Research

Author for correspondence:
E. L. Ionides
e-mail: ionides@umich.edu

Monte Carlo profile confidence intervals for dynamic systems

E. L. Ionides[1], C. Breto[1], J. Park[1], R. A. Smith[2] and A. A. King[3,4]

[1]Department of Statistics, [2]Department of Bioinformatics, [3]Department of Ecology and Evolutionary Biology, and [4]Department of Mathematics, The University of Michigan, Ann Arbor, MI, USA

(iD) ELI, 0000-0002-4190-0174

Monte Carlo methods to evaluate and maximize the likelihood function enable the construction of confidence intervals and hypothesis tests, facilitating scientific investigation using models for which the likelihood function is intractable. When Monte Carlo error can be made small, by sufficiently exhaustive computation, then the standard theory and practice of likelihood-based inference applies. As datasets become larger, and models more complex, situations arise where no reasonable amount of computation can render Monte Carlo error negligible. We develop profile likelihood methodology to provide frequentist inferences that take into account Monte Carlo uncertainty. We investigate the role of this methodology in facilitating inference for computationally challenging dynamic latent variable models. We present examples arising in the study of infectious disease transmission, demonstrating our methodology for inference on nonlinear dynamic models using genetic sequence data and panel time-series data. We also discuss applicability to nonlinear time-series and spatio-temporal data.

1. Introduction

This paper develops profile likelihood inference methodology for situations where computationally intensive Monte Carlo methods are employed to evaluate and maximize the likelihood function. If the profile log-likelihood function can be computed with a Monte Carlo error small compared to one unit, carrying out statistical inference from the Monte Carlo profile as if it were the true profile will have relatively small effects on resulting confidence intervals. Sometimes, however, no reasonable amount of computation can reduce the Monte Carlo error in evaluating the profile to levels at or below one log unit. This predicament typically arises with large datasets and complex models. However, to investigate large datasets in the context of complex models there is little alternative to the use of Monte Carlo methods. Monte Carlo approximation of the profile likelihood function provides opportunities to assess Monte Carlo variability and make appropriate compensations. We use this approach to construct profile likelihood confidence intervals with statistical behaviour properly adjusted for Monte Carlo uncertainty.

Our paper is organized as follows. First, we set up mathematical notation to formalize the task of Monte Carlo profile likelihood estimation via a metamodel. Section 2 puts this task in the context of some previous work on likelihood-based inference for intractable models. Section 3 develops our methodological approach. Section 4 presents a dynamic latent variable modeling framework of broad applicability for which the methodology is appropriate. We demonstrate the capabilities of our methodology by solving two inferential problems for which scientific progress has been limited by the lack of effective statistical methodology. These examples arise from the study of transmissible human diseases, a field characterized by extensive and diverse data, indirect observation of the underlying infection processes, strongly nonlinear stochastic dynamics and public health importance. Infectious disease data therefore provide many inference opportunities and challenges. Section 4.1 concerns inference on population
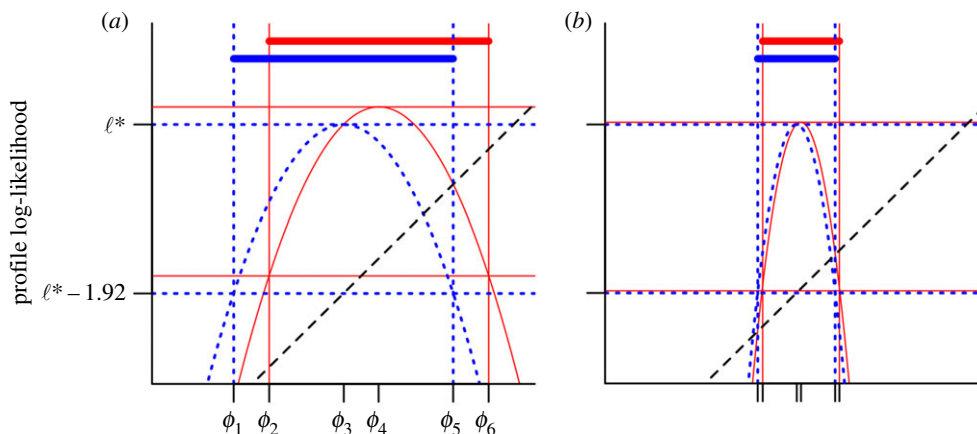
Figure 1. The effect of bias on confidence intervals for a quadratic profile log-likelihood function. (*a*) The blue dotted quadratic represents a log-likelihood profile. The maximum-likelihood estimator of the profiled parameter is ϕ_3, with corresponding log likelihood $\ell^* = \ell(\hat{\theta}^*\,;\,y^*) = \ell^P(\hat{\phi}^*\,;\,y^*)$. The 95% CI [$\phi_1$, ϕ_5] is constructed, via the horizontal and vertical blue dotted lines, as the set of parameter values with profile log likelihood higher than $\ell^* - 1.92$. The red quadratic is the sum of the blue dotted quadratic and linear bias (black dashed line). Horizontal and vertical red lines construct the resulting approximate confidence interval [ϕ_2, ϕ_6] and point estimate ϕ_4. (*b*) The same construction, but with higher curvature of the profile log likelihood leading to diminishing effect of the bias on the confidence interval. (Online version in colour.)

dynamics from genetic data. Section 4.2 concerns fitting nonlinear partially observed Markov models to panel data. Section 4.3 discusses the role of our methodology for nonlinear time-series and spatio-temporal data analysis. Section 5 investigates our methodology via a simulation study on a toy example. Section 6 is a concluding discussion which situates our paper within the broader goal of inference for large datasets and complex models.

We consider a general statistical inference framework in which data are a real-valued vector, y^*, modelled as a realization of a random variable Y having density $f_Y(y\,;\,\theta)$, where θ is an unknown parameter in $\mathbb{R}^p$. We are concerned with inference on θ in situations where the data analyst cannot directly evaluate $f_Y(y\,;\,\theta)$. Instead, we suppose that approximate evaluation of $f_Y(y\,;\,\theta)$ is possible through Monte Carlo approaches. One situation in which this arises is when the statistician can simulate draws from the density $f_Y(y\,;\,\theta)$ despite being unable to directly evaluate it [1]. In addition to a simulator for the full joint distribution of Y, we might also have access to simulators for various marginal and conditional distributions related to $f_Y(y\,;\,\theta)$. For example, this can arise if Y has the structure of a fully or partially observed Markov process [2]. Simulation-based methods are growing in usage, motivated by advances in the availability of complex data and the desire for statistical fitting of complex models to these data. Although we cannot calculate them, we can nevertheless define the log-likelihood function,

$$\ell(\theta\,;\,y^*) = \log f_Y(y^*\,;\,\theta) \qquad (1.1)$$

and a maximum-likelihood estimate (MLE),

$$\hat{\theta}^* = \hat{\theta}(y^*) = \arg\max_{\theta} \ell(\theta\,;\,y^*). \qquad (1.2)$$

To formalize the task of constructing marginal confidence intervals, we suppose that $\theta = (\phi, \psi)$ with $\phi \in \mathbb{R}^1$ and $\psi \in \mathbb{R}^{p-1}$. Here, ϕ is a focal parameter for which we are interested in obtaining a confidence interval using the data, y^*. As the choice of focal parameter is arbitrary, we are solving the general problem of obtaining a marginal confidence interval for each component of a parameter vector. The profile log-likelihood function for ϕ is defined as

$$\ell^P(\phi\,;\,y^*) = \max_{\psi} \ell((\phi, \psi)\,;\,y^*). \qquad (1.3)$$

The profile log likelihood is maximized at a marginal MLE,

$$\hat{\phi}^* = \hat{\phi}(y^*) = \arg\max_{\phi} \ell^P(\phi\,;\,y^*). \qquad (1.4)$$

