66	(2.08,3.14,4.61)
2	1	μ_{I}/β	0.43	0.46	0.28	(0.02,0.43,1.02)
2	2	ν‡	2.30	2.79	1.05	(1.13,2.63,5.17)
2	2	μ_{I}	0.91	1.78	1.15	(0.14,1.63,4.44)
2	2	μ_L‡	5.38	9.92	7.25	(0.42,8.42,28.9)
2	2	β	4.64	4.09	0.69	(2.43,4.24,4.96)
2	2	γ	0.80	0.89	0.28	(0.37,0.87,1.47)
2	2	ν/γ‡	2.85	3.15	0.64	(2.06,3.09,4.58)
2	2	μ_{I}/β	0.23	0.44	0.28	(0.03,0.40,1.01)
2	3	ν‡	4.74	5.53	2.20	(2.24,5.27,10.5)
2	3	μ_{I}	2.51	3.42	2.25	(0.15,3.14,8.83)
2	3	μ_L‡	8.87	19.4	14.6	(0.69,16.4,56.3)
2	3	β	9.16	8.10	1.46	(4.15,8.42,9.91)
2	3	γ	1.43	1.79	0.62	(0.75,1.74,3.09)
2	3	ν/γ‡	3.01	3.11	0.64	(2.00,3.06,4.53)
2	3	μ_{I}/β	0.40	0.42	0.27	(0.02,0.41,1.00)

†Shown are the posterior mode, mean, standard deviation and the (2.5%,50%,97.5%) quantiles.

‡Estimates for these parameters have been multiplied by 100.

by the prior. Even though β and μ_{I} cannot be identified individually, the ratio β/μ_{I} is less sensitive to the prior distribution. (Note that the prior distribution of the ratio of two independent uniform variates on $(0,a)$ and $(0,b)$ is flat on $(0,a/b)$ and decays thereafter). The scatter plots of these two parameters for the three prior distributions are shown in Fig. 4 (the posteriors have been scaled to the unit square and β is shifted upwards to distinguish between the three priors). It can be seen from Fig. 4 that μ_{I} and β are positively correlated and the scaled joint densities are insensitive to the prior. This suggests that the ratio of these parameters can be estimated more precisely than the individual parameters. The posterior distributions from using a prior of $U(0,2)$ for β are shown in Fig. 5 as an example.

Simulations from the model at the posterior modes for this model are insensitive to the prior. These simulations indicate that the immunity variable in the large majority of cases only ever assumes the values of 0 or 1 throughout the simulation. This is largely due to the very small estimated rate of acquisition of immunity. Furthermore, we found that the immunity realizes the value 1 only rarely during the simulation and remains at 1 for a relatively short period of time. Therefore there is no immune response for the large majority of the simulations, which thus provide little information for the parameters β and μ_{I} (both of which are interpreted as rates per unit of immunity). This explains the relatively uninformative posterior distributions for these parameters. These two parameters are coupled in the stochastic model in that an increase in μ_{I} reduces the immunity and the effect of this on the larvae can be compensated by a corresponding increase in β. The dependence of these parameters on the priors for μ_{I} and β reflects the invariance of the model to the simultaneous scaling of these parameters.

During the short time intervals where the immunity is at a level of 1, the immunity has a devastating effect on the juveniles, and it is the time of onset of this immunity which produces highly variable mature parasite counts (see Riley *et al.* (2003) for more details).

In model 2, γ (the maturation rate) is allowed to vary. From Table 2, γ is estimated to be significantly larger than 0.04. This increase in γ is offset by an increase in ν (acquisition of immunity), μ_{L} (natural larvae death) and β (death of larvae due to immunity), which means that a loss of larvae event is not always a maturation event (it is evident from the sample correlation matrix

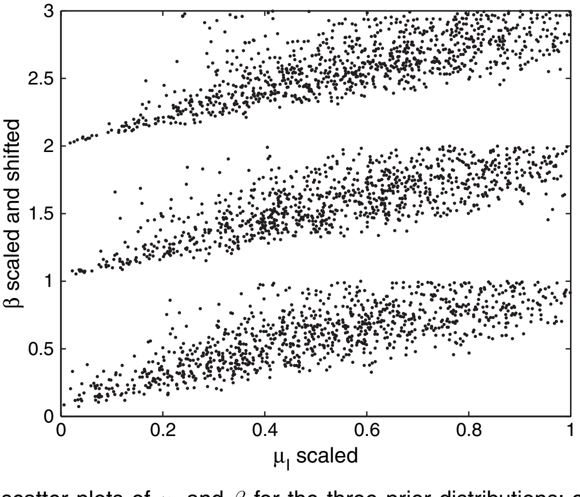


Fig. 4. Joint posterior scatter plots of μ_{I} and β for the three prior distributions: each posterior has been scaled to the unit square and β has been shifted upwards to distinguish between the three priors (the bottom plot is based on prior 1)

