

Ecological non-linear state space model selection via adaptive particle Markov chain Monte Carlo (AdPMCMC)

Gareth W. Peters

UNSW Mathematics and Statistics Department, Sydney, 2052, Australia.

Geoffrey R. Hosack

CSIRO Mathematics, Informatics and Statistics, GPO Box 1538, Hobart

Keith R. Hayes

CSIRO Mathematics, Informatics and Statistics, GPO Box 1538, Hobart

Summary. We develop a novel advanced Particle Markov chain Monte Carlo algorithm that is capable of sampling from the posterior distribution of non-linear state space models for both the unobserved latent states and the unknown model parameters. We apply this novel methodology to five population growth models, including models with strong and weak Allee effects, and test if it can efficiently sample from the complex likelihood surface that is often associated with these models. Utilising real and also synthetically generated data sets we examine the extent to which observation noise and process error may frustrate efforts to choose between these models. Our novel algorithm involves an Adaptive Metropolis proposal combined with an SIR Particle MCMC algorithm (AdPMCMC). We show that the AdPMCMC algorithm samples complex, high-dimensional spaces efficiently, and is therefore superior to standard Gibbs or Metropolis Hastings algorithms that are known to converge very slowly when applied to the non-linear state space ecological models considered in this paper. Additionally, we show how the AdPMCMC algorithm can be used to recursively estimate the Bayesian Cramér-Rao Lower Bound of [Tichavský et al., 1998]. We derive expressions for these Cramér-Rao Bounds and estimate them for the models considered. Our results demonstrate a number of important features of common population growth models, most notably their multi-modal likelihood surfaces and dependence between the static and dynamic parameters. We conclude by sampling from the posterior distribution of each of the models, and use Bayes factors to highlight how observation noise significantly diminishes our ability to select among some of the models, particularly those that are designed to reproduce an Allee effect. These result have important ramifications for ecologists searching for evidence of all forms of density dependence in population time series.

Key words: Adaptive Metropolis, Particle Markov chain Monte Carlo, Sequential Monte Carlo, Particle Filtering, recursive Bayesian Cramér-Rao Lower Bound, Model selection

1. Introduction

Ecologists in recent years have begun to fit Bayesian state space models to ecological time series, [Millar and Meyer, 2000, Clark and Bjornstad, 2004, Ward, 2006]. State space models (SSMs) avoid biases that can occur when either the observations are assumed to be perfect or the process model is treated deterministically [Carrol et al., 1995, Shenk et al., 1998, Calder et al., 2003, Freckleton et al., 2006]. SSM inference involves jointly estimating the latent state vector and the static parameter vector of the models for observations and states. Many ecological studies that estimate density dependence using time series data, either do not include jointly estimated observation error or exclude it altogether [Sibly et al., 2005, Sæther et al., 2008, Gregory et al., 2010]. Studies that do incorporate process error and observation noise into ecological inference sometimes use very special cases of linear state and observation equations [Lindley, 2003, Williams et al., 2003, Linden and Knape, 2009]. More realistic, non-linear models may be implemented sub-optimally in an Extended Kalman Filter [Wang, 2007, Zeng et al., 1998, Welch and Bishop, 1995]. Linear approximations of non-linear state equations [Wang, 2007] can allow for a combination of suboptimal Kalman filtering and Gibbs sampling to be performed. Linear approximation, however, can lead to inaccurate parameter posterior inference for non-linear SSM with non-negligible observation noise and process error.

Studies that employ the Metropolis-Hastings within Gibbs (MHG) algorithm avoid these restrictive linear assumptions or local linearizations, [Millar and Meyer, 2000, Ward, 2006, Clark and Bjornstad, 2004, Sæther et al., 2007], but the performance of MHG is known to be very sensitive to the process model and observation model parameterisation as discussed in [Andrieu et al., 2010]. The standard MHG algorithm can be difficult to tune for non-linear ecological SSMs, [Millar and Meyer, 2000], for two reasons. Firstly, non-negligible posterior correlations can occur in the high dimensional parameter space that is composed of both the latent state process and the static parameters. Secondly, the joint likelihood surfaces of the static parameters for popular non-linear population growth models (irrespective of observation error) can be multi-modal and have strong sharp flat ridges, [Polansky et al., 2009]. Furthermore, an MHG algorithm that has not sufficiently mixed over modes and ridges will corrupt all model selection and model averaging routines.

In this paper, we describe a novel technical development that is specifically designed to address non-linear dynamics in SSMs. The method uses a Particle MCMC algorithm for joint process and parameter estimation in non-linear and non-Gaussian SSMs [Andrieu et al., 2010], coupled to an adaptive Metropolis proposal which we denote by the Adaptive Particle Markov chain Monte Carlo algorithm (AdPMCMC). We apply this novel methodology to synthetic and real data sets. Our objectives are to examine the performance of the new algorithm in non-linear population growth models, with complex likelihood surfaces, and to re-examine the problems of model selection with Bayes factors estimated via Markov chain samples. We find that the novel methodology is capable of efficiently sampling highly non-linear models without careful tuning in very high dimensional posterior models.

Complicated likelihood surfaces with multimodal characteristics present a serious challenge to Bayesian posterior samplers, [Polansky et al., 2009]. The failure of the basic MHG sampler to mix over the support

of the posterior in such situations is a known concern in this case for ecological modellers and in general a well known problem in the statistics literature, see discussions and illustrations in [Carlin and Louis, 1997, Robert, 2001, Bishop et al., 2006]. Our AdPMCMC methodology will be shown to be capable of efficiently exploring the posterior distributions obtained from typical ecological growth models, overcoming many of the poor mixing properties of the basic MHG sampler in this context. Additionally, we use Bayes factors to highlight the ambiguity in model structure when fitting models to short ecological time series and therefore affirm that predictions outside the observed data made by a single best fitting model can be misleading. This is especially important when the intention is to capture effects that are difficult to observe, such as the Allee effect. It is precisely in estimation of quantities such as Bayes factors, that one requires a sampling methodology capable of efficiently delivering samples from the entire support of the posterior distribution.

1.1. Contribution

This paper addresses three main concerns about model fitting to ecological time series. The first of these is the inclusion of observation error. Wang [2007], argues that observation error should be included in ecological modelling when estimating abundance of animals because measurement errors arise from several sources, for example unaccounted effects of weather on the probability of detecting animals or random failures of trapping devices. Ignoring observation error can lead to biased estimates of the parameters [Carrol et al., 1995]. We argue it will also have an important effect on model selection and the ability to detect ecological effects such as density dependence and the strength of an Allee effect.

The second issue involves model simplification. It is difficult to jointly estimate the latent process with the model parameters, especially when the state or observation process are non-linear and there is strong dependence between model parameters and the latent process. The approach we present will allow us to work directly with realistic non-linear models in the presence of arbitrarily high process and observation noise settings. Moreover, we do not introduce any approximation error as we have no need to linearise such models to perform filtering, such as in the approaches discussed in [Wang, 2007].

The third issue relates to efficient estimation and robust model selection. To date relatively simple Metropolis-Hastings and Gibbs sampling routines have been used to fit ecological state space models [Millar and Meyer, 2000, Clark and Bjornstad, 2004, Ward, 2006, Sæther et al., 2007]. These algorithms are prone to convergence difficulties for two reasons: the first is due to non-negligible posterior correlations that will occur in the high dimensional parameter space that is composed of both the latent state process and the static parameters; the second is due to posterior multimodality. The AdPMCMC we propose is specifically designed for joint process and parameter estimation in non-linear and non-Gaussian states space models.

1.2. Structure and Notation

The paper is structured as follows. In Section 2, we present five common population growth models, including four that are non-linear in the states, parameters or both. Two of these models include weak or strong Allee effects which can be important in ecological populations [Dennis, 2002]. In Section 3, we summarise a number of issues regarding Bayesian estimation, prior selection and the likelihood surface that are important and

relevant to the issues identified above. Section 4 presents the details of the estimation and model selection procedure, including the AdPMCMC algorithm. Section 5 presents the results and discussion which is split into two parts. The first part considers results for synthetic data and the performance of the algorithm. The second part presents the results for two real data sets under the newly developed sampling methodology. The paper concludes with brief discussion and recommendations for future developments including more sophisticated adaption routines and non-Gaussian error structures.

The notation used throughout this paper will involve: capitals to denote random variables and lower case to denote realizations of random variables; bold face to denote vectors and non-bold for scalars; subscript will denote discrete time, where $n_{1:T}$ denotes n_1, \dots, n_T . In the sampling methodology combining MCMC and particle filtering (Sequential Monte Carlo - SMC), we use the notation $[\cdot](j, i)$ to denote the j -th state of the Markov chain for the i -th particle in the particle filter. Note, the index i is dropped for quantities not involving the particle filter. We will define a proposed new state prior to acceptance at iteration j of the Markov chain by $[\cdot](j)'$. In addition, we note that $\boldsymbol{\theta}$ generically represents the particular model parameters under consideration, for example M_0 has $\boldsymbol{\theta} = (b_0, \sigma_\epsilon, \sigma_w)$ and M_4 will have $\boldsymbol{\theta} = (b_5, b_6, b_7, , \sigma_\epsilon, \sigma_w)$.

2. Ecological State Space Models

This section presents the ecological state space models considered, with discussion from the ecological modelling perspective, regarding density dependence and weak or strong Allee effects.

2.1. Observation equations

The generic observation equation we consider is,

$$y_t = g_t(n_t) + w_t \quad (1)$$

where $w_t \sim N(0, \sigma_w)$. The observation model acknowledges that we typically do not observe the entire population of interest, but rather a sample realization and that the observation mechanisms we use are imperfect. It therefore reflects all sources of variability introduced by the data-generating mechanism. The Gaussian observation error assumption for log transformed abundances is common and widely applied in ecological contexts, e.g. [Clark and Bjornstad, 2004]. Importantly, this assumption can be easily extended to include more flexible classes of observation noise, such as α - stable models [Peters et al., 2009] or others recently proposed in the ecological literature [Stenseth et al., 2003, Ward, 2006], under the methodological framework that we develop in this paper.