A profile likelihood confidence interval with cut-off δ is defined as

$$\{\phi : \ell^P(\phi\,;\,y^*) > \ell^P(\hat{\phi}^*\,;\,y^*) - \delta\}. \qquad (1.5)$$

Profile likelihood confidence intervals are a widespread inference approach with some favourable properties, including asymptotic efficiency and natural transformation under reparameterization [3]. Modifications can lead to higher-order asymptotic performance [4] but these are not routinely available. In our context, (1.3)–(1.5) are not directly accessible to the data analyst. Instead, we work with independent Monte Carlo profile likelihood evaluations at a sequence of points $\phi_{1:K} = (\phi_1, \ldots, \phi_K)$. We denote the evaluations as $(\breve{\ell}_k^P(y^*), k \in 1:K)$, using a breve accent to distinguish Monte Carlo quantities. Without loss of generality we can write

$$\breve{\ell}_k^P(y^*) = \ell^P(\phi_k\,;\,y^*) + \beta_k(y^*) + \epsilon_k(y^*), \quad k \in 1:K, \qquad (1.6)$$

where $\epsilon_{1:K}(Y)$ are Monte Carlo random variables which are, by construction, mean zero and independent conditional on Y. In (1.6), $\beta_{1:K}(y^*)$ gives the Monte Carlo bias of each profile log-likelihood evaluation. Motivation for the decomposition into target quantity, bias and additive error on the log scale in (1.6) is that this is a proper scale for Monte Carlo central limit theory relevant to our subsequent examples [5] as well as an appropriate scale for inference.

The amount of information about ϕ in the data is represented by the curvature of the profile log likelihood, and if this is large then statistically relevant region with high profile likelihood is narrow. As represented pictorially in figure 1, increasing the curvature of the profile log likelihood reduces the consequence of non-constant bias in the construction of profile confidence intervals. A useful simplification arises when it is reasonable to treat the distribution of the Monte Carlo bias and error in (1.6) as constant across the statistically relevant region having high-profile likelihood. This leads us to consider a metamodel with $\beta_k(y^*) = \beta(y^*)$ and with $\epsilon_{1:K}(y^*)$ independent and identically distributed (i.i.d.)

with variance $\sigma^2(y^*) < \infty$. The resulting metamodel is

$$\overset{P}{\ell}_k(y^*) = \ell^P(\phi_k \,;\, y^*) + \beta(y^*) + \epsilon_k,$$
$$\epsilon_{1:K} \text{ i.i.d. with } \mathrm{Var}(\epsilon_k) = \sigma^2(y^*). \quad (1.7)$$

Empirical evidence for non-constant Monte Carlo variance could motivate the inclusion of heteroskadistic errors in (1.7). The assumption in (1.7) of approximately constant Monte Carlo bias is hard to quantify empirically on computationally challenging problems, since one cannot readily obtain an estimate with negligible bias. Although the bias on the Monte Carlo profile likelihood estimate may be intractable, the coverage of a constructed confidence interval can be checked by simulation at a specific parameter value such as an MLE, as demonstrated in §5.

2. Previous work on likelihood-based inference via simulation

A prescient paper by Diggle & Gratton [1] developed Monte Carlo maximum-likelihood methodology with similar motivation to our current goals. However, Diggle & Gratton [1] did not work with profile likelihood and did not show how to correct the resulting confidence intervals for Monte Carlo error. Further, Diggle & Gratton [1] assumed that the Monte Carlo methods would involve simulating from the modelled joint distribution of the entire dataset, whereas modern computationally efficient Monte Carlo algorithms may be based on simulating sequentially from conditional distributions in a carefully crafted decomposition of the entire joint distribution. Bretó *et al.* [2] and He *et al.* [6] introduced the term *plug-and-play* to describe statistical methodology for which the model (viewed as an input to an inference algorithm) is specified via a simulator in this broader sense. The term *likelihood-free* has been used similarly, in the context of Markov chain Monte Carlo (MCMC) [7] and sequential Monte Carlo (SMC) [8,9]. The term *equation-free* has been used for the related concept of simulation-based model investigations in the physical sciences [10]. Related terms *implicit* [1] and *doubly intractable* [11] have been used to describe models for which only plug-and-play algorithms are practical. From the point of view of categorizing statistical methodology, it is natural to view the way in which an inference algorithm accesses the statistical model as a property of the algorithm rather than a property of the model.

Rubio & Johansen [12] investigated non-parametric estimation of a likelihood surface via approximate Bayesian computing (ABC). These authors also provided a literature review of previous approaches to carry out statistical inferences in situations where likelihood evaluation and maximization necessarily involve computationally intensive and noisy Monte Carlo procedures. We are not aware of previous work developing Monte Carlo profile likelihood methodology. Profile methodology focuses the computational effort on parameters of key interest—specifically, parameters for which one computes a profile. The process of constructing a profile requires computation of a relevant feature of the likelihood surface in the region of inferential interest. Studying the likelihood surface on this scale, rather than focusing exclusively on a point estimate such as the maximum-likelihood estimate, has some theoretical justification [13]. In the general theory of stochastic simulation-based optimization, building metamodels describing the response surface is a standard technique [14]. Our goal is to develop metamodel methodology that takes advantage of the statistical properties of the profile likelihood and constructs confidence intervals correcting properly for Monte Carlo variability.

3. Profile cut-off correction via a local quadratic metamodel

Local asymptotic normality (LAN) provides a general theoretical framework in which the log-likelihood function is asymptotically well approximated by a quadratic [15]. Under sufficient regularity conditions, this quadratic approximation is inherited by the profile log likelihood [16]. Here, we write the marginal ϕ component of the LAN property as a finite sample normal approximation given by

$$\ell^P(\phi \,;\, Y) - \ell^P(\phi_0 \,;\, Y) \approx Z(\phi - \phi_0)\sqrt{I} - \frac{(\phi - \phi_0)^2 I}{2}, \quad (3.1)$$

where $Y \sim f_Y(y \,;\, \theta_0)$ for $\theta_0 = (\phi_0, \psi_0)$, and $Z \sim N[0, 1]$ is a normal random variable with mean 0 and variance 1. In (3.1), $\approx$ indicates approximate equality in distribution. Under regular asymptotics, the curvature of the quadratic approximation in LAN is the Fisher information, and LAN is therefore a similar property to asymptotic normality of the maximum-likelihood estimate. The quantity I in (3.1) can be interpreted as the marginal Fisher information for ϕ, also known as the ϕ-aspect of the Fisher information [4, Section 3.4]. Specifically, if we write the inverse of the full Fisher information as

$$V_\theta = \begin{bmatrix} V_\phi & V_{\phi\psi} \\ V_{\psi\phi} & V_\psi \end{bmatrix},$$

then $I = V_\phi^{-1}$. In this article, we focus on developing and demonstrating statistical methodology rather than on presenting theoretical results. Therefore, the formal mathematical representation of the approximations in this paper as asymptotic limit theorems is postponed to subsequent work.