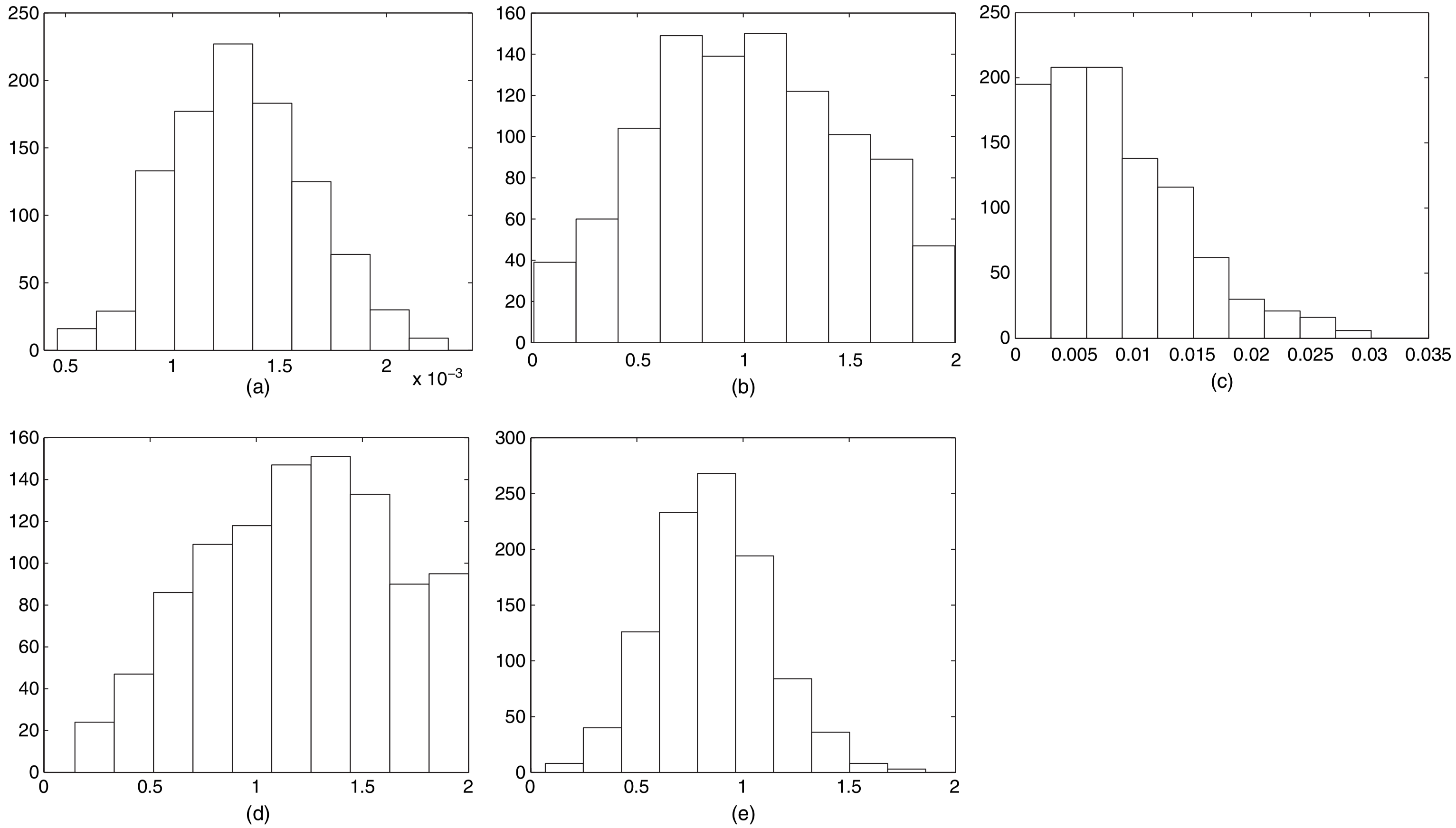


Fig. 5. Posterior densities for model 1 based on the first prior configuration: (a) posterior for ν; (b) posterior for μ_{I}; (c) posterior for μ_{L}; (d) posterior for β; (e) posterior for μ_{I}/β

of the posterior that γ is unsurprisingly positively correlated with each of these parameters, in particular ν). Unfortunately, owing to correlated parameters and imprecisely determined μ_{I} and β, the posteriors are very sensitive to the three prior configurations chosen here, as is apparent in Table 2. Therefore it is not possible in the analysis that is performed here to obtain informative inferences on all the parameters of this model. It is, however, possible to obtain reliable posterior distributions of ratios of these parameters. For example, we found that ν/γ was relatively unaffected by the priors, and again μ_{I}/β was relatively insensitive to these priors. The posterior distributions for this model by using the first prior configuration are shown in Fig. 6 as an example.