2.2. Process equations

We consider five different stochastic population dynamic models that encapsulate different ecological effects such as density dependent mortality and strong and weak Allee effects. Each of these process models describes how the number of individuals in a population affect the subsequent growth of the population. In these models N_t represents a continuous latent random variable for the population size. The process error in

these models ($\epsilon_t \sim N(0, \sigma_\epsilon^2)$) reflects all sources of variability in the underlying population growth process that is not captured by the model. Process error in all of the models is assumed to behave multiplicatively on the natural scale, see [Clark, 2007].

2.2.1. The exponential growth equation (M_0)

The exponential growth model provides a simple density independent model for comparison with potentially more realistic density dependent models. The discrete, exponential growth log transformed equation is

$$\log N_{t+1} = \log N_t + b_0 + \epsilon_t. \quad (2)$$

where $b_0 = r$ is the maximum per-individual growth rate (defined as the difference between the per individual birth and death rates).

2.2.2. The Ricker equation (M_1)

The discrete Ricker equation is a density dependent model that is similar to the familiar logistic growth model. After log transformation it is given by,

$$\log N_{t+1} = \log N_t + b_0 + b_1 N_t + \epsilon_t, \quad (3)$$

where b_0 is the density-independent growth rate and b_1 governs the strength of density dependence. For populations that exhibit negative density dependence ($b_1 < 0$), the “carrying capacity” of the environment (usually denoted K) is defined by the stable equilibrium at $-\frac{b_0}{b_1}$ so long as the density independent growth rate is positive ($b_0 > 0$). If $b_0 < 0$ then the only stable equilibrium is located at 0. Whilst linear in its parameters, this model is non-linear in the latent state because it contains an exponential term in N_t .

2.2.3. The theta-logistic equation (M_2)

[Lande et al., 2003] recommend the theta-logistic equation when the form of the density dependence in the population dynamics is unknown, given here on the log scale

$$\log N_{t+1} = \log N_t + b_0 + b_2 (N_t)^{b_3} + \epsilon_t, \quad (4)$$

where b_3 determines the form of density dependence. For the carrying capacity ($K = \left(-\frac{b_0}{b_2}\right)^{\frac{1}{b_3}}$) to exist, b_0 and b_2 must be of opposing sign, and it is stable only if b_0 and b_3 are of the same sign.

2.2.4. The “mate-limited” logistic equation (M_3)

[Morris and Doak, 2002] propose a “mate-limited” Allee effect equation, presented here under a log transformation,

$$\log N_{t+1} = 2 \log N_t - \log(b_4 + N_t) + b_0 + b_1 N_t + \epsilon_t, \quad (5)$$

where $b_4 > 0$ represents the population size at which the per-individual birth rate is half of what it would be if mating was not limited, thereby controlling the population size at which Allee effects are “noticeable”. [Dennis, 2002] notes that mate-limited models can describe, in a phenomenological fashion, Allee effects from other biological mechanisms besides mate limitation.

2.2.5. The “Flexible-Allee” logistic equation (M_4)

This modified version of the Ricker model allows for both strong and weak Allee effects,

$$\log N_{t+1} = \log N_t + b_5 + b_6 N_t + b_7 N_t^2 + \epsilon_t, \quad (6)$$

where $K = \frac{(-b_6 - \sqrt{b_6^2 - 4b_5 b_7})}{2b_7}$, and $C = \frac{(-b_6 + \sqrt{b_6^2 - 4b_5 b_7})}{2b_7}$. If these roots are real, then C represents the threshold of population size below which per capita population growth is negative. In the deterministic model, when $0 < C < K$, the Allee effect is strong and it is possible for an unstable equilibrium at $N = C$ to occur between two stable equilibria, $N = 0$ and $N = K$. When $C < 0$, the Allee effect is weak and a single stable equilibria occurs at $N = K$. If C and K are not real, then $N = 0$ is the only stable equilibrium. This model is a discrete version of a continuous time model previously used in theoretical studies of spatial population dynamics [Lewis and Kareiva, 1993]; see [Boukal and Berec, 2002] for further discussion of Allee effects in discrete time population models. Note that model M_4 differs from model M_3 in that the latter can only exhibit a strong Allee effect.

The process models M_0 to M_4 are nested models. Specifically, setting $b_7 = 0$ in model M_4 and $b_4 = 0$ in model M_3 returns model M_1 . Setting $b_3 = 1$ in model M_2 also returns model M_1 . Setting $b_1 = 0$ in model M_1 returns the density independent model M_0 .

3. Bayesian Estimation

Our approach estimates the latent process and model parameters for models M_0 through to M_4 using advanced Bayesian inference. The Markov stochastic process $N_{0:T}$ is unobserved (latent) and must be estimated over the time interval $[0, T]$ at discrete time points $t = 0$ through to T . In this setting we are interested in estimating the entire data set as opposed to a sequential, real-time estimates of the latent states jointly with the model parameters. As such the posterior of interest to this problem is given by $p(\boldsymbol{\theta}, n_{1:T} | y_{1:T}, M_i)$, where we have absorbed the initial state value (N_0) into the vector of static parameters $\boldsymbol{\theta}$.

3.1. Priors, Likelihood and Posterior

Table 1 summarises the priors selected for the process and observation model parameters. The process model parameters $b_0, b_1, b_2, b_3, b_5, b_6$ and b_7 are given Gaussian distribution priors. The prior for b_4 is assumed to be nonnegative [Morris and Doak, 2002] and is given a Gamma prior distribution. In addition, we give the noise variances $\sigma_\epsilon^2, \sigma_w^2$ inverse gamma priors with the parameters $\alpha_\epsilon = \alpha_w = \frac{T}{2}$ and $\beta_\epsilon = \beta_w = \frac{2(\alpha_\epsilon - 1)}{10}$.

Table 1. Priors for the parameters.

Parameters	Priors
$b_{0,1,2,3,5,6,7}$	$N(0, 1)$
b_4	$G(\text{shape} = 1, \text{scale} = 10)$
σ_ϵ^2	$IG(\alpha_\epsilon, \beta_\epsilon)$
σ_w^2	$IG(\alpha_w, \beta_w)$

We also assume that observations are independent and identically distributed. Hence, the generic likelihood is given by

$$\mathcal{L}(n_{1:T}, \boldsymbol{\theta}; y_{1:T}) = \prod_{t=1}^T p(y_t | n_t, \sigma_w^2), \quad (7)$$

where $p(y_t | n_t, \sigma_w^2)$ is a Gaussian distribution with mean n_t and unknown variance σ_w^2 . We assume that the observations are conditionally independent given the latent state, with latent states under a first order Markov dependence. This produces a target posterior distribution given by

$$p(\boldsymbol{\theta}, n_{1:T} | y_{1:T}) \propto p(\boldsymbol{\theta}) \prod_{t=1}^T p(y_t | n_t, \boldsymbol{\theta}) p(n_t | n_{t-1}, \boldsymbol{\theta}). \quad (8)$$

3.2. Estimation and Model Selection

Having specified the posterior distribution, inference and model selection proceeds by estimating popular Bayesian statistics such as the Minimum Mean Square Error (MMSE = posterior mean) and evaluating the posterior evidence for each model. In our context this requires estimates of the following quantities

$$\begin{aligned} \text{MMSE: } & (\hat{N}_{1:T}^{MMSE}, \hat{\boldsymbol{\Theta}}^{MMSE}) = \int \boldsymbol{\Theta} \int N_{1:T} p(\boldsymbol{\theta}, n_{1:T} | y_{1:T}, M_i) dN_{1:T} d\boldsymbol{\Theta}, \\ \text{Posterior evidence } M_i: & p(y_{1:T} | M_i) = \int p(y_{1:T} | \boldsymbol{\theta}, n_{1:T}, M_i) p(\boldsymbol{\theta}, n_{1:T} | M_i) dN_{1:T} d\boldsymbol{\theta}. \end{aligned} \quad (9)$$

Estimating these quantities for non-linear state space models presents a considerable statistical challenge. The dimension of the posterior distribution is $\dim(\boldsymbol{\Theta}) + T$, hence for moderate T (≈ 50 to 100) the dimension is large. Controlling the variance of these Bayesian estimators requires efficient posterior sampling methods from the posterior distribution, which we achieve via our adaptive version of the PMCMC algorithm of [Andrieu et al., 2010].

3.3. Adaptive MCMC within Particle MCMC (AdPMCMC).

The aim of this section is to present a novel methodology to sample from the posterior distribution given in Equation (8), based on a version of a recent sampler specifically developed for use in state space models; see [Andrieu et al., 2010]. The methodology is known as PMCMC, and it represents a state of the art sampling framework for state space problems. These samples can then be used to form Monte Carlo estimates for Equations (9). The fundamental innovation we present in this paper is to combine an Adaptive MCMC algorithm within the PMCMC framework.

The key advantage of the PMCMC algorithm is that it allows one to jointly update the entire set of posterior parameters $(\boldsymbol{\Theta}, N_{1:T})$ and only requires calculation of the marginal acceptance probability in the Metropolis-Hastings algorithm. PMCMC achieves this by embedding a particle filter estimate of the optimal proposal distribution for the latent process into the MCMC algorithm. This allows the Markov chain to mix efficiently in the high dimensional posterior parameter space because the particle filter approximation of the optimal proposal distribution in the MCMC algorithm, thereby allowing high-dimensional parameter block updates even in the presence of strong posterior parameter dependence. The models considered in this paper for example have between three and seven static parameters to be estimated jointly with the latent state space, resulting in an additional T parameters, where T can range from the tens to the hundreds, or even thousands, depending on the data set considered.