The LAN property suggests that the Monte Carlo profile log likelihood evaluated at $\phi_{1:K}$ can be approximated, in a neighbourhood of its maximum, by a quadratic metamodel,

$$\overset{P}{\ell}_k(y) = -\hat{a}(y)\phi_k^2 + \hat{b}(y)\phi_k + \hat{c}(y) + \epsilon_k,$$
$$\epsilon_{1:K} \text{ i.i.d. with } \mathrm{Var}(\epsilon_k) = \sigma^2(y^*). \quad (3.2)$$

This local quadratic metamodel is a special case of (1.7). The unknown coefficients $\hat{a}(y^*)$, $\hat{b}(y^*)$ and $\hat{c}(y^*)$, corresponding to equation (3.2) evaluated at $y = y^*$, describe a quadratic approximation to the numerically intractable likelihood surface. We can use standard linear regression to estimate $\hat{a}(y^*)$, $\hat{b}(y^*)$ and $\hat{c}(y^*)$ from the Monte Carlo profile evaluations. Writing $\epsilon = \epsilon_{1:K}$, we denote the resulting linear regression coefficients as $\breve{a}^* = \breve{a}(y^*, \epsilon)$, $\breve{b}^* = \breve{b}(y^*, \epsilon)$ and $\breve{c}^* = \breve{c}(y^*, \epsilon)$. The Monte Carlo quadratic profile likelihood approximation is

$$\breve{\ell}^Q(\phi \,;\, y^*) = -\breve{a}^* \phi^2 + \breve{b}^* \phi + \breve{c}^*. \quad (3.3)$$

The marginal MLE $\hat{\phi}^*$ can be approximated by the maximum of $\breve{\ell}^Q(\phi \,;\, y^*)$, which is given by

$$\breve{\phi}^Q(y^*, \epsilon) = \frac{\breve{b}(y^*, \epsilon)}{2\breve{a}(y^*, \epsilon)}.$$

Now, for $Y \sim f_Y(y;\theta_0)$, we separate the variability in $\breve{\phi}^Q(Y,\epsilon)$ into two components:

(1) *Statistical error* is the uncertainty resulting from randomness in the data, viewed as a draw from the statistical model. This is the error in the ideal quadratic profile approximation estimate $\hat{b}(y^*)/2\hat{a}(y^*)$ as an estimate of ϕ_0.

(2) *Monte Carlo* error is the uncertainty resulting from implementing a Monte Carlo estimator. This is the error in $\breve{b}(y^*,\epsilon)/2\breve{a}(y^*,\epsilon)$ as a Monte Carlo estimate of $\hat{b}(y^*)/2\hat{a}(y^*)$.

The LAN approximation in (3.1) suggests a normal approximation for the distribution of the marginal MLE $\hat{\phi}^*$ which we write as

$$\hat{\phi}(Y) \approx N[\phi_0, \text{SE}_{\text{stat}}^2]. \tag{3.4}$$

The usual statistical standard error, $1/\sqrt{2\hat{a}(y^*)}$, is not available to us, but we can instead use its Monte Carlo estimate,

$$\text{SE}_{\text{stat}} = \frac{1}{\sqrt{2\breve{a}(y^*,\epsilon)}}. \tag{3.5}$$

To quantify the Monte Carlo error, we first note that standard linear model methodology provides variance and covariance estimates $\breve{\text{Var}}[\breve{a}(y^*,\epsilon)]$, $\breve{\text{Var}}[\breve{b}(y^*,\epsilon)]$ and $\breve{\text{Cov}}(\breve{a}(y^*,\epsilon),\breve{b}(y^*,\epsilon))$. The regression errors represent only Monte Carlo variability conditional on $Y = y^*$, i.e. $\breve{\text{Var}}[\breve{a}(y^*,\epsilon)] = \text{Var}[\breve{a}(Y,\epsilon)|Y=y^*]$. A standard central limit approximation for regression coefficient estimates is

$$\begin{pmatrix} \breve{a}(y^*,\epsilon) \\ \breve{b}(y^*,\epsilon) \end{pmatrix} \approx N$$

$$\left[\begin{pmatrix} \hat{a}(y^*) \\ \hat{b}(y^*) \end{pmatrix}, \begin{pmatrix} \breve{\text{Var}}[\breve{a}(y^*,\epsilon)] & \breve{\text{Cov}}(\breve{a}(y^*,\epsilon),\breve{b}(y^*,\epsilon)) \\ \breve{\text{Cov}}(\breve{a}(y^*,\epsilon),\breve{b}(y^*,\epsilon)) & \breve{\text{Var}}[\breve{b}(y^*,\epsilon)] \end{pmatrix} \right].$$

An application of the delta method gives a central limit approximation for the maximum, conditional on $Y = y^*$, given by

$$\frac{\breve{b}(y^*,\epsilon)}{2\breve{a}(y^*,\epsilon)} \approx N\left[\left(\frac{\hat{b}(y^*)}{2\hat{a}(y^*)}\right), \text{SE}_{\text{mc}}^2\right], \tag{3.6}$$

where:

$$\text{SE}_{\text{mc}}^2 = \breve{\text{Var}}\left[\frac{\breve{b}(y^*,\epsilon)}{2\breve{a}(y^*,\epsilon)}\right] \approx \frac{1}{4\breve{a}^2(y^*,\epsilon)}\left\{\breve{\text{Var}}[\breve{b}(y^*,\epsilon)] - \frac{2\breve{b}(y^*,\epsilon)}{\breve{a}(y^*,\epsilon)}\breve{\text{Cov}}[\breve{a}(y^*,\epsilon),\breve{b}(y^*,\epsilon)] + \frac{\breve{b}^2(y^*,\epsilon)}{\breve{a}^2(y^*,\epsilon)}\breve{\text{Var}}[\breve{a}(y^*,\epsilon)]\right\}. \tag{3.7}$$

To obtain the combined statistical and Monte Carlo error, we write

$$\text{Var}\left[\frac{\breve{b}(Y,\epsilon)}{2\breve{a}(Y,\epsilon)}\right] = \mathbb{E}\left\{\text{Var}\left[\frac{\breve{b}(Y,\epsilon)}{2\breve{a}(Y,\epsilon)}\bigg|Y\right]\right\}$$
$$+ \text{Var}\left\{\mathbb{E}\left[\frac{\breve{b}(Y,\epsilon)}{2\breve{a}(Y,\epsilon)}\bigg|Y\right]\right\}. \tag{3.8}$$

Now, from (3.1), the curvature of the profile log likelihood is approximately constant, independent of Y. We suppose that the profile points used to obtain $\breve{\phi}^Q(Y,\epsilon)$ are approximately centred on $\breve{\phi}^Q(Y,\epsilon)$ regardless of the value of Y. This assumption can be satisfied by construction, for example, by fitting the quadratic metamodel in (3.2) using local weights (as in the Monte Carlo-adjusted profile (MCAP) algorithm below). Further, we suppose that $\text{Var}[\epsilon_k(Y)] \approx \text{Var}[\epsilon_k(y^*)]$. Together, these approximations imply

$$\text{Var}\left[\frac{\breve{b}(Y,\epsilon)}{2\breve{a}(Y,\epsilon)}\bigg|Y\right] \approx \text{Var}\left[\frac{\breve{b}(y^*,\epsilon)}{2\breve{a}(y^*,\epsilon)}\right]. \tag{3.9}$$

Also, from the central limit approximation in (3.6), we have

$$\mathbb{E}\left[\frac{\breve{b}(Y,\epsilon)}{2\breve{a}(Y,\epsilon)}\bigg|Y\right] \approx \hat{\phi}(Y). \tag{3.10}$$