5.2.4. *Model sensitivities*

With an increased rate of acquisition of immunity it is now more plausible to obtain an immunity count greater than 1 but the immunity counts are still very low and 0 most of the time. Owing to increased rates of ν, μ_{L} and β in model 2, the juvenile parasite population dies off much faster than in model 1. However, to ensure that mature parasites develop during this time the maturation rate is significantly larger. Similarly, the rates that are associated with prior configuration 3 are larger than the rates that are associated with prior configuration 1 and thus the maximum mature parasite count is reached more quickly in the former instance.

To analyse this phenomenon in more detail we recorded the results of means from 10000 simulations from model 2 at the posterior modes using prior configurations 1 and 3. The trajectories of the juvenile larvae population and the immunity are quite different when these two posterior modes are considered. The mature parasite count trajectories are quite different before 10 days have elapsed but are similar thereafter. Therefore, the mature parasite counts that are generated by both parameter sets are consistent during the observed necropsy times (at least 24 days). Therefore the parameters would be expected to be better identified and less sensitive to their prior distributions if some data were collected at earlier necropsy times. Furthermore, it would be useful if additional information were available on the number of juveniles (although this appears to be difficult in practice). Finally, expert biological knowledge on plausible (or implausible) values of the parameters would assist in the identification process. We assess the effect on inferences for the parameters if mature parasite counts were available at earlier necropsy times in Section 5.3.

5.2.5. *Model choice and goodness of fit*

Regardless of the prior configuration, simulations of mature parasite counts (at least in the time span of the observed data) from each model were similar. In Fig. 7, 95% predictions intervals (by using parameter estimates from posterior modes) were obtained for model 1 (Fig. 7(a)) and model 2 (Fig. 7(b)) shown up to 300 days (predictions were similar thereafter). It can be seen from Fig. 7 that both models provide a reasonable fit to the data. The model 1 parameters predict a maximum peak mature parasite proportion at around day 70 and cannot explain a few high mature parasite counts that occur around days 20–30 (providing motivation for allowing the maturation rate to be tuned to the data). Model 2 suggests that the maximum mature parasite count is reached at or before 24 days but the prediction intervals do not encompass some of the low mature parasite counts observed that are between roughly 100 and 400 days. Such observations could be explained by an underrecording of the true mature parasite counts.

Finally, we consider the final tolerances that are reached by each model and the individual discrepancies of the auxiliary model that are contributing to these tolerances. The final tolerance for model 1 was 4.9, and 3.4 for model 2. Since model 1 is a special case of model 2, we