In the state space setting, the Particle MCMC algorithm used to sample from the target distribution $p(\boldsymbol{\theta}, N_{1:T} | y_{1:T})$ proceeds by mimicking the marginal Metropolis-Hastings algorithm in which the acceptance probability is given by,

$$\alpha([\boldsymbol{\theta}, n_{1:T}](j-1), [\boldsymbol{\theta}, n_{1:T}](j)') = \min \left(1, \frac{p([\boldsymbol{\theta}](j)' | y_{1:T}) q([\boldsymbol{\theta}](j)', [\boldsymbol{\theta}](j-1))}{p([\boldsymbol{\theta}](j-1) | y_{1:T}) q([\boldsymbol{\theta}](j-1), [\boldsymbol{\theta}](j)')} \right).$$

where $q([\boldsymbol{\theta}](j-1), [\boldsymbol{\theta}](j))$ is the proposal distribution of the PMCMC generated Markov chain for the static parameters to propose a move from state $[\boldsymbol{\theta}](j-1)$ at iteration $j-1$ to a new state $[\boldsymbol{\theta}](j)$ at iteration j .

To achieve this we split the standard Metropolis Hastings proposal distribution into two components. The first constructs a proposal kernel via an adaptive Metropolis scheme ([Roberts and Rosenthal, 2009], [Atchadé and Rosenthal, 2005]) that is used to sample the static parameters $\boldsymbol{\Theta}$. Introducing an adaptive MCMC proposal kernel into the Particle MCMC setting allows the Markov chain proposal distribution to adaptively learn the regions of the marginal posterior distribution of the static model parameters that have the most mass. This significantly improves the acceptance probability of the proposal distribution and enables much more rapid and efficient mixing of the Markov chain for a small number of particles L and a simple particle filter algorithm. The second component of the proposal kernel constructs an estimate of the posterior distribution of the latent states, $N_{1:T}$, allowing us to sample a proposed trajectory. This proposal kernel is constructed via a Sequential Monte Carlo algorithm that is based on a simple SIR filter, ([Gordon et al., 1993], [Doucet and Johansen, 2009]). Note the SIR filter is suitable for PMCMC applications since it is only used as a proposal distribution and not as an empirical estimate of the posterior as is more common; see discussion of standard application in [Doucet and Johansen, 2009]. In particular the proposal kernel approximates the optimal choice

$$q([\boldsymbol{\theta}, n_{1:T}](j-1), [\boldsymbol{\theta}, n_{1:T}](j)') = q([\boldsymbol{\theta}](j-1), [\boldsymbol{\theta}](j)') p([n_{1:T}](j)' | y_{1:T}, [\boldsymbol{\theta}](j)'), \quad (10)$$

via an adaptive MCMC proposal for $q([\boldsymbol{\theta}](j-1), [\boldsymbol{\theta}](j)')$ and a particle filter (SMC) estimate for $p([n_{1:T}](j)' | y_{1:T}, [\boldsymbol{\theta}](j)')$. The SMC algorithm proposal samples (approximately) from the sequence of distributions, $\{p(n_{1:t} | y_{1:t}, \boldsymbol{\theta})\}_{t=1:T}$. For a recent review of SMC methodology, of which there are several different methodologies, see [Doucet and Johansen, 2009].

When this proposal is substituted into the standard Metropolis Hastings acceptance probability, several terms cancel to produce,

$$\alpha([\boldsymbol{\theta}, n_{1:T}](j-1), [\boldsymbol{\theta}, n_{1:T}](j)') = \min\left(1, \frac{p(y_{1:T} | [\boldsymbol{\theta}](j)') p([\boldsymbol{\theta}](j)') q([\boldsymbol{\theta}](j)', [\boldsymbol{\theta}](j-1))}{p(y_{1:T} | [\boldsymbol{\theta}](j-1)) p([\boldsymbol{\theta}](j-1)) q([\boldsymbol{\theta}](j-1), [\boldsymbol{\theta}](j)')}\right).$$

We can now detail a generic version of the AdPMCMC methodology we developed.

3.3.1. Generic Particle MCMC

One iteration of the generic AdPMCMC algorithm proceeds as follows:

- (a) Sample $[\boldsymbol{\theta}](j)' \sim q([\boldsymbol{\theta}](j-1), \cdot)$ from an Adaptive MCMC proposal.(Appendix 1. Algorithm 2).
- (b) Run an SMC algorithm with L particles to obtain:

$$\begin{aligned} \hat{p}(n_{1:T} | y_{1:T}, [\boldsymbol{\theta}](j)') &= \sum_{i=1}^L W_{1:T}^{(i)} \delta_{[n_{1:T}](j,i)'} (n_{1:T}) \\ \hat{p}(y_{1:T} | [\boldsymbol{\theta}](j)') &= \prod_{t=1}^T \left(\frac{1}{L} \sum_{i=1}^L w_t ([n_t](j,i)') \right) \end{aligned} \quad (11)$$

Then sample a candidate path $[N_{1:T}](j)' \sim \hat{p}(n_{1:T} | y_{1:T}, [\boldsymbol{\theta}](j)').$ (see Appendix 1. Algorithm 1)

- (c) Accept the proposed new Markov chain state comprised of $[\boldsymbol{\theta}, N_{1:T}](j)'$ with acceptance probability given by

$$\alpha([\boldsymbol{\theta}, n_{1:T}](j-1), [\boldsymbol{\theta}, n_{1:T}](j)') = \min\left(1, \frac{\hat{p}(y_{1:T} | [\boldsymbol{\theta}](j)') p([\boldsymbol{\theta}](j)') q([\boldsymbol{\theta}](j)', [\boldsymbol{\theta}](j-1))}{\hat{p}(y_{1:T} | [\boldsymbol{\theta}](j-1)) p([\boldsymbol{\theta}](j-1)) q([\boldsymbol{\theta}](j-1), [\boldsymbol{\theta}](j)')}\right) \quad (12)$$

where $\hat{p}(y_{1:T} | [\boldsymbol{\theta}](j-1))$ is obtained from the previous iteration of the PMCMC algorithm.

The key advantage of this approach is that the difficult problem of designing high dimensional proposals has been replaced with the simpler problem of designing low dimensional mutation kernels in the Sequential Monte Carlo algorithm embedded in the MCMC algorithm. This sampling approach in which Sequential Monte Carlo is used to approximate the marginal likelihood in the acceptance probability has been shown to have several theoretical convergence properties. In particular the empirical law of the particles converges to the true filtering distribution at each iteration as a bounded linear function of time t and the number of particles L ; see [Andrieu et al., 2010]. This means it is possible to construct and efficiently sample approximately optimal path space proposals with linear cost.

3.3.2. Generic Adaptive MCMC

We utilise the adaptive MCMC algorithm to learn the proposal distribution for the static parameters in our posterior Θ . There are several classes of adaptive MCMC algorithms; see [Andrieu and Atchade, 2006]. The distinguishing feature of adaptive MCMC algorithms, (compared to standard MCMC), is that the Markov chain is generated via a sequence of transition kernels. Adaptive algorithms utilise a combination of time or state inhomogeneous proposal kernels. Each proposal in the sequence is allowed to depend on the past history of the Markov chain generated, resulting in many possible variants.

When using inhomogeneous Markov kernels it is particularly important to ensure the generated Markov chain is ergodic, with the appropriate stationary distribution. Several recent papers proposing theoretical conditions that must be satisfied to ensure ergodicity of adaptive algorithms include, [Andrieu and Moulines, 2006] and [Haario et al., 2005]. In particular [Roberts and Rosenthal, 2009] proved ergodicity of adaptive MCMC under conditions known as *Diminishing Adaptation* and *Bounded Convergence*.

It is non-trivial to develop adaption schemes that are easily verified to satisfy these two conditions. In this paper use a mixture proposal kernel known to satisfy these two ergodicity conditions when unbounded state spaces and general classes of target posterior distributions are utilised; see [Roberts and Rosenthal, 2009].

3.4. Adaptive Metropolis within SIR Particle MCMC (AdPMCMC).

In this section (and Appendix 1) we present the details of the Adaptive Metropolis within Particle MCMC (AdPMCMC) algorithm used to sample from the posterior on the path space of our latent states and model parameters. This involves specifying the details of the proposal distribution in Equation 10. The proposal, $q([\boldsymbol{\theta}](j-1), [\boldsymbol{\theta}](j)')$, involves an adaptive Metropolis proposal comprised of a mixture of Gaussians, one component of which has a covariance structure that is adaptively learnt on-line as the algorithm progressively explores the posterior distribution. The mixture proposal distribution for parameters $\boldsymbol{\theta}$ is given at iteration j of the Markov chain by,

$$q_j([\boldsymbol{\theta}](j-1), \cdot) = w_1 N\left(\boldsymbol{\theta}; [\boldsymbol{\theta}](j-1), \frac{(2.38)^2}{d} \Sigma_j\right) + (1 - w_1) N\left(\boldsymbol{\theta}; [\boldsymbol{\theta}](j-1), \frac{(0.1)^2}{d} I_{d,d}\right). \quad (13)$$

Here, Σ_j is the current empirical estimate of the covariance between the parameters of $\boldsymbol{\theta}$ estimated using samples from the Particle Markov chain up to time j . The theoretical motivation for the choices of scale factors 2.38, 0.1 and dimension d are all provided in [Roberts and Rosenthal, 2009] and are based on optimality conditions presented in [Roberts and Rosenthal, 2001].

The proposal kernel for the latent states $N_{1:T}$, given by $p([n_{1:T}](j)' | y_{1:T}, [\boldsymbol{\theta}](j)')$, uses the simplest SMC algorithm known as the SIR algorithm. Therefore the mutation kernel, is given by the process model, in our case $N(N_t; f_t(N_{t-1}, \boldsymbol{\theta}), \sigma_\epsilon^2)$.

4. Results and Analysis

This section is split into four subsections, the first two subsections study the performance and estimation accuracy of the AdPMCMC algorithm using synthetic data generated from models (M_0, \dots, M_4) with known parameters. We gauge the accuracy of the algorithm by comparing the marginalised average estimated mean square error (MSE) of the MMSE estimate of the latent state process $N_{1:T}$ (estimated over 20 blocks per data set with 20 data sets for the average) to a recursively defined Bayesian Cramer-Rao Lower Bound (BCRLB). In the third subsection, we study the performance of Bayes factor estimates for each of the models using evidence estimates derived from Markov chain samples obtained via the AdPMCMC algorithm. In the final sub-section, we perform parameter estimation and model selection using two real data sets.