Putting (3.9) and (3.10) into (3.8), and using the approximations in (3.4) and (3.6), we get

$$\frac{\breve{b}(Y,\epsilon)}{2\breve{a}(Y,\epsilon)} \approx N[\phi_0, \text{SE}_{\text{total}}^2],$$

where

$$\text{SE}_{\text{total}} = \sqrt{\text{SE}_{\text{mc}}^2 + \text{SE}_{\text{stat}}^2}.$$

The usual asymptotic profile likelihood confidence interval cut-off value can be obtained by converting the standard error of the MLE into an equivalent cut-off on a quadratic approximation to the profile log likelihood. In our setting, the asymptotic $(1-\alpha)$ confidence interval, $\breve{\phi}^Q \pm z_\alpha \times \text{SE}_{\text{total}}$, where $\mathbb{P}[Z > z_\alpha] = \alpha/2$, is equivalent to a Monte Carlo adjusted profile cut-off for the quadratic approximation $\breve{\ell}^Q(\phi; y^*)$ of

$$\delta = \breve{a}^* \times (z_\alpha \times \text{SE}_{\text{total}})^2 = z_\alpha^2\left(\breve{a}^* \times \text{SE}_{\text{mc}}^2 + \frac{1}{2}\right). \tag{3.11}$$

Note that, if $\text{SE}_{\text{mc}} = 0$, the calculation in (3.11) for $\alpha = 0.05$ reduces to

$$\delta = \frac{1.96^2}{2} = 1.92,$$

the usual cut-off to construct a 95% CI for an exact profile likelihood.

Confidence intervals based on a quadratic approximation to the exact log likelihood are asymptotically equivalent to using the same cut-off δ with a smoothed version of the likelihood, so long as an appropriate smoother is used [13]. An appropriate smoother should return a quadratic when the points do indeed lie on a quadratic, a property satisfied, for example, by local quadratic smoothing such as the R function `loess` [17]. We, therefore, propose using δ as an appropriate cut-off on a profile likelihood estimate obtained by applying a suitable smoother to the Monte Carlo evaluations in (1.6). A smoother, $S(\phi; \phi_{1:K}, \breve{\ell}^P_{1:K}, \lambda)$, generates a value at ϕ based on fitting a smooth curve through the points $\{(\phi_k, \breve{\ell}^P_k), k \in 1:K\}$ with an algorithmic parameter λ determining the smoothness of the fit. A resulting maximum

smoothed Monte Carlo profile likelihood estimate is

$$\breve{\phi}^S = \arg \max_{\phi} S(\phi; \, \phi_{1:K}, \breve{\ell}^P_{1:K}, \lambda). \tag{3.12}$$

A corresponding Monte Carlo profile likelihood confidence interval for a cut-off δ is

$$\{\phi : S(\phi; \, \phi_{1:K}, \breve{\ell}^P_{1:K}, \lambda) > S(\breve{\phi}^S; \, \phi_{1:K}, \breve{\ell}^P_{1:K}, \lambda) - \delta\}. \tag{3.13}$$

Here, we suppose that $S(\phi; \, \phi_{1:K}, \breve{\ell}^P_{1:K}, \lambda)$ is evaluated at ϕ via local quadratic regression with weight $w_k(\phi)$ on the point $(\phi_k, \breve{\ell}^P_k)$, where $w_k(\phi)$ depends on the proximity of ϕ to ϕ_k. Specifically, we take S to be the widely used local quadratic smoother of [18] as implemented by the function `loess` in R3.3.3. In this case, λ is the *span* of the smoother, defined as the fraction of the data used to construct the weights in the local regression at any point ϕ. In practice, the statistician needs to specify λ. While automated choices of smoothing parameter have been proposed, it remains standard practice to choose the smoothing parameter based on some experimentation and looking at the resulting fit. In our experience, the default `loess` choice of $\lambda = 0.75$ in R3.3.3 has been appropriate in most cases. However, a larger value of λ is needed when the profile is evaluated at very few points (as demonstrated in the electronic supplementary material, Section S1). When the exact profile is not far from quadratic, one can expect local quadratic smoothing of the Monte Carlo profile likelihood to be insensitive to the choice of λ.

Just as the local quadratic regression smoother has weights $w_{1:K}(\phi)$, the quadratic metamodel in (3.2) can be fitted using regression weights. A natural choice of these weights for obtaining a profile confidence interval cut-off for $S(\phi; \, \phi_{1:K}, \breve{\ell}^P_{1:K}, \lambda)$ is $w_{1:K}(\breve{\phi}^S)$. This choice is used for the MCAP algorithm below. For our numerical results, we used the implementation of this MCAP algorithm given as the electronic supplementary material, Section S3.

Algorithm MCAP

input:

Monte Carlo profile $\breve{\ell}^P_{1:K}$ evaluated at $\phi_{1:K}$

Local quadratic regression smoother, S

Smoothing parameter, λ

Confidence level, $1 - \alpha$

output:

Cut-off, δ, for a Monte Carlo profile likelihood confidence interval

algorithm:

Fit a local quadratic smoother, $\breve{\ell}^S(\phi) = S(\phi; \, \phi_{1:K}, \breve{\ell}^P_{1:K}, \lambda)$

Obtain $\breve{\phi}^S = \arg \max \breve{\ell}^S(\phi)$

Obtain regression weights $w_{1:K}$ for the evaluation of
$\quad S(\phi; \, \phi_{1:K}, \breve{\ell}^P_{1:K}, \lambda)$ at $\phi = \breve{\phi}^S$

Fit a linear regression model, $\breve{\ell}^P_k = -a\phi_k^2 + b\phi_k + c + \epsilon_k$, with
$\quad$ weights $w_{1:K}$

Obtain regression estimates $\breve{a}$ and $\breve{b}$

Obtain regression covariances $\widehat{\mathrm{Var}}[\breve{a}], \widehat{\mathrm{Var}}[\breve{b}], \widehat{\mathrm{Cov}}[\breve{a}, \breve{b}]$

Let $\mathrm{SE}^2_{mc} = \frac{1}{4\breve{a}^2}\{\widehat{\mathrm{Var}}[\breve{b}] - \frac{2\breve{b}}{\breve{a}}\widehat{\mathrm{Cov}}[\breve{a}, \breve{b}] + \frac{\breve{b}^2}{\breve{a}^2}\widehat{\mathrm{Var}}[\breve{a}]\}$

Let χ_{α} be the $(1 - \alpha)$ quantile of the χ-square distribution on one
$\quad$ degree of freedom

Let $\delta = \chi_{\alpha}(\breve{a} \times \mathrm{SE}^2_{mc} + 1/2)$

---

4. Example: inference for partially observed dynamic systems

Many dynamic systems with indirectly observed latent processes can be modelled within the partially observed Markov process (POMP) framework. A general POMP model, also known as a hidden Markov model or a state space model, consists of a latent Markov process $\{X(t)\}$, with $X(t)$ taking values in a space $\mathbb{X}$, together with a sequence of observable random variables $Y_{1:N} = (Y_1, \ldots, Y_N)$. The observation Y_n is modelled as a measurement of $X(t_n)$ by requiring that Y_n is conditionally independent of the other observations and of $\{X(t)\}$ given $X(t_n)$. We will suppose that Y_n takes values in $\mathbb{Y} = \mathbb{R}^d$, the space of d-dimensional real vectors. When $d = 1$ (or d is small) $Y_{1:N}$ is called a univariate (or multivariate) time-series model. The POMP framework provides a fundamental approach for nonlinear time-series analysis, with innumerable applications [2]. When d becomes large, the POMP framework allows for non-linear panel data and spatio-temporal data, as well as other complex data structures. Unless the POMP model is linear and Gaussian, or $\mathbb{X}$ is a sufficiently small finite set, Monte Carlo techniques such as SMC are required to evaluate the likelihood function. For our examples, we focus on likelihood maximization by iterated filtering [19]. Similar issues arise with alternative computational approaches, including Monte Carlo Expectation-Maximization algorithms [20, ch. 11]. Even for the relatively simple case of time-series POMP models (discussed further in §4.3) numerical issues can be computationally demanding for currently available methodology, giving opportunity for MCAP methodology to facilitate data analysis. However, to demonstrate the capabilities of our methodology, we first present two high-dimensional POMP inference challenges that become computationally tractable using MCAP.