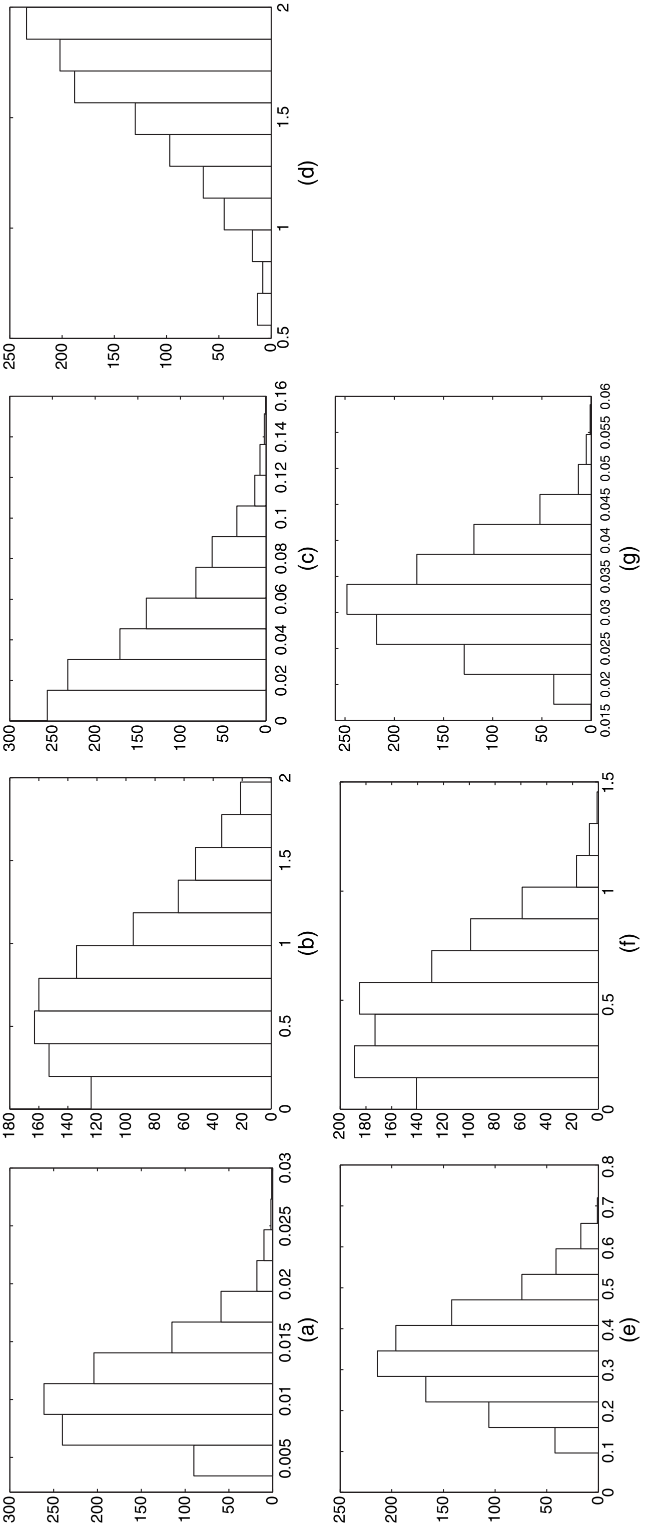


Fig. 6. Posterior densities for model 2 based on the first prior configuration: (a) posterior for ν; (b) posterior for μ_{I}; (c) posterior for μ_{L}; (d) posterior for β; (e) posterior for γ; (f) posterior for μ_{I}/β; (g) posterior for ν/γ

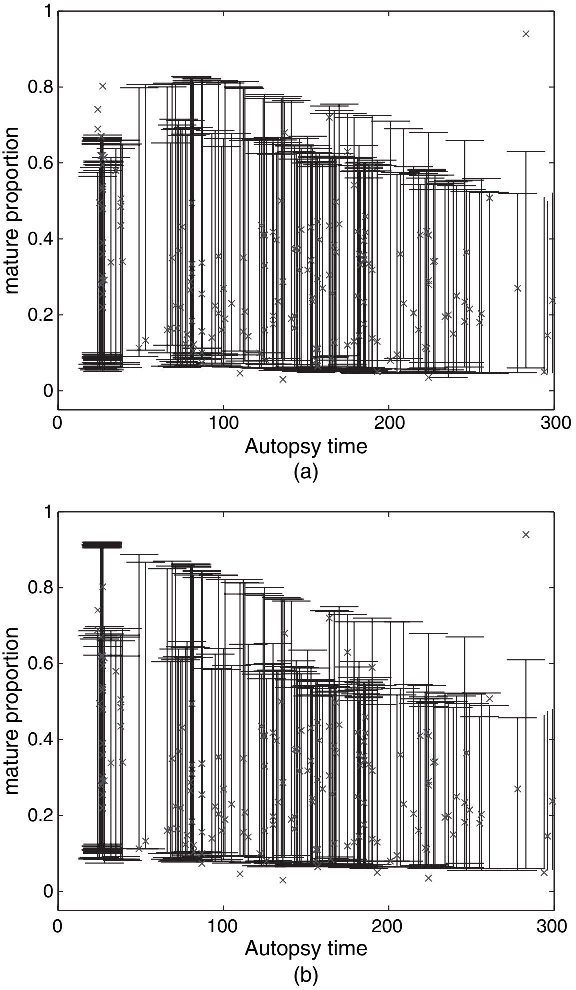


Fig. 7. 95% prediction intervals for (a) model 1 and (b) model 2 based on prior configuration 3

would expect model 2 to be able to achieve a lower tolerance. Therefore it may appear difficult to compare models by using such a criterion since there is no explicit penalty for the extra parameter of model 2. However, the stopping rule that we used for the SMC ABC algorithm was that the MCMC acceptance rate fell to around 2–3%. Therefore there is an implicit penalty for model 2 since we can expect a lower MCMC acceptance rate as there is an extra parameter in the model space. On the basis of this argument, we suggest that model 2 be preferred over model 1.