Each iteration of the AdPMCMC algorithm an approximation to the optimal proposal is constructed and sampled to produce a proposed Markov chain state update in dimension $\dim(\boldsymbol{\Theta}) + T$. Additionally, all simulation studies in this paper involved generation of a Markov chain via the AdPMCMC algorithm with the following three stages:

stage 1: ANNEALED PMCMC - This stage involves a random initialization of the posterior parameters $(\boldsymbol{\theta}, N_{1:T})$ followed by an annealed AdPMCMC algorithm, using a sequence of distributions

$$p_n(\boldsymbol{\theta}, N_{1:T} | y_{1:T}) = p(\boldsymbol{\theta}, N_{1:T})^{1-\gamma_n} p(y_{1:T} | \boldsymbol{\theta}, N_{1:T})^{\gamma_n},$$

where γ_n increases linearly from 10^{-5} to 1 for the first 5,000 AdPMCMC steps. The annealing must also be integrated into the construction of the particle filter proposal, and impacts directly on the estimate of the marginal likelihood for $\boldsymbol{\theta}$;

stage 2: NON ADAPTIVE PMCMC - This stage involves a burn-in chain of 5,000 iterations in which a non-adaptive proposal is utilised for the static parameters $\boldsymbol{\theta}$ based on the non-adaptive mixture component in proposal Equation (13) and the SIR particle filter for the latent process states. These burn-in samples are used to form an initial estimate of the covariance matrix for the first iteration of the adaptive MCMC proposal;

stage 3: ADAPTIVE AdPMCMC - This stage involves generating samples using the SIR particle filter proposal and the mixture adaptive Metropolis proposal Equation (13) for 50,000 samples that are subsequently used in estimating model parameters.

In addition to these burnin stages one could utilise a combination of tempering throughout the PMCMC chain. However, we note that when performing the tempering stage on the marginal likelihood, one should be careful to avoid bias in the acceptance probability. One way to overcome this problem and to still maintain all the Markov chain samples would be to perform post processing of the tempered samples, with temperature not equal to one, under the Importance Tempering framework of [Gramacy et al., 2010] within a tempered PMCMC algorithm. It should also be noted, that if performing this additional tempering, one must be very careful about the inclusion of adaptive MCMC for the static parameters and the conditions of bounded convergence and diminishing adaptation should be verified. Finally, we point out that adaption of the mutation kernel in the SMC component of the algorithm can be performed arbitrarily without any constraint on the adaption rate and one possible approach to consider involves the methodology of [Cornebise et al., 2008].

4.1. *Simulation Study - Synthetic Data M_2 - Theta Logistic.*

Here we study the performance of the AdPMCMC algorithm in the context of a challenging non-linear process model: the theta-logistic (M_2). We randomly generate data sets from the model, and we show that the AdPMCMC sampling methodology produces a Markov chain that mixes efficiently and can accurately obtain estimates of Equation (9), even in the presence of significant observation noise.

The parameters used for the synthetic studies in this first subsection are ($b_0 = 0.15, b_2 = -0.125, b_3 = 0.1, (K = 6.2, \theta = 0.1), \sigma_w = 0.39, \sigma_\epsilon = 0.47, n_0 = 1.27$). We produce $T = 50$ observations using the simulated values of the latent process. We begin the analysis of the AdPMCMC algorithm for a range of particle numbers $L \in \{20, 50, 100, 200, 500, 1000, 5000\}$. Figure 1 shows the average acceptance probability of the AdPMCMC algorithm, which proposes updates of the entire Markov chain state ($N_{1:T}, \boldsymbol{\theta}$) in 56 dimensions, at each iteration of the AdPMCMC sampler, as a function of L . As expected, increasing the number of particles results in an improved estimate of the optimal proposal distribution $\hat{p}(n_{1:T}|y_{1:T}, \boldsymbol{\theta}')$ in Equation (10) and a lower variance estimate for marginal likelihood $\hat{p}(y_{1:T}|\boldsymbol{\theta}')$ in Equation (11), ultimately improving the acceptance probability. The results demonstrate that it is reasonable to perform simulations with $L = 500$ particles which produces efficient mixing in the AdPMCMC sampler.

The bottom panel of Figure 2 presents the autocorrelation function (ACF) for σ_w^2 for different numbers of

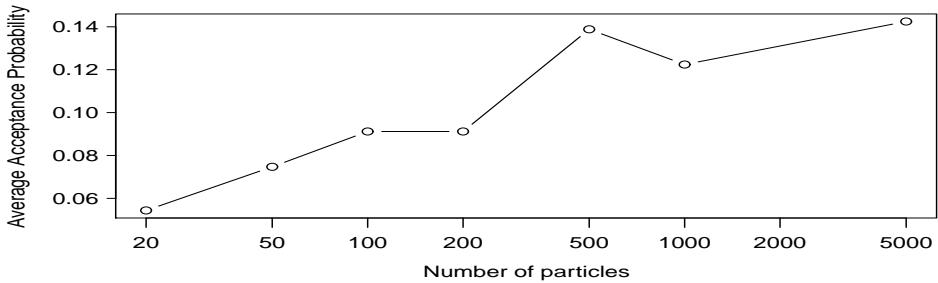


Fig. 1. Average acceptance probabilities of the AdPMCMC algorithm proposing updates of the MCMC in 56 dimensional space at each iteration, averaged over 50,000 Markov chain samples post burn-in, as a function of the number of particles L .

particles L . The top panel plots the Geweke Z-score time series diagnostic [Geweke et al., 1992] for the σ_ϵ^2 parameter as a function of L . The settings for this diagnostic considered the difference between the means of the first 10% and the last 50%. This is calculated for the Markov chain at lengths of $t \in \{5k, 10k, \dots, 50k\}$. We note that technically this diagnostic is derived for a non-adaptive Markov chain proposal. We apply this result here as a guide to convergence arguing it still produces informative results once the rate of adaption slows down, because the covariance structure used in the proposal becomes close to constant. This occurs once the chain has mixed sufficiently over the posterior support.

Figure 3 shows the trace plots of the Markov chain sample paths for the static parameters $\sigma_\epsilon^2, \sigma_w^2, N_0, b_0, b_2, b_3$. The first 5,000 samples demonstrate that the annealing stage can handle initializations of the Markov chain far from the true parameter values used to generate the data. The second stage, from samples 5,000 to 10,000, demonstrates the slow mixing performance of the untuned non-adaptive chain. Finally the samples from 10,000 to 60,000 clearly show the significant improvement of including an adaption stage in the AdPMCMC algorithm.

The top panel of Figure 4 presents the true generated latent process $N_{1:T}^{TRUE}$ and the observations. The bottom panel shows a comparison of the $N_{1:T}^{TRUE}$ versus boxplots of each state element N_i obtained

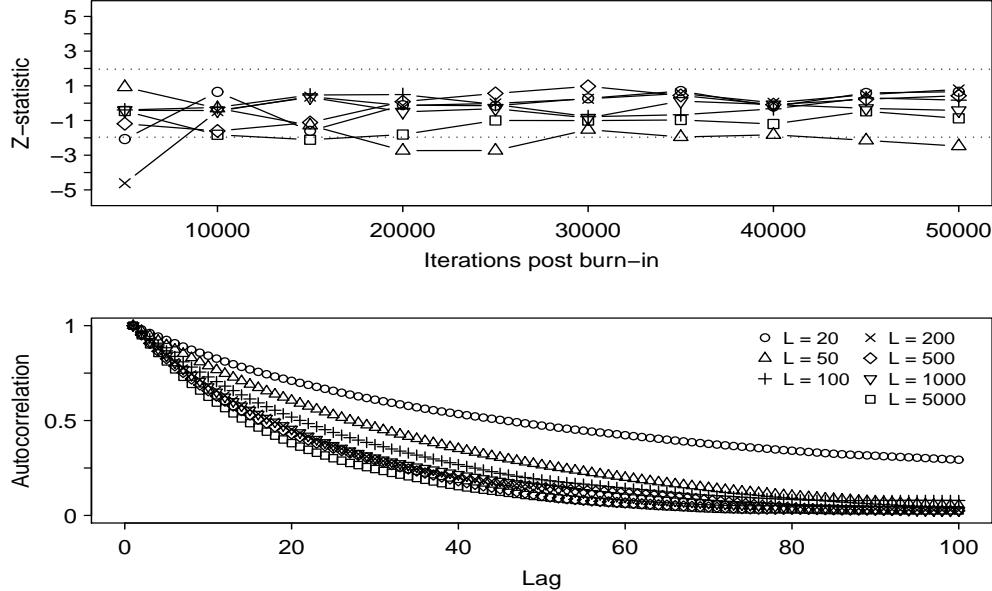


Fig. 2. Diagnostics for the AdPMCMC algorithm calculated post burn-in over 50,000 Markov chain samples. TOP SUBPLOT: Geweke Z-score statistics for noise σ_w versus L . BOTTOM SUBPLOT: ACF function of the noise σ_w versus the number of particles L .