4.1. Inferring population dynamics from genetic sequence data

Genetic sequence data on a sample of individuals in an ecological system has potential to reveal population dynamics. Extraction of this information has been termed *phylodynamics* [21]. Likelihood-based inference for joint models of the molecular evolution process, population dynamics and measurement process is a challenging computational problem. The bulk of extant phylodynamic methodology has therefore focused on inference for population dynamics conditional on an estimated phylogeny and replacing the population dynamic model with an approximation, called a *coalescent model* that is convenient for calculations backwards in time [22]. Working with the full joint likelihood is not entirely beyond modern computational capabilities; in particular it can be done using the `genPomp` algorithm of Smith *et al.* [23]. The `genPomp` algorithm is an application of iterated filtering methodology [19] to phylodynamic models and data. To the best of our knowledge, `genPomp` is the first algorithm capable of carrying out full joint likelihood-based inference for population-level phylodynamic inference. However, the `genPomp` algorithm leads to estimators with high Monte Carlo variance, indeed, too high for reasonable amounts of computation resources to reduce Monte Carlo variability to negligibility.
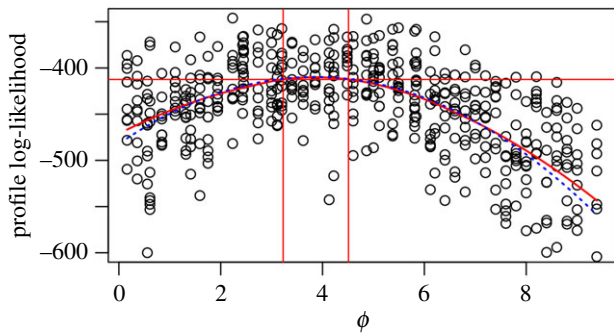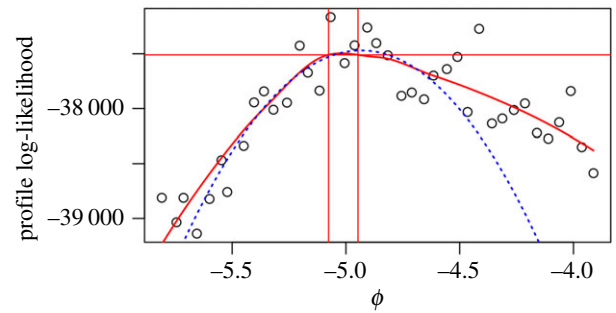
Figure 2. Profile likelihood for an infectious disease transmission parameter inferred from genetic data on pathogens. The smoothed profile likelihood and corresponding MCAP 95% CI are shown as solid red lines. The quadratic approximation in a neighbourhood of the maximum is shown as a blue dotted line. (Online version in colour.)

Figure 3. Profile likelihood for a nonlinear partially observed Markov process model for a panel of time series of historical state-level polio incidence in the USA. The smoothed profile likelihood and corresponding MCAP 95% CI are shown as solid red lines. The quadratic approximation in a neighbourhood of the maximum is shown as a dotted blue line. (Online version in colour.)

This, therefore, provides a useful scenario to demonstrate our methodology.

Figure 2 presents a Monte Carlo profile computed by Smith et al. [23], with confidence intervals constructed by applying the MCAP algorithm implemented by the mcap procedure (electronic supplementary material, Section S3) with default smoothing parameter. The model and data concern HIV transmission in Southeast Michigan, but details of the model and computations are not of immediate interest since all we need to consider are the estimated profile likelihood points. The profiled parameter quantifies HIV transmission from recently infected, diagnosed individuals—it is ε_{J_0} in the notation of Smith et al. [23] but we rename it as ϕ for the current paper. The computations for figure 2 took approximately 10 days using 500 cores on a Linux cluster. To scale this methodology to increasingly large datasets and more complex models, it is apparent that one may be limited by the computational effort required to control Monte Carlo error. The MCAP procedure gives a Monte Carlo standard error of $SE_{mc} = 0.151$ on the value maximizing the smoothed Monte Carlo profile, based on the quadratic approximation at the maximum. The statistical error is $SE_{stat} = 0.32$. Combining these sources of uncertainty gives a total standard error of $SE_{total} = 0.354$. From (3.11), the resulting 95% CI cut-off is $\delta = 2.35$. We see in figure 2 that the smoothed profile is close to its quadratic approximation in the neighbourhood of the maximum statistically supported by the data. We also see that the Monte Carlo uncertainty in the profile confidence interval is rather small, leading to a profile cut-off not much bigger than the value of 1.92 for zero Monte Carlo error, despite the large Monte Carlo variability in the evaluation of any one point on the profile.

4.2. Panel time-series analysis

Panel data consist of a collection of time series which have some shared parameters, but negligible dynamic dependence. We consider inference using mechanistic models for panel data, i.e. equations for how the process progresses through time derived from scientific principles about the system under investigation. In principle, statistical methods for mechanistic time-series analysis [2] extend to the panel situation [24]. However, extensive data add computational challenges to Monte Carlo inference schemes. In particular,

with increasing amounts of data, it must eventually become infeasible to calculate the likelihood with an error as small as one log unit. The MCAP procedure nevertheless succeeds so long as the signal-to-noise ratio in the Monte Carlo profile is adequate. In a simple situation, where each time series is modelled as i.i.d. and each time-series model contains the same parameters, we can check how this ratio scales. The Fisher information scales linearly with the number of time series in the panel, and therefore the curvature of the log-likelihood profile also scales linearly. The Monte Carlo standard error on the likelihood scales at a square-root rate. In this case, we, therefore, expect the MCAP methodology to scale successfully with the number of time series in the panel.

Investigations of population-level infectious disease transmission lead to highly nonlinear, stochastic, partially observed dynamic models. The great majority of disease transmission is local, despite the importance of spatial transmission to seed the local epidemics [25]. Fitting models to panels of epidemiological time-series data, such as incidence data for collections of cities or states, offers potential to elucidate the similarities and differences between these local epidemics.