Each particle in the SMC approximation contains values of the stochastic model parameters and the indirect model parameter values that are estimated from the simulated data (generated by the stochastic model parameters). We can obtain information on which of the auxiliary parameters of each of the models is contributing to the overall tolerance by considering the distributions of the differences between each of the particle's auxiliary parameter values and the optimal indirect parameter estimates that were obtained in Section 3.2.2. It is apparent that neither model can reproduce the estimate of the parameter β_2 (although model 2 does slightly better). The raw data in Fig. 1 show a tendency for the mean to increase slightly near the end of the observation period, and this is possibly due to a paucity of observations at late necropsy times. At this stage of the process the stochastic model can only predict a non-increasing count of mature parasites so this represents a slight inconsistency between the raw data and the stochastic model. All other discrepancies at least include 0 in their distributions. Model 2 can more precisely obtain the parameters β_0 and η_{200} of the indirect model; therefore model 2 can account for the right amount of variability when the initial larvae count is approximately 200. Furthermore, the discrepancy of η_{200} for model 2 is centred on zero. However, the discrepancy of β_1 for model 1 is approximately centred about zero and hence this model reproduces this summary statistic more consistently than model 2.

5.3. Information from early time points

The results of Section 5.2.4 suggest that data collected at earlier times than 24 days would provide more information about the parameters whose posteriors were relatively uninformative.

5.3.1. Data and auxiliary model

In this section we use simulated data to assess the effect on inferences if mature parasite counts were available at such earlier time points. In total, 100 independent observations were simulated by using the parameter estimates of Riley *et al.* (2003), $\nu = 0.00084$, $\mu_I = 0.31$, $\mu_L = 0.0011$, $\beta = 1.1$, $\gamma = 0.04$ and $\mu_M = 0.0015$, all with an initial larvae burden of 200. Five observations were recorded at each necropsy time of 1, 2, 5, 10, 15, 20, 30, 40, 50, 60, 80, 100, 120, 140, 160, 180, 200, 300, 400 and 500 days, so that 30 observations were collected before day 24 which were not available in the original experimental data. We fitted a beta–binomial distribution to these simulated data where the functions $f_p(t_i, l_i)$ and $f_\theta(t_i, l_i)$ in model (3) were both quadratic in log-time, producing six parameter estimates ($\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ for $f_p(t_i, l_i)$ and $\hat{\eta}_0$, $\hat{\eta}_1$ and $\hat{\eta}_2$ for $f_\theta(t_i, l_i)$). The mean fit of this model together with the data is shown in Fig. 8.

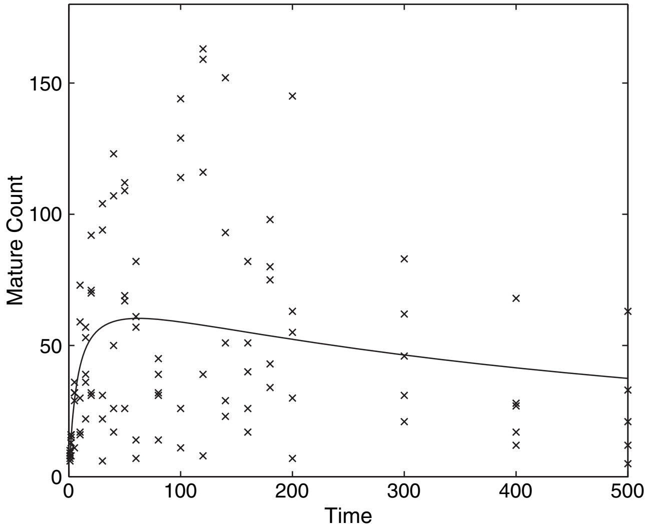


Fig. 8. Simulated data with a mean fitted by using a beta–binomial distribution (see the text for more details)

5.3.2. Results

We applied the SMC ABC replenishment algorithm to investigate the reproducibility of the first five parameters while holding μ_{M} fixed. The results below were based on the algorithm tuning parameters of $N=1000$, $\alpha=0.5$, $c=0.01$, $\varepsilon_1=40$ and $\varepsilon_T=2$. We again considered three prior configurations. The individual prior distributions for ν, μ_{L} and γ were the same over the three prior configurations. These were uniform on (0,0.5), (0,0.5) and (0,1). μ_{I} and β had the same individual priors but they changed over the three prior configurations. These were uniform over (0,2), (0,5) and (0,10) for the three prior configurations.

The posterior summaries for the three prior configurations are shown in Table 3. Unfortunately, again μ_{I} and β are not identified precisely owing to the low immunity issue that was discussed in Section 5.2.3. However, inferences about ν and γ are insensitive to the priors, at least compared with those from the original data set. This indicates that the addition of data at earlier necropsy times is beneficial and provides additional information for these two parameters. Furthermore, the true values of these parameters are recovered relatively accurately, demonstrating the utility of the estimation methodology. The posterior for μ_{L} increased in width and shifted slightly to the right as the upper limits on the priors for μ_{I} and β increased. However, inferences for this parameter are far less sensitive to the priors in comparison with the original data set (when five parameters were considered). This shows that the data at early necropsy times support more precise inference for this parameter, but additional accuracy would almost certainly be obtained if larvae counts were also available.