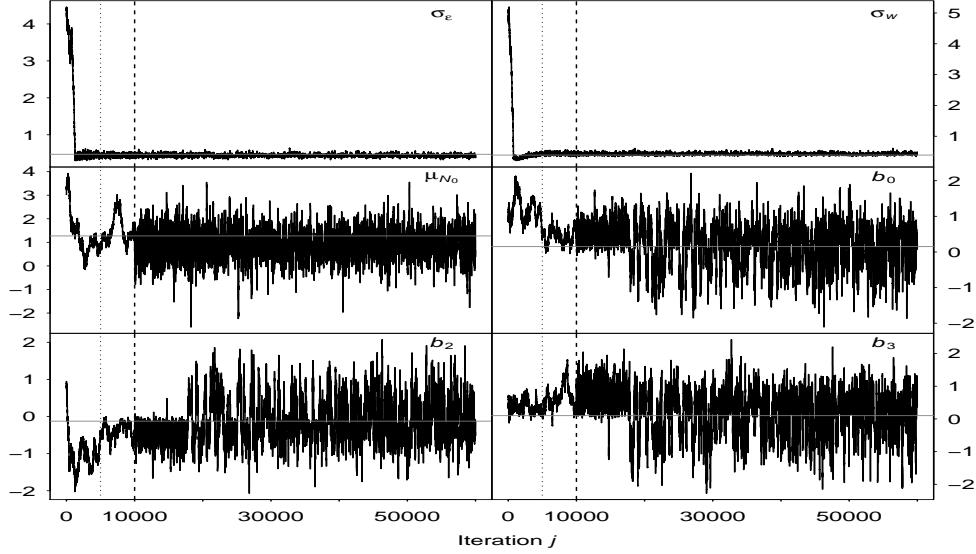


Fig. 3. Trace plots of the sample paths for the marginal Markov chain parameters $\sigma_e^2, \sigma_w^2, N_0, b_0, b_2, b_3$ based on $L = 5000$ with STAGE 1: the first 5,000 samples from the annealed stage (to the left of the vertical dotted line); STAGE 2: the samples from 5,001 to 10,000 involve the non-adaptive PMCMC stage (between the dotted and dashed lines) and; STAGE 3: the samples from 10,001 to 60,000.

using 50,000 post burn-in samples from the AdPMCMC algorithm. We also show the estimated path space MMSE, $N_{1:T}^{MMSE}$ and the 95% posterior predictive intervals shaded around this MMSE estimate. This plot demonstrates that the MMSE estimate is accurate.

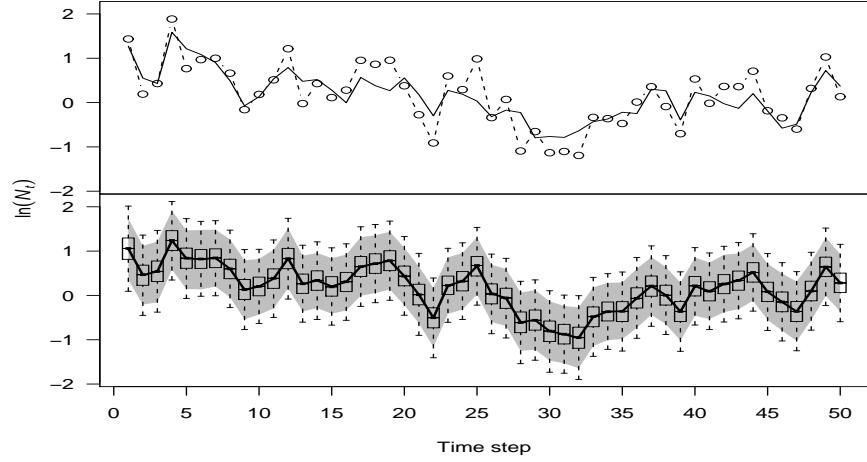


Fig. 4. TOP PANEL: The true latent process (solid line) and the observations (dashed line). BOTTOM PANEL: Box plots for the path space estimates $\{[N_{1:T}](j)\}_{j=1}^J$ obtained from the PMCMC algorithm. In addition, we present the estimated $N_{1:T}^{MMSE}$ as a solid line and the grey colouring presents the 95% posterior predictive intervals for each N_t .

Figure 5 demonstrates, for $L = 5,000$ particles, scatter plots of the pairwise marginal distributions for each static parameter (lower triangular region of the matrix plot), the kernel density estimate of the marginal posterior distribution for each parameter (diagonal of the matrix plot) and the estimated posterior correlation coefficient between the posterior parameters (upper triangular matrix plot).

Figure 6 shows, for $L = 5,000$ particles, a heat map for the estimated posterior correlation matrix obtained using samples from the AdPMCMC for the path space and static parameters, $p(\boldsymbol{\theta}, n_{1:T}|y_{1:T})$.

4.2. Synthetic Data - Mean Square Error (MSE) Analysis

In this section we study the MSE estimates of the latent process $N_{1:T}$, after integrating out the uncertainty in the static parameters $\boldsymbol{\theta}$, for a range of SMC particle counts L . This will illustrate the accuracy of the AdPMCMC estimates of the true underlying process for a given signal to noise ratio. We also derive a recursive expression for the Bayesian Cramér-Rao Lower Bound (BCRLB) for each model M_0 to M_4 as a lower bound comparison. We demonstrate how, for a set of D data sets each of length T , the BCRLB can be trivially estimated at no additional computational cost in our model framework, recursively for each time step t , via the AdPMCMC algorithm and a modified recursion from [Tichavský et al., 1998].

Cramér-Rao Lower Bound for the Path Space Proposal in PMCMC

To estimate the Cramér-Rao Lower Bound in the Bayesian context we do not require that the estimator of interest, in our case $N_{1:T}^{MMSE}$, be unbiased. However we do require that the model is specified such that the following two conditions hold.

Condition 1: for each of the static model parameters $\theta^{(i)} \in [a^{(i)}, b^{(i)}]$, the prior model $p(\theta^{(i)})$ satisfies that $\lim_{\theta^{(i)} \rightarrow a^{(i)}} p(\theta^{(i)}) \rightarrow 0$ and $\lim_{\theta^{(i)} \rightarrow b^{(i)}} p(\theta^{(i)}) \rightarrow 0$

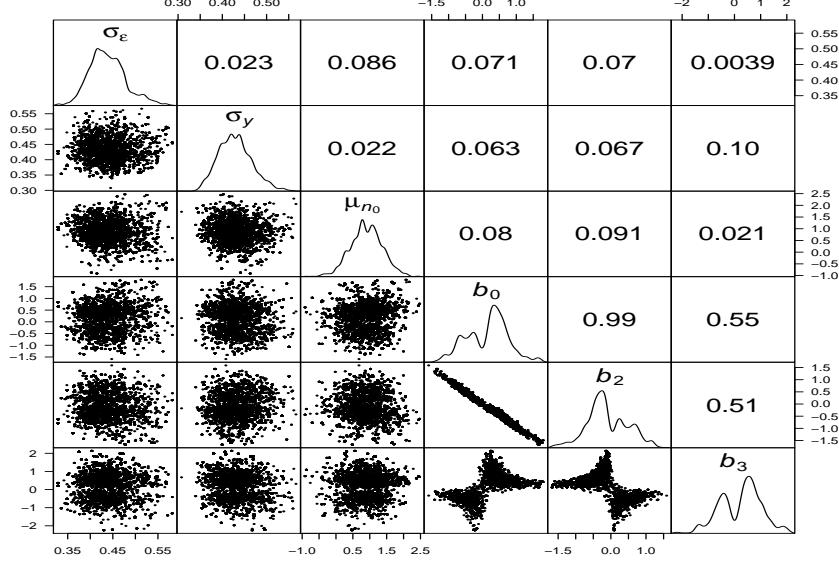


Fig. 5. Scatter plots of the posterior distributions for the static parameters θ , this plot also contains the smoothed estimated marginal posterior distributions for each static parameter, followed by the estimated linear posterior correlation coefficient between each static parameter.

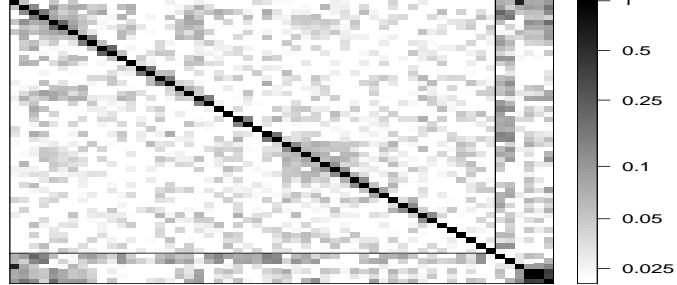


Fig. 6. Heat map in grey scale for the posterior correlation matrix between static parameters and the path space, $N_{1:T}$, given in the first 50 rows and columns, followed by the state parameters in the last six.

Condition 2: The following smoothness properties of the likelihood hold $\theta^{(i)}$,

$$\int \frac{\partial f_{\theta^{(i)}}(y_{1:T})}{\partial \theta^{(i)}} dy_{1:T} = 0$$

Under these conditions we may use the results of [Tichavský et al., 1998] in which recursive expressions for the BCRLB are derived for general non-linear state space models. We modify these results to integrate out the posterior uncertainty in the joint estimates of the static parameters θ . We derive results that perform

this marginalization numerically utilizing the existing AdPMCMC framework for each data set.

In particular we estimate for models M_0 through to M_4 the following mean square error (mse), recursively in time t ,

$$\begin{aligned} & \int \cdots \int \left\{ \left[N_{1:T} - \widehat{N}_{1:T} \right] \left[N_{1:T} - \widehat{N}_{1:T} \right]^T \right\} p(n_{1:T}, y_{1:T}, \boldsymbol{\theta}) dn_{1:T} dy_{1:T} d\boldsymbol{\theta} \\ &= \int \cdots \int \mathbb{E}_{p(n_{1:T}, y_{1:T} | \boldsymbol{\theta})} \left\{ \left[N_{1:T} - \widehat{N}_{1:T} \right] \left[N_{1:T} - \widehat{N}_{1:T} \right]^T \right\} p(\boldsymbol{\theta}) d\boldsymbol{\theta}, \end{aligned} \quad (14)$$

where in this paper we focus on the MMSE estimator $\widehat{N}_t = \widehat{N}_t^{MMSE}$. The BCRLB provides a lower bound on the MSE used to estimate the path space parameters which correspond in our model to the estimation of the latent process states $N_{1:T}$. We denote the Fisher Information Matrix (FIM), used in the BCRLB, on the path space by $[J_{1:T}(N_{1:T})](j)$ and marginally by $[J_t(N_t)](j)$ for time t , conditional on the proposed static parameters at iteration j of the PMCMC algorithm. Here we derive an analytic recursive expression for this quantity. For M_0 we can get an analytic solution whereas for the other models we will resort to AdPMCMC based online approximations with a novel estimation method based on the particle filter proposal distribution of our AdPMCMC algorithm for each data set.