We demonstrate the MCAP procedure on a panel estimate of the reporting rate of paralytic polio in the pre-vaccination era USA. Reporting rate has important consequences for understanding the system: conditional on observed incidence data, reporting rate determines the extent of the unreported epidemic. Yet, in the presence of many uncertainties about this complex disease transmission system, a single disease incidence time series often cannot conclusively pin down this epidemiological parameter. The profile evaluations in figure 3 were obtained by Bretó et al. [24] in an extension of the analysis of Martinez-Bakker et al. [26]. Martinez-Bakker et al. [26] analysed state-level paralytic polio incidence data in order to study the role of unobserved asymptomatic polio infections in disease persistence. Here, the reporting rate parameter ($\log(\rho)$ in the terminology of [24]) is denoted by ϕ. The MCAP procedure gives a Monte Carlo standard error of $SE_{mc} = 0.033$ and a statistical error of $SE_{stat} = 0.013$. Combining them gives a total standard error of $SE_{total} = 0.035$. The resulting profile cut-off is $\delta = 13.6$. The profile decreases slowly to the right of the smoothed MLE, since higher reporting rates can be compensated for by lower transmission intensities. The model struggles to explain reporting rates much lower than the smoothed MLE, since the reporting rate must be sufficient to explain the observed number

of cases in a situation where almost all individuals acquire non-paralytic polio infections. This asymmetrical trade-off may explain why the profile log likelihood shows some noticeable deviation from its quadratic approximation in a neighbourhood of the maximum. A consequence of this changing curvature is that the quadratic approximation used to construct the Monte Carlo profile at its maximum (figure 3, dotted blue line) does not share this maximum.

The computations for figure 3 required approximately 24 h on 300 cores. At this level of computational intensity, we see that the majority of uncertainty about the parameter ϕ is due to Monte Carlo error rather than statistical error. For this large panel dataset, in the context of the fitted model, the parameter ϕ would be identified very accurately by the data if we had access to the actual likelihood surface. Additional computation could, therefore, reduce the uncertainty on our estimate of ϕ by a factor of three. However, the data analyst may decide the available computational effort is better used exploring other parameters or alternative model specifications.

4.3. Applications to time-series and spatio-temporal data analysis

The examples in §§4.1 and 4.2 demonstrate applications which were computationally intractable without MCAP. Applications of the POMP framework to nonlinear time-series analysis typically involve smaller data sets, and a relatively simple dependence structure, and are therefore less computationally demanding. This consideration has facilitated the utilization of Monte Carlo profile likelihood, without the benefits of MCAP, as a technique at the cutting edge of nonlinear time-series analysis. In the context of infectious disease dynamics, Dobson [27] wrote, 'Powerful new inferential fitting methods [28] considerably increase the accuracy of outbreak predictions while also allowing models whose structure reflects different underlying assumptions to be compared. These approaches move well beyond time series and statistical regression analyses as they include mechanistic details as mathematical functions that define rates of loss of immunity and the response of vector abundance to climate.' Examples showing a central role for Monte Carlo profile likelihood in such analyses are given by King *et al.* [29, Fig. 2], Camacho *et al.* [30, Figs. S3 and S8A], Blackwood *et al.* [31, Fig. 3A], Shrestha *et al.* [32, Figs. 2B-2G and 4L-4P] and Blake *et al.* [33, Figs. S1, S4 and S5]. The main practical limitation of this approach is computational resources [6]. We have shown that our methodology can both quantify and dramatically reduce the Monte Carlo error in computationally intensive inferences for POMP models. The MCAP procedure therefore improves the accessibility and scalability of inference for nonlinear time-series models.

Spatio-temporal data consists of time series collected at various locations. Models for partially observed spatio-temporal dynamics extend the panel models of §4.2 by allowing for dynamic dependence between locations. SMC methods, that provide a foundation for likelihood-based inference relating POMP models to time-series data, struggle with spatio-temporal data since they scale poorly with spatial dimension [34]. Theoretically, SMC methods with sub-exponential scaling can be developed for weakly coupled spatio-temporal systems [35]. Nevertheless, practical methodology for fitting nonlinear non-Gaussian spatio-temporal models continues to be constrained by high Monte Carlo variance [36]. Thus, this class of inference challenges stands to benefit from our MCAP methodology. The electronic supplementary material, Section S1, presents an example for fitting a coupled spatio-temporal model to measles incidence in twenty cities.

5. A simulation study of the Monte Carlo-adjusted profile procedure

We look for a numerically convenient toy scenario that generates Monte Carlo profiles resembling figures 2 and 3. Our simulated data are an independent, identically distributed lognormal sample $Y_{1:N}$, where $\log(Y_n) \sim N[\phi, 2\sigma^2]$ for $n \in 1:N$. We consider a profile likelihood confidence interval for the log mean parameter, ϕ. The lognormal distribution leads to log-likelihood profiles that deviate from quadratic. To set up a situation with Monte Carlo error in evaluating and maximizing the likelihood, we supposed that the likelihood is accessed via Monte Carlo integration of a latent variable. Specifically, we write $Y_n|X_n \sim$ lognormal(X_n, σ^2) with $X_n \sim N[\phi, \sigma^2]$. Then, our Monte Carlo density estimator is

$$\breve{f}_Y(y \; ; \; \phi, \sigma, s, J) = \frac{1}{J} \sum_{j=1}^{J} f_{\mathrm{LN}}(y \; ; \; \phi + \sigma \epsilon_j, \sigma^2), \quad (5.1)$$

where $f_{\mathrm{LN}}(y \; ; \; \mu, \tau^2)$ is the lognormal density,

$$f_{\mathrm{LN}}(y \; ; \; \mu, \tau^2) = \frac{1}{y\tau\sqrt{2\pi}} \exp\left\{ \frac{-(\log y - \mu)^2}{2\tau^2} \right\}$$

and $\epsilon_{1:J}$ is a sequence of standard normal pseudo-random numbers corresponding to a seed s. We suppose that we are working with a parallel random number generator such that pseudo-random sequences corresponding to different seeds behave numerically like independent random sequences. Our Monte Carlo log-likelihood estimator is

$$\breve{\ell}(\phi, \sigma \; ; \; y_{1:N}, s, J) = \sum_{n=1}^{N} \log \breve{f}_Y(y_n \; ; \; \phi, \sigma, s + n - 1, J). \quad (5.2)$$

Our Monte Carlo profile is calculated at $\phi \in \phi_{1:K}$. We maximize the likelihood numerically, at a fixed seed, to give a corresponding estimate of σ given by

$$\breve{\sigma}_k^P(y_{1:N}, s, J) = \arg\max_\sigma \ell(\phi_k, \sigma \; ; \; y_{1:N}, s + N(k - 1), J). \quad (5.3)$$

We do not wish to imply that practical examples will generally result from a fixed-seed Monte Carlo likelihood calculation. Seed fixing is an effective technique for removing Monte Carlo variability from relatively small calculations, but can become difficult or impossible to implement effectively for complex, coupled, nonlinear systems.

The following numerical results used $N = 50$ and $J = 3$ with true parameter values $\phi_0 = 0$ and $\sigma_0^2 = 1$. There are two ways to increase the Monte Carlo error in the log likelihood for this toy example, by increasing the sample size, N, and decreasing the Monte Carlo effort, J. The Monte Carlo variance of the log-likelihood estimate increases linearly with N, but at the same time the curvature of the log likelihood increases and, within the inferentially relevant region, the profile log likelihood becomes increasingly close to quadratic. Thus, in the context of our methodology, increasing N
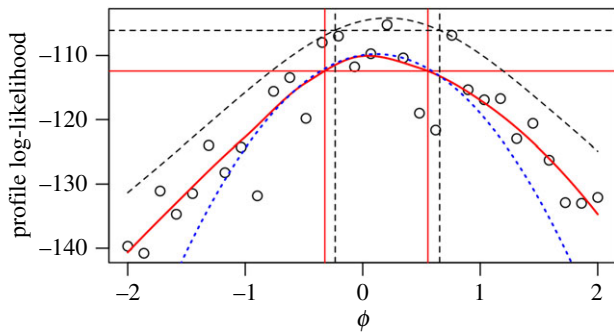
Figure 4. Profile construction for the toy model. The exact profile and its asymptotic 95% CI are constructed with black dashed lines. Points show Monte Carlo profile evaluations. The MCAP is constructed in solid red lines, using the default $\lambda = 0.75$ smoothing parameter. The quadratic approximation used to calculate the MCAP profile cut-off is shown as a dotted blue line. (Online version in colour.)