The stochastic process model in this case could capture the main characteristics of the data described by the beta–binomial distribution owing to modal discrepancies that are close to 0 as can be seen in Fig. 9.

Table 3. Posterior summaries for the simulated data example†

Prior	*Parameter*	*Mode*	*Mean*	*Standard deviation*	*(2.5%,50%,97.5%) quantiles*
1	ν‡	0.06	0.07	0.02	(0.03,0.06,0.11)
1	μ_{L}‡	0.67	1.08	0.71	(0.05,0.97,2.70)
1	β	1.63	1.41	0.36	(0.66,1.44,1.97)
1	γ‡	3.74	3.78	0.43	(3.03,3.76,4.67)
1	μ_{I}/β	0.23	0.26	0.17	(0.02,0.24,0.66)
2	ν‡	0.06	0.07	0.02	(0.03,0.06,0.11)
2	μ_{I}	0.74	1.12	0.74	(0.09,0.96,3.00)
2	μ_{L}†	1.05	1.29	0.83	(0.08,1.20,3.15)
2	β	3.28	3.35	0.98	(1.25,3.41,4.84)§
2	γ†	3.69	3.81	0.44	(3.00,3.78,4.80)
2	μ_{I}/β	0.27	0.33	0.18	(0.03,0.31,0.72)
3	μ_{I}	1.84	2.27	1.29	(0.24,2.14,5.00)
3	μ_{L}‡	1.32	1.37	0.86	(0.09,1.30,3.26)
3	β	8.43	6.76	2.05	(2.24,6.87,9.77)§
3	γ‡	3.78	3.76	0.42	(2.96,3.74,4.63)
3	μ_{I}/β	0.31	0.34	0.17	(0.04,0.34,0.69)

†Shown are the posterior mode, mean, standard deviation and the (2.5%, 50%,97.5%) quantiles.

‡Estimates for these parameters have been multiplied by 100.

§The credible interval does not contain the true value of the corresponding parameter.