Conditional on the previous Markov chain state $[\boldsymbol{\theta}, n_{1:T}] (j-1)$ and the new sampled Markov chain proposal for the static parameters at iteration j , $[\boldsymbol{\theta}] (j)$, we obtain the following modified recursive expression for the FIM based on Eq. (21) in [Tichavský et al., 1998]:

$$[J_t(\widehat{N}_t)](j) = [D_{t-1}^{22}(\widehat{N}_t)](j) - [D_{t-1}^{21}(\widehat{N}_t)](j) \left([J_{t-1}(\widehat{N}_t)](j) + [D_{t-1}^{11}(\widehat{N}_t)](j) \right)^{-1} [D_{t-1}^{12}(\widehat{N}_t)](j), \quad (15)$$

where we obtain the following matrix decompositions of our system model, via Eqs. (34-36) of [Tichavský et al., 1998] under the model assumptions of additive Gaussian process and observation noise (note derivatives here are taken under the log transformed models for N_t as specified in the additive Gaussian error SSMS in Section 2.2):

$$\begin{aligned} [J_0(\widehat{N}_t)](j) &= -\mathbb{E} \left[\nabla_{\log N_0} \{ \nabla_{\log N_0} \log p(N_0) \}^T \right]; \\ [D_{t-1}^{11}](j) &= -\mathbb{E} \left[\nabla_{\log N_{t-1}} \{ \nabla_{\log N_{t-1}} \log p(N_t | N_{t-1}) \}^T \right] = \mathbb{E} \left\{ [\nabla_{\log N_{t-1}} f(N_{t-1}; \boldsymbol{\theta})] Q_{t-1}^{-1} [\nabla_{\log N_{t-1}} f(N_{t-1}; \boldsymbol{\theta})]^T \right\}; \\ [D_{t-1}^{12}](j) &= [D_{t-1}^{21}](j) = -\mathbb{E} \left[\nabla_{\log N_t} \{ \nabla_{\log N_{t-1}} \log p(N_t | N_{t-1}) \}^T \right] = -\mathbb{E} \left[\nabla_{\log N_{t-1}} f(N_{t-1}; \boldsymbol{\theta}) \right] Q_{t-1}^{-1}; \\ [D_{t-1}^{22}](j) &= -\mathbb{E} \left[\nabla_{\log N_t} \{ \nabla_{\log N_t} \log p(N_t | N_{t-1}) \}^T \right] + -\mathbb{E} \left[\nabla_{\log N_t} \{ \nabla_{\log N_t} \log p(y_t | N_t) \}^T \right] \\ &= Q_{t-1}^{-1} + \mathbb{E} \left\{ [\nabla_{\log N_t} h(N_t; \boldsymbol{\theta})] R_t^{-1} [\nabla_{\log N_t} h(N_t; \boldsymbol{\theta})]^T \right\} \end{aligned}$$

where $f(N_{t-1}; \boldsymbol{\theta})$ is the state model with process noise covariance Q_t and $h(N_t; \boldsymbol{\theta})$ is the observation model with observation noise covariance R_t . Next we derive these quantities for each model, summarised in Table 2 where the expectations terms $[J_0(\widehat{N}_0)](j) = \left[\frac{1}{\sigma_e^2} \right](j)$ and $[D_{t-1}^{22}](j) = \left[\frac{1}{\sigma_e^2} + \frac{1}{\sigma_w^2} \right](j)$ are common to all models and evaluate as shown for iteration j of the AdPMCMC algorithm.

Remark 1: *The key point about utilising this recursive evaluation for the FIM matrix is that in the majority of cases one can not evaluate the required expectations in Table 2 analytically. However, since we are constructing a particle filter proposal distribution for the AdPMCMC algorithm to target the filtering distribution $p(n_t | y_{1:T}, [\boldsymbol{\theta}](j))$ we can use this particle estimate to evaluate the expectations at each iteration*

Table 2. Derivation of the BCRLB for each model.

Model	$[D_{t-1}^{11}] (j)$	$[D_{t-1}^{12}] (j)$
M_0	$\frac{1}{\sigma_\epsilon^2}$	$-\frac{1}{\sigma_\epsilon^2}$
M_1	$\mathbb{E} \left[(1 + \exp(N_{t-1}))^2 \frac{1}{\sigma_\epsilon^2} \right]$	$-\mathbb{E} \left[(1 + \exp(N_{t-1})) \frac{1}{\sigma_\epsilon^2} \right]$
M_2	$\mathbb{E} \left[(1 + b_2 b_3 \exp(b_3 N_{t-1}))^2 \frac{1}{\sigma_\epsilon^2} \right]$	$-\mathbb{E} \left[(1 + b_2 b_3 \exp(b_3 N_{t-1})) \frac{1}{\sigma_\epsilon^2} \right]$
M_3	$\mathbb{E} \left[\left(2 - \frac{\exp(N_{t-1})}{(b_4 + \exp(N_{t-1}))} + b_1 \exp(N_{t-1}) \right)^2 \frac{1}{\sigma_\epsilon^2} \right]$	$-\mathbb{E} \left[\left(2 - \frac{\exp(N_{t-1})}{(b_4 + \exp(N_{t-1}))} + b_1 \exp(N_{t-1}) \right) \frac{1}{\sigma_\epsilon^2} \right]$
M_4	$\mathbb{E} \left[(1 + b_6 \exp(N_{t-1}) + 2b_7 \exp(2N_{t-1}))^2 \frac{1}{\sigma_\epsilon^2} \right]$	$-\mathbb{E} \left[(1 + b_6 \exp(N_{t-1}) + 2b_7 \exp(2N_{t-1})) \frac{1}{\sigma_\epsilon^2} \right]$

t. It is important to note that this recursion avoids calculating the expectations using the entire empirical estimate of the path space, and only requires the marginal filter density estimates for each data set, which will not suffer from degeneracy as a path space empirical estimate would. Hence, for example we approximate the expectation at each time recursion t with $\mathbb{E} \left[(1 + \exp(N_{t-1}))^2 \frac{1}{\sigma_\epsilon^2} \right] \approx \sum_{i=1}^L \left[W_t (1 + \exp(N_{t-1}))^2 \frac{1}{\sigma_\epsilon^2} \right] (j, i)$ using the current Markov chain realisation of the parameters $[\boldsymbol{\theta}](j)$ and the particle estimate $[N_{t-1}](j, i)$ for the i -th particle at iteration j of the AdPMCMC algorithm, for each data set.

Remark 2: We can estimate accurately the BCRLB at each stage of the filter whilst simultaneously integrating out the static parameters $\boldsymbol{\theta}$. In addition we note that for the Exponential model (M_0), the BCRLB is analytic and optimal in this recursion and is given by the classic information filter, see [Harvey, 1991]. That is, in the AdPMCMC algorithm for M_0 , the particle filtering proposal can be replaced via Rao-Blackwellization with the Kalman filter to obtain the marginal likelihood evaluations [Carter and Kohn, 1994] for this special case of the linear Gaussian system.

Table 3 presents simulation results for models M_0 through to M_4 . In particular using the estimates obtained by the AdPMCMC algorithm for the Bayesian MMSE estimates in Equation (9) for the latent process $N_{1:T}$ we estimate the following mean square error (mse) quantity for each model

$$\begin{aligned} & \frac{1}{T} \int \sum_{t=1}^T \mathbb{E}_{p(n_{1:T}, y_{1:T} | \boldsymbol{\theta})} \left\{ \left[\widehat{N}_t^{MMSE} - n_t^{TRUE} \right] \left[\widehat{N}_t^{MMSE} - n_t^{TRUE} \right]^T | \boldsymbol{\theta} \right\} p(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ & \approx \frac{1}{JT} \sum_{j=1}^J \sum_{t=1}^T \mathbb{E} \left\{ \left[\left[\widehat{N}_t^{MMSE} \right] (j) - n_t^{TRUE} \right] \left[\left[\widehat{N}_t^{MMSE} \right] (j) - n_t^{TRUE} \right]^T | [\boldsymbol{\theta}](j) \right\} \\ & = \frac{1}{JT} \sum_{j=1}^J \sum_{t=1}^T \mathbb{E} \left\{ \left[\frac{1}{L} \sum_{i=1}^L [W_t N_t](j, i) - n_t^{TRUE} \right] \left[\frac{1}{L} \sum_{i=1}^L [W_t N_t](j, i) - n_t^{TRUE} \right]^T | [\boldsymbol{\theta}](j) \right\} \end{aligned} \quad (16)$$

where $\boldsymbol{\theta}$ denotes the static parameters. We then obtain an average of this estimate over $D = 20$ independent data realizations and report this as the average MSE for the latent state process estimate for each model and in (brackets) the standard deviation of the MSE estimate is reported. We then compare these average MSE estimates to the BCRLB estimates derived above to provide a performance comparison of our methodology as a function of the number of particles, L .

The results are presented here for each model with $L \in \{20, 100, 500\}$ particles, $T = 50$ and 50,000 iterations post burn-in ($J = 50,000$). We report the results for 20 blocks per AdPMCMC chain in calculation

Table 3. Average RMSE for $N_{1:T}^{MMSE}$ for 20 independent data realizations ($T = 50$).

Average Estimated Root Mean Square Error (20 blocks, J=50,000)					
L - number of particles	M_0	M_1	M_2	M_3	M_4
20	0.31 (0.03)	0.32 (0.04)	0.31 (0.03)	0.33 (0.03)	0.32 (0.03)
100	0.31 (0.04)	0.31 (0.03)	0.30 (0.03)	0.31 (0.04)	0.32 (0.04)
500	0.31 (0.03)	0.32 (0.04)	0.31 (0.04)	0.31 (0.04)	0.30 (0.04)
Average Estimated BCRLB					
500	0.34 (0.01)	0.38 (0.01)	0.34 (0.01)	0.33 (0.01)	0.35(0.02)

Table 4. Average RMSE for $N_{1:T}^{MMSE}$ for 20 independent data realizations ($T = 50$).