8

actually makes inference easier despite the increasing Monte Carlo noise. This avoids a paradoxical difficulty of Monte Carlo inference for big data: more data should be a help for a statistician, not a hindrance! Decreasing J represents a situation where Monte Carlo variability increases without increasing information about the parameter of interest. In this case, the Monte Carlo variability and the Monte Carlo bias on the log likelihood due to Jensen's inequality both increase. Also, likelihood maximization becomes more erratic for small J since the maximization error due to the fixed seed becomes more important. However, figure 4 shows that, even when there is considerable bias and variance in the Monte Carlo profile evaluations, the Monte Carlo profile confidence intervals can be little wider than the exact interval.

We computed intervals with nominal coverage of 95%. The MCAP coverage here was 93.4%, compared to 94.3% for the asymptotically exact profile (with a simulation study Monte Carlo standard error of 0.2%). The MCAP intervals were, on average, 12.5% larger than the corresponding exact profile interval, with the increased width accounting for the additional Monte Carlo uncertainty.

Two alternative approaches to generating confidence intervals based on a maximum-likelihood estimator are observed Fisher information and the bootstrap method. Comparisons with these methods on the toy example are presented in the supplement (electronic supplementary material, Section S2). Profile likelihood confidence intervals were found to perform favourably on this example, measured by interval width for a given coverage and given computational effort.

6. Discussion

This paper has focused on likelihood-based confidence intervals. An alternative to likelihood-based inference is to compare the data with simulations using some summary statistic. Various plug-and-play methodologies of this kind have been proposed, such as synthetic likelihood [37] and nonlinear forecasting [38]. For large nonlinear systems, it can be hard to find low-dimensional summary statistics that capture a good fraction of the information in the data. Even summary statistics derived by careful scientific or statistical reasoning have been found surprisingly uninformative compared to

the whole data likelihood in both scientific investigations [39] and simulation experiments [40].

Much attention has been given to scaling Bayesian computation to complex models and large data. Bayesian computation is closely related to likelihood inference for stochastic dynamic models: the random variables generating a dynamic system are typically not directly observed, and these latent random variables are therefore similar to Bayesian parameters. We refer to these latent random variables as random effects since they have a similar role as linear model random effects. To carry out inference on the structural parameters of the model (i.e. the vector θ in this article) the Bayesian approach looks for the marginal posterior of θ, which involves integration over the random effects. Likelihood-based inference for θ similarly involves integrating out the random effects. Numerical methods such as expectation propagation (EP) [41] and variational Bayes [42] are effective for some model classes. Another approach is to combine MCMC computations on subsets of the data, as in the posterior interval estimation (PIE) method of Li *et al.* [43]. The above approaches (EP, VB and PIE) all emphasize situations where the joint density of the data and latent variables can be conveniently split up into conditionally independent chunks, such as a hierarchical model structure. Our methodology has no such requirement. The panel model example above does have a natural hierarchical structure, with individual panels being independent (in the frequentist model sense) or conditionally independent given the shared parameters (in the Bayesian model sense). Our genetic example, and the spatio-temporal example of the electronic supplementary material, Section S1, do not have such a representation.

Some simulation-based Bayesian computation methodologies have built on the observation that unbiased Monte Carlo likelihood computations can be used inside an MCMC algorithm [44]. For large systems, high Monte Carlo variability of likelihood estimates is a concern, in this context, since it slows down MCMC convergence [45]. Doucet *et al.* [46] found that, for a given computational budget, the optimal balance between number of MCMC iterations and time spent on each likelihood evaluation occurs at a Monte Carlo likelihood standard deviation of one log unit. For the systems we demonstrate, Monte Carlo errors that small are not computationally feasible.

Our simple and general approach permits inference when the signal-to-noise ratio in the Monte Carlo profile log likelihood is sufficient to uncover the main features of this function, up to an unimportant vertical shift. For large datasets in which the signal (quantified as the curvature of the log likelihood) is large, the methodology can be effective even when the Monte Carlo noise is far too big to carry out standard MCMC techniques. Although the frequentist motivation for likelihood-based inference differs from the goal of Bayesian posterior inference, both approaches can be used for deductive scientific reasoning [47,48].

Our methodology builds on the availability of Monte Carlo algorithms to evaluate and maximize the likelihood. If these Monte Carlo algorithms are completely overwhelmed by the problem at hand, our method will fail. Geometric features of the likelihood surface, such as nonlinear ridges and multimodality, can lead to challenges for all numerical methods including Monte Carlo approaches. High dimensionality can also be problematic, particularly if combined with difficult geometry. The presence of challenging

characteristics leads to the high Monte Carlo error that motivates and necessitates methodology such as ours. However, if the Monte Carlo component of the MCAP standard error is large relative to the statistical component ($SE_{mc} \gg SE_{stat}$) and also too large to be useful for the scientific application, we diagnose that our method has failed.

References

1. Diggle PJ, Gratton RJ. 1984 Monte Carlo methods of inference for implicit statistical models. *J. R. Stat. Soc. B (Stat. Methodol.)* **46**, 193–227.

2. Bretó C, He D, Ionides EL, King AA. 2009 Time series analysis via mechanistic models. *Ann. Appl. Stat.* **3**, 319–348. (doi:10.1214/08-AOAS201)

3. Pawitan Y. 2001 *In all likelihood: statistical modelling and inference using likelihood*. Oxford, UK: Clarendon Press.

4. Barndorff-Nielsen OE, Cox DR. 1994 *Inference and asymptotics*. London, UK: Chapman and Hall.

5. Bérard J, Del Moral P, Doucet A. 2014 A lognormal central limit theorem for particle approximations of normalizing constants. *Electron. J. Probab.* **19**, 1–28. (doi:10.1214/ejp.v19-3428)

6. He D, Ionides EL, King AA. 2010 Plug-and-play inference for disease dynamics: measles in large and small towns as a case study. *J. R. Soc. Interface* **7**, 271–283. (doi:10.1098/rsif.2009.0151)

7. Marjoram P, Molitor J, Plagnol V, Tavaré S. 2003 Markov chain Monte Carlo without likelihoods. *Proc. Natl Acad. Sci. USA* **100**, 15 324–15 328. (doi:10.1073/pnas.0306899100)

8. Sisson SA, Fan Y, Tanaka MM. 2007 Sequential Monte Carlo without likelihoods. *Proc. Natl Acad. Sci. USA* **104**, 1760–1765. (doi:10.1073/pnas.0607208104)

9. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MP. 2009 Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* **6**, 187–202. (doi:10.1098/rsif.2008.0172)

10. Kevrekidis IG, Gear CW, Hummer G. 2004 Equation-free: the computer-assisted analysis of complex, multiscale systems. *Am. Inst. Chem. Eng. J.* **50**, 1346–1354. (doi:10.1002/aic.10106)

11. Lyne A-M, Girolami M, Atchade Y, Strathmann H, Simpson D. 2015 On Russian roulette estimates for Bayesian inference with doubly-intractable likelihoods. *Stat. Sci.* **30**, 443–467. (doi:10.1214/15-STS523)

12. Rubio FJ, Johansen AM. 2013 A simple approach to maximum intractable likelihood estimation. *Electron. J. Stat.* **7**, 1632–1654. (doi:10.1214/13-EJS819)

13. Ionides EL. 2005 Maximum smoothed likelihood estimation. *Stat. Sin.* **15**, 1003–1014.

14. Barton RR, Meckesheimer M. 2006 Metamodel-based simulation optimization. *Handb. Oper. Res. Manage. Sci.* **13**, 535–574.