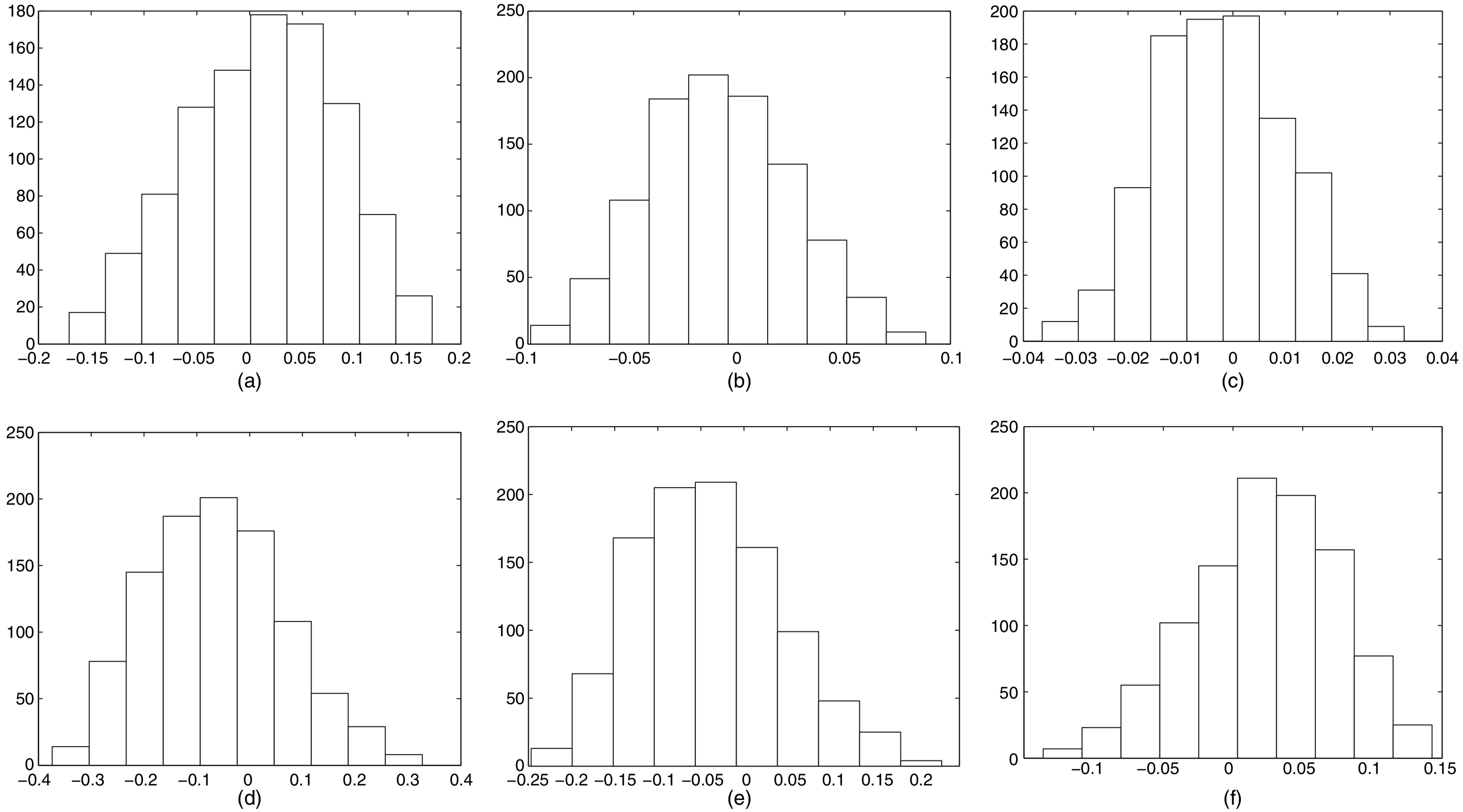


Fig. 9. Discrepancies of the beta–binomial auxiliary parameter estimates when the stochastic model involving five parameters is fitted to the simulated data (the results are based on the first prior configuration): here the discrepancies are for (a) β_0, (b) β_1, (c) β_2, (d) η_0, (e) η_1 and (f) η_2

6. Discussion

In this paper we presented an approach for obtaining summary statistics for use in ABC algorithms based on indirect inference. A simpler yet flexible model is proposed that has a tractable likelihood function, and it is the estimates of the parameters of this indirect model from fitting to the observed data that become the summary statistics. The motivation for such an innovation was to analyse the full parasite data set by using the Riley *et al.* (2003) stochastic model, rather than the subset of the data that were included in Drovandi and Pettitt (2010), as the goodness-of-fit approach that was used in Drovandi and Pettitt (2010) could only accommodate data with a fixed number of initial larvae and thus a substantial amount of the data had to be discarded. The approach that we have presented here is suitable for any application without independent and/or identically distributed data where an indirect model can be proposed whose likelihood function is available and where suitable parameter estimates can be obtained straightforwardly.

This procedure was computationally intensive as it involved model simulation and a likelihood maximization for each new parameter proposal. We used the Nelder–Mead method for maximization owing to its simplicity and the convenient property that the algorithm is derivative free. For the beta–binomial model that was used here the log-likelihood derivatives are analytically available and so it is possible for a Newton-type maximization approach to be applied. An alternative to maximum likelihood estimation for this model could be based on the method of moments. This technique would be computationally much quicker than maximizing the likelihood, but it is not immediately clear how parameter estimates that are generated in this way would compare with those obtained by maximum likelihood.

We simulated data at exactly the same necropsy times as the observed data set. The speed of the simulation step would be enhanced by simulating fewer observations, since there is no requirement for the observed and simulated data sets to be of the same dimension. This would be particularly useful at higher tolerances in the ABC algorithm but the size of the simulated data set should be increased as lower tolerances are reached; otherwise additional uncertainty is introduced. Furthermore, an experimental design that produces optimal necropsy times for a given sized data set may have positive effects, such as fewer observations being required to produce precise auxiliary model parameter estimates.

There appear to be some similarities between ABC and Bayesian calibration of computer models. The likelihood function of the computer model is typically intractable. However, as in Henderson *et al.* (2009), the intractable part of the model is not the observed data likelihood but rather the observed data being dependent on the simulated data. The evaluation of the likelihood can be avoided in an MCMC setting by drawing independent proposals from this likelihood, in the same way as MCMC ABC. But technically no ABC step (i.e. comparing summary statistics) is required as the observed data likelihood given the simulated data is tractable and is present in the Metropolis–Hastings ratio. Furthermore, often simulation from the computer model is also time consuming and an emulator is required as an approximation to the computer model by using a Gaussian process (Kennedy and O'Hagan, 2001).

Acknowledgements

The authors thank Edwin Michael and David Denham for access to the data. The authors also thank Chris Glasbey for bringing the work of Riley *et al.* (2003) to their attention in the context of ABC and Steven Riley for his helpful communications on his work. The authors are also grateful to the high performance computing and research support team of Queensland University of Technology for their assistance on supercomputing issues. The comments of an

Associate Editor and two referees led to improvements in this paper, especially the presentation. The research of all three authors was supported by the Australian Research Council.