Average Estimated Root Mean Square Error (20 blocks, J=50,000, L = 500)					
σ_w^2	M_0	M_1	M_2	M_3	M_4
2×	0.40 (0.05)	0.39 (0.05)	0.39 (0.03)	0.39 (0.05)	0.39 (0.06)
4×	0.50 (0.07)	0.50 (0.05)	0.51 (0.05)	0.49 (0.07)	0.51 (0.07)
Average Estimated BCRLB					
2×	0.38 (0.02)	0.42 (0.02)	0.36 (0.02)	0.35 (0.02)	0.38 (0.02)
4×	0.43 (0.03)	0.47 (0.03)	0.41 (0.03)	0.38 (0.02)	0.43 (0.03)

of the estimated Root Mean Square Error for each chain. However, we also explored varying the number of blocks between (10, 20, 40) in the RMSE calculations which we found to have only a minor effect on the average MSE estimate.

We present two tables of results, the first in Table 3 utilises data generated from model settings which reflect typical observation and process noise variance levels and the second in Table 4 presents results in which the noise variance is increased by a factor of two and four. Each data set is generated from the same model parameters in each of the models M_0 to M_4 for comparison purposes.

The results demonstrate the accuracy of our methodology since the MMSE estimates obtained from the simulations are not (statistically) significantly different from the estimated average BCRLB. This provides confidence in the AdPMCMC methodology. In addition we see that increasing the noise variance of the observations, results in a relatively larger increase in the estimated RMSE compared to the estimated BCRLB. In other words as the signal-to-noise ratio decreases, the estimator accuracy also decreases, therefore producing a larger difference between the average BCRLB and the estimated average RMSE. We see this effect in Table 4, though it is not overly pronounced. The results in most models at high noise settings no longer achieve the BCRLB on average, though they are still reasonably close.

4.3. Bayes Factors Model Selection Assessment.

The ideal case for choosing between posterior models occurs when integration of the posterior (with respect to the parameters) is possible. For two models M_i and M_j , parameterized by $\theta_{1:k}$ and $\alpha_{1:j}$ respectively, the

posterior odds ratio of M_i over M_j is given by

$$\frac{p(y_{1:T}|M_i)p(M_i)}{p(y_{1:T}|M_j)p(M_j)} = \frac{p(M_i) \int p(y_{1:T}|\theta_{1:k}, M_i)p(\theta_{1:k}|M_i)d\theta_{1:k}}{p(M_j) \int p(y_{1:T}|\alpha_{1:j}, M_j)p(\alpha_{1:j}|M_j)d\alpha_{1:j}} = \frac{p(M_i)}{p(M_j)} BF_{ij}$$

which is the ratio of posterior model probabilities, having observed the data, where $p(M_i)$ is the prior probability of model M_i and BF_{ij} is the Bayes factor. This quantity is the Bayesian version of a likelihood ratio test, and a value greater than one indicates that model M_i is more likely than model M_j (given the observed data, prior beliefs and choice of the two models).

To demonstrate the performance of the AdPMCMC in this context we generated 10 data sets using the flexible-Allee model, M_4 each with the same model parameters. Then we estimated each of the models M_0 through to M_4 using this data, and calculated estimates of the Bayes factors. A representative data realization for the parameters of M_4 chosen in this simulation is presented in Figure 7. In this example we

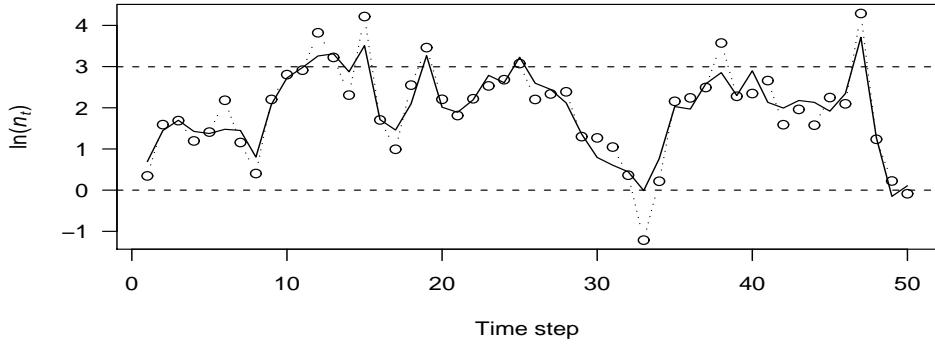


Fig. 7. A representative data realization for the BF analysis generated from model M_4 the Flex-Ricker model. The upper dashed line is the carrying capacity (K) and the lower dashed line is the Allee threshold (C).

considered $N_0 = \ln(2)$, $b_5 = -0.05$, $b_6 = 0.0525$, $b_7 = -0.0025$ with process and observation noise variances drawn randomly from the priors as detailed in Section 3.1. These model parameters were selected to generate population trajectories that fluctuate between the population carrying capacity K and the Allee threshold C . This makes for a challenging model selection scenario and should result in ambiguity over the actual true model, as all models we consider can potentially capture this form of population growth behaviour in the presence of the high process and observation noise used here. The results in Table 5 confirm that for any given realization of the data from M_4 with these model parameters, we see switching between the most plausible model to explain the particular realization. We see from the Bayes Factors that there is a strong ambiguity between the true model, the flexible Allee M_4 used to generate the data and the Ricker model M_1 . The reason for this is that the process and observation noise obscures the Allee effect but the signal of negative density dependence remains, and this is most parsimoniously represented by the Ricker model M_1 .

This synthetic example highlights the challenges faced in choosing between models, when realistic levels of observation and latent process noise are present in the population counts data. It also emphasises

Table 5. Bayes Factors (rounded to integers) with true model as M_4 and (T = 50, L=500, J=50k).

Data Set	BF_{01}	BF_{02}	BF_{03}	BF_{04}	BF_{12}	BF_{13}	BF_{14}	BF_{23}	BF_{24}	BF_{34}
1	0	1	3	0	32	185	6	6	0	0
2	0	0	0	0	411	24	3	0	0	0
3	40	31	11873	0	1	296	0	379	0	0
4	0	0	0	0	18986	62	5	0	0	0
5	0	0	0	0	2	66	3	27	1	0
6	0	0	0	0	8	15	5	2	1	0
7	0	0	0	0	65903	44	5	0	0	0
8	0	0	0	0	27	97	137	4	5	1
9	0	0	0	0	697	98	4	0	0	0
10	0	0	0	0	75	5229	3	70	0	0

Table 6. Bayes Factors (rounded to integers) with true model as M_4 and (T = 50, L = 500, J = 50k). Observation noise variance and process noise variance both decreased by two orders of magnitude.

Data Set	BF_{01}	BF_{02}	BF_{03}	BF_{04}	BF_{12}	BF_{13}	BF_{14}	BF_{23}	BF_{24}	BF_{34}
1	0	2	0	0	68	13	0	0	0	0
2	0	0	0	0	41	1	0	0	0	0
3	0	3	0	0	2207	2	1	0	0	0
4	0	3	0	0	275	4	0	0	0	0
5	0	0	0	0	83	2	0	0	0	0
6	0	2	0	0	65	0	0	0	0	0
7	0	2	0	0	33	1	0	0	0	0
8	0	2	0	0	144	1	0	0	0	0
9	0	7	0	0	39	1	1	0	0	1
10	0	1	0	0	116	2	0	0	0	0

the importance of efficient posterior MCMC samplers for each model to ensure that the Bayes Factors are accurate. These results also highlight that observation noise and process noise modelling can be critical when determining the presence of density dependent mortality or Allee effects in real data sets.

When the standard deviations of both the process and observation noise were decreased by an order of magnitude, then the model selection exercise produces Bayes Factors that identify the correct model, (Table 6). In this reduced noise case, the Bayes Factors for all models (other than M_4) versus model M_4 were less than one. Table 7 explores the effect of observation and process noise on the model selection under different signal to noise ratios. We focus on the Bayes factors for model M_1 versus M_4 , as these two models are most likely to be ambiguous in this case as they are each capable of capturing an Allee effect to varying degrees.

The results in Table 7 demonstrate several important points relating to model selection in the presence of differing severities of process and observation noise. In low process error settings we see that in the majority

Table 7. Bayes Factor BF_{14} (rounded to integers) with true model as M_4 and ($T = 50$, $L=500$, $J=50k$).

Noise Levels	Data Set									
	1	2	3	4	5	6	7	8	9	10
$\sigma_{\epsilon,w}^2 \sim IG(T/2, (T-2)/10)$	6	3	0	5	3	5	5	137	4	3
$\sigma_\epsilon^2/10$	0	7	0	0	10	10	0	1	3	0
$\sigma_w^2/10$	6	2	0	1	121	0	2	7	1	0
$\sigma_{\epsilon,w}^2/10$	0	2	0	0	0	0	0	1	0	0
$\sigma_{\epsilonpsilon}/100$	0	1	0	1	4	0	0	0	1	0
$\sigma_w^2/100$	10	1	0	39	1	0	7	2	0	2
$\sigma_{\epsilon,w}^2/100$	0	0	1	0	0	0	0	0	1	0

of cases, the correct model has strong evidence for its selection. Secondly, in the case in which the observation noise is significantly reduced, we see a reduction in the evidence for the incorrect model M_1 . However, this is not as marked as when the process error is decreased, indicating that perhaps the process error may have a stronger influence on the ability to distinguish the two closely related models. Finally, we confirm that for large reductions in both process and observation noise, there is clear selection of the appropriate model M_4 .

4.4. Real Data Analysis

In the following section we present analysis of two real data sets. For each data set, we extended the length of the AdPMCMC chains so that the mixing results in a Geweke times series diagnostic approximate Z-score in the interval [-2,2] for all parameters of all models. This ensures accurate estimation of the Bayes factors.