15. LeCam L, Yang GL. 2000 Asymptotics in statistics, 2nd edn. New York, NY: Springer.

16. Murphy SA, van der Vaart AW. 2000 On profile likelihood. *J. Am. Stat. Assoc.* **95**, 449–465. (doi:10.1080/01621459.2000.10474219)

17. R Core Team. 2017 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

18. Cleveland WS, Grosse E, Shyu WM. 1993 Local regression models. In *Statistical Models in S* (eds JM Chambers, TJ Hastie), pp. 309–376. London, UK: Chapman and Hall.

19. Ionides EL, Nguyen D, Atchadé Y, Stoev S, King AA. 2015 Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proc. Natl Acad. Sci. USA* **112**, 719–724. (doi:10.1073/pnas.1410597112)

20. Cappé O, Moulines E, Rydén T. 2005 *Inference in hidden Markov models*. New York, NY: Springer.

21. Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, Holmes EC. 2004 Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327–332. (doi:10.1126/science.1090727)

22. Karcher MD, Palacios JA, Lan S, Minin VN. 2016 phylodyn: an R package for phylodynamic simulation and inference. *Mol. Ecol. Resour.* **17**, 96–100. (doi:10.1111/1755-0998.12630)

23. Smith RA, Ionides EL, King AA. 2017 Infectious disease dynamics inferred from genetic data via sequential Monte Carlo. *Mol. Biol. Evol.* pre-published online. (doi:10.1093/molbev/msx124)

24. Bretó C, Ionides EL, King AA. In preparation. Panel data analysis via mechanistic models.

25. Bjornstad O, Grenfell B. 2008 Hazards, spatial transmission and timing of outbreaks in epidemic metapopulations. *Environ. Ecol. Stat.* **15**, 265–277. (doi:10.1007/s10651-007-0059-3)

26. Martinez-Bakker M, King AA, Rohani P. 2015 Unraveling the transmission ecology of polio. *PLoS Biol.* **13**, e1002172. (doi:10.1371/journal.pbio.1002172)

27. Dobson A. 2014 Mathematical models for emerging disease. *Science* **346**, 1294–1295. (doi:10.1126/science.aaa3441)

28. Ionides EL, Bretó C, King AA. 2006 Inference for nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **103**, 18 438–18 443. (doi:10.1073/pnas.0603181103)

29. King AA, Ionides EL, Pascual M, Bouma MJ. 2008 Inapparent infections and cholera dynamics. *Nature* **454**, 877–880. (doi:10.1038/nature07084)

30. Camacho A, Ballesteros S, Graham AL, Carrat F, Ratmann O, Cazelles B. 2011 Explaining rapid reinfections in multiple-wave influenza outbreaks: Tristan da Cunha 1971 epidemic as a case study. *Proc. R. Soc. B* **278**, 3635–3643. (doi:10.1098/rspb.2011.0300)

31. Blackwood JC, Cummings DAT, Broutin H, lamsirithaworn S, Rohani P. 2013 Deciphering the impacts of vaccination and immunity on pertussis epidemiology in Thailand. *Proc. Natl Acad. Sci. USA* **110**, 9595–9600. (doi:10.1073/pnas.1220908110)

32. Shrestha S, Foxman B, Weinberger DM, Steiner C, Viboud C, Rohani P. 2013 Identifying the interaction between influenza and pneumococcal pneumonia using incidence data. *Sci. Transl. Med.* **5**, 191ra84. (doi:10.1126/scitranslmed.3005982)

33. Blake IM, Martin R, Goel A, Khetsuriani N, Everts J, Wolff C, Wassilak S, Aylward RB, Grassly NC. 2014 The role of older children and adults in wild poliovirus transmission. *Proc. Natl Acad. Sci. USA* **111**, 10 604–10 609. (doi:10.1073/pnas.1323688111)

34. Bengtsson T, Bickel P, Li B. 2008 Curse-of-dimensionality revisited: collapse of the particle filter in very large scale systems. In *Probability and Statistics: Essays in Honor of David A. Freedman* (eds T Speed, D Nolan), pp. 316–334. Beachwood, OH: Institute of Mathematical Statistics.

35. Rebeschini P, van Handel R. 2015 Can local particle filters beat the curse of dimensionality? *Ann. Appl. Probab.* **25**, 2809–2866. (doi:10.1214/14-AAP1061)

36. Beskos A, Crisan D, Jasra A, Kamatani K, Zhou Y. 2014 A stable particle filter in high-dimensions. (http://arxiv.org/abs/1412.3501)

37. Wood SN. 2010 Statistical inference for noisy nonlinear ecological dynamic systems. *Nature* **466**, 1102–1104. (doi:10.1038/nature09319)

38. Ellner SP, Bailey BA, Bobashev GV, Gallant AR, Grenfell BT, Nychka DW. 1998 Noise and nonlinearity in measles epidemics: combining mechanistic and

statistical approaches to population modeling. *Am. Nat.* **151**, 425 – 440. (doi:10.1086/286130)

39. Shrestha S, King AA, Rohani P. 2011 Statistical inference for multi-pathogen systems. *PLoS Comput. Biol.* **7**, e1002135. (doi:10.1371/journal.pcbi.1002135)

40. Fasiolo M, Pya N, Wood SN. 2016 A comparison of inferential methods for highly nonlinear state space models in ecology and epidemiology. *Stat. Sci.* **31**, 96 – 118. (doi:10.1214/15-STS534)

41. Gelman A, Vehtari A, Jylänki P, Robert C, Chopin N, Cunningham JP. 2014 Expectation propagation as a way of life. (http://arxiv.org/abs/1412.4869).

42. Hoffman MD, Blei DM, Wang C, Paisley JW. 2013 Stochastic variational inference. *J. Mach. Learn. Res.* **14**, 1303 – 1347.

43. Li C, Srivastava S, Dunson DB. 2016 Simple, scalable and accurate posterior interval estimation. (http://arxiv.org/abs/1605.04029)

44. Andrieu C, Roberts GO. 2009 The pseudo-marginal approach for efficient computation. *Ann. Stat.* **37**, 697 – 725. (doi:10.1214/07-AOS574)

45. Bardenet R, Doucet A, Holmes C. 2015 On Markov chain Monte Carlo methods for tall data. (http://arxiv.org/abs/1505.02827)

46. Doucet A, Pitt MK, Deligiannidis G, Kohn R. 2015 Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika* **102**, 295 – 313. (doi:10.1093/biomet/asu075)

47. Gelman A, Shalizi CR. 2013 Philosophy and the practice of Bayesian statistics. *British. J. Math. Stat. Psychol.* **66**, 8 – 38. (doi:10.1111/j.2044-8317.2011.02037.x)

48. Ionides EL, Giessing A, Ritov Y, Page SE. 2017 Response to the ASA's statement on *p*-values: Context, process, and purpose. *Am. Stat.* **71**, 88 – 89. (doi:10.1080/00031305.2016.1234977)