References

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

Denham, D., Ponnudurai, T., Nelson, G., Guy, F. and Rogers, R. (1972) Studies with *Brugia pahangi:* I, parasitological observations on primary infections of cats (*Felis catus*). *Int. J. Parasitol.*, **2**, 239–247.

Drovandi, C. C. and Pettitt, A. N. (2010) Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. *Biometrics*, to be published, doi: 10.1111/j.1541-0420.2010.01410.x.

Garcia, L. S. (2007) *Diagnostic Medical Parasitology*. Washington DC: American Society for Microbiology.

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

Grimmett, G. and Stirzaker, D. (2001) *Probability and Random Processes*, 3rd edn. New York: Oxford University Press.

Heggland, K. and Frigessi, A. (2004) Estimating functions in indirect inference. *J. R. Statist. Soc.* B, **66**, 447–462.

Henderson, D. A., Boys, R., Krishnan, K. J., Lawless, C. and Wilkinson, D. J. (2009) Bayesian emulation and calibration of a stochastic computer model of mitochondrial dna deletions in substantia nigra neurons. *J. Am. Statist. Ass.*, **104**, 76–87.

Kennedy, M. C. and O'Hagan, A. (2001) Bayesian calibration of computer models (with discussion). *J. R. Statist. Soc.* B, **63**, 425–464.

Krishnarajah, I., Cook, A., Marion, G. and Gibson, G. (2005) Novel moment closure approximations in stochastic epidemics. *Bull. Math. Biol.*, **67**, 855–873.

Langhammer, J., Birk, H. W. and Zahner, H. (2003) Renal disease in lymphatic filariasis: evidence for tubular and glomerular disorders at various stages of the infection. *Trop. Med. Int. Hlth*, **2**, 875–884.

Luciani, F., Sisson, S. A., Jiang, H., Francis, A. R. and Tanaka, M. M. (2009) The epidemiological fitness cost of drug resistance in *Mycobacterium tuberculosis*. *Proc. Natn. Acad. Sci. USA*, **106**, 14711–14715.

Marjoram, P., Molitor, J., Plagonal, V. and Tavaré, S. (2003) Markov chain Monte Carlo without likelihoods. *Proc. Natn. Acad. Sci. USA*, **100**, 15324–15328.

Marjoram, P. and Tavaré, S. (2006) Modern computational approaches for analysing molecular genetic variation data. *Nature Rev. Genet.*, **7**, 759–770.

McKinley, T., Cook, A. R. and Deardon, R. (2009) Inference in epidemic models without likelihoods. *Int. J. Biostatist.*, **5**, article 24.

Michael, E., Grenfell, B., Isham, V., Denham, D. and Bundy, D. (1998) Modelling variability in lymphatic filariasis: macro filarial dynamics in the *Brugia pahangi* cat model. *Proc. R. Soc. Lond.* B, **265**, 155–165.

Nelder, J. and Mead, R. (1965) A simplex method for function minimization. *Comput. J.*, **7**, 308.

Ratmann, O., Andrieu, C., Wiuf, C. and Richardson, S. (2009) Model criticism based on likelihood-free inference with an application to protein network evolution. *Proc. Natn. Acad. Sci. USA*, **106**, 10576–10581.

Reeves, R. W. and Pettitt, A. N. (2005) A theoretical framework for approximate Bayesian computation. In *Proc. 20th Int. Wrkshp Statistical Modelling, Sydney* (eds A. R. Francis, K. M. Matawie, A. Oshlack and G. K. Smyth), pp. 393–396. Sydney: University of Western Sydney.

Riley, S., Donnelly, C. L. and Ferguson, N. M. (2003) Robust parameter estimation techniques for stochastic within-host macroparasite models. *J. Theoret. Biol.*, **225**, 419–430.

Sisson, S., Fan, Y. and Tanaka, M. (2007) Sequential Monte Carlo without likelihoods. *Proc. Natn. Acad. Sci. USA*, **104**, 1760–1765.

Suswillo, R., Denham, D. and McGreevy, P. (1982) The number and distribution of *Brugia pahangi* in cats at different times after primary infection. *Acta Trop.*, **39**, 151–156.

Tanaka, M., Francis, A., Luciani, F. and Sisson, S. (2006) Estimating tuberculosis transmission parameters from genotype data using approximate Bayesian computation. *Genetics*, **173**, 1511–1520.

Weiss, G. and von Haeseler, A. (1998) Inference of population history using a likelihood approach. *Genetics*, **149**, 1539–1546.