We study the nutria data from [Gosling, 1981]. Nutria are a widespread invasive species with populations now established around the world [Woods et al., 1992]. This data is available as data set 9833 in the Global Population Database [NERC Centre for Population Biology, Imperial College, 1999], and presents a time series of female nutria abundance in East Anglia at monthly intervals, obtained by retrospective census for a feral population.

This data set is of interest because [Drake, 2005] fit models $M_0 \dots M_3$ without observation error, and found that the AIC selected the strong Allee model M_3 as the best overall model whereas BIC selected the Ricker model M_1 as the best overall model. Here we revisit this data set, include observation error, add an additional model (M_4) that allows for both strong and weak Allee effects, and use Bayes factors for model selection.

In Figure 8 we present the two most plausible models for the nutria data, according to the Bayes factor analysis in Table 8 with the following settings: $T = 120$, $L=500$, $J = 350,000$ post burn-in, (burn-in = 155,000 and annealing = 5,000). In this analysis, the default scalings proposed by [Roberts and Rosenthal, 2009] for the non-adaptive component in Algorithm 2 proved suboptimal for models $M_{1,3,4}$. We subsequently scaled down the non-adaptive components in these models by factors of 10^{-3} to 10^{-7} and made a similar adjustment in the annealing phase. The MMSE estimates for the path space $N_{1:T}^{MMSE}$, thinned by 10%,

are shown for models M_4 and M_2 , together with the unthinned Markov chain trace plots of the noise and process model variances. The solid lines are the MMSE of the path space and the points are observations with the gray shading corresponding to the posterior 95% credibility intervals.

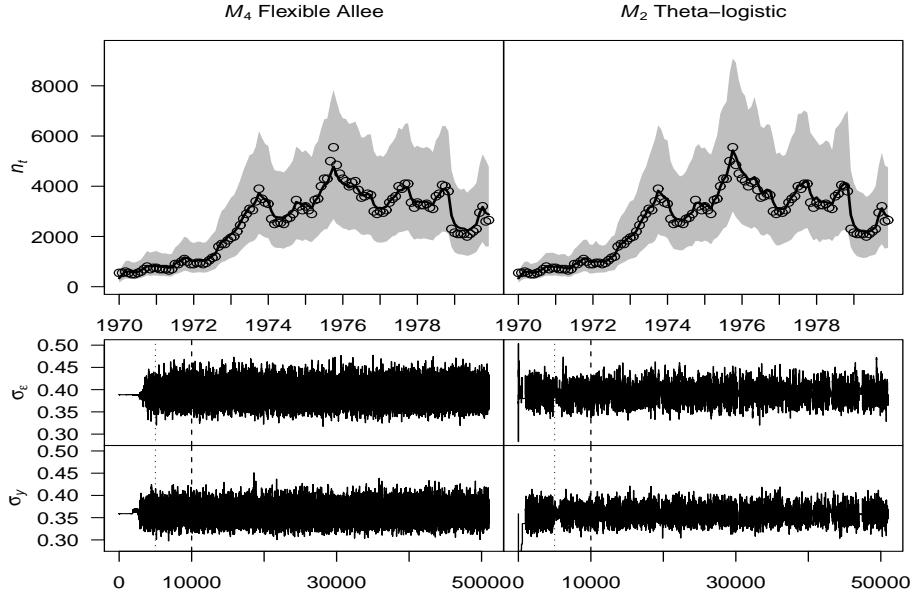


Fig. 8. Nutria data analysis. Left plots - model M_4 and Right plots - model M_2 . Top panel - estimates of the MMSE for the path space (solid lines), observations from data (points) and posterior 95% predictive interval (shaded). Lower panels - trace plots of noise standard deviations for observation and state process.

Our findings demonstrate posterior evidence for the selection of model M_4 , over all other models, in the presence of jointly estimated observation and state noise. With the selection of model M_4 , it is interesting to comment on the MMSE estimate for the model parameters from this fit. There is a small probability (< 0.01) that the only stable equilibrium is zero; that is, there is no stable positive equilibrium K and the population will go extinct. Conditional on the existence of a positive per capita population growth rate, that is positive for some population density, the estimate for the Allee threshold C in model M_4 is -2484 (mean) [-38814, 796] (0.95 CI) relative to a carrying capacity $K = 3403$ [2369, 4437]. This suggests that a weak Allee effect is present in the nutria population.

If we ignore the new model M_4 , the Bayes factors (Table 8) identify model M_1 , which has no Allee effect, as the best fitting model, over model M_3 , which only has a strong Allee effect. In contrast, the results of Drake [2005] are ambiguous, where AIC results suggest that M_3 is the best model but BIC suggests M_1 is the best. Drake [2005] notes that the differences in both AIC and BIC results are minor and do not provide compelling support for one model over the other. These results show the utility of considering models (such as M_4) that allow for a weak Allee effect.

The second real data set we analyse is a time series of a population of sparrowhawks (*Accipiter nisus*) in south Scotland. This is a well studied population [Newton and Rothery, 1997] that shows evidence of density dependence [Newton and Marquiss, 1986]. This data set (number 6575 in the Global Population Database

[NERC Centre for Population Biology, Imperial College, 1999]) is of interest because [Polansky et al., 2009] showed that the likelihood surface of the theta-logistic model applied to this data, with both process and observation noise, has strong ridges and multiple modes. [Polansky et al., 2009] calculate the maximum likelihood estimates of the theta-logistic's parameters and warn that its complex likelihood surface may derail methods of model comparison. They calculated that the parameter controlling the form of density dependence (equivalent to parameter b_3 in M_2) equals -4.83 at the global maximum of the likelihood surface. A second local maximum was found when $b_3 = 0.04$.

The Bayes factors (Table 9) with algorithm settings ($T = 18$, $L = 500$, $J = 150,000$ post burn-in, burn-in = 50,000 and annealing = 5,000), however, suggest that the Ricker model M_1 is the best fitting model. This is equivalent to the theta-logistic M_2 with $b_3 = 1$. In our analysis the theta-logistic model is the worst of the five models. Note also that the AdPMCMC sampler mixes suitably well over the complicated multimodal posterior support (Figure 9). Here we have again used the Geweke diagnostic to determine the number of post burn-in samples, but we recognise that we may require a larger number of samples in order to accurately estimate the weighting or proportion of posterior mass in each of the two modes. The fact remains that the sampler is able to mix over the two modes, enabling model selection via Bayes factors even in the presence of ridges and multiple modes in the likelihood surface, assuming that the sampler mixes sufficiently well over the posterior support.

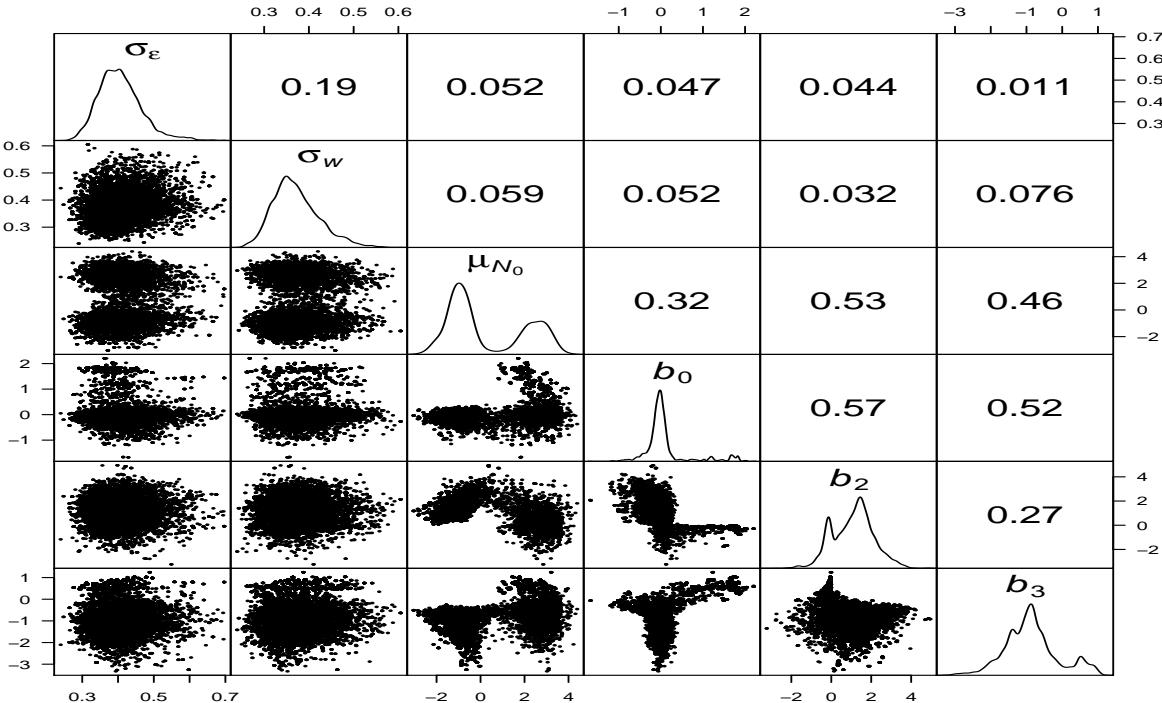


Fig. 9. Scatter plots of the posterior distributions for the static parameters θ for *A. nisus* data set under model M_2 . This plot also contains the smoothed estimated marginal posterior distributions for each static parameter, followed by the estimated linear posterior correlation between each static parameter.

Table 8. Bayes Factors (rounded) for nutria.

Model	M_0	M_1	M_2	M_3	M_4
$BF_0.$	1	0	7	0	0
$BF_1.$	6432	1	44366	14	0
$BF_2.$	0	0	1	0	0
$BF_3.$	450	0	3104	1	0
$BF_4.$	32229	5	222302	72	1

Table 9. Bayes Factors (rounded) for A .*nibus.*

Model	M_0	M_1	M_2	M_3	M_4
$BF_0.$	1	0	5	0	0
$BF_1.$	22	1	118	10	6
$BF_2.$	0	0	1	0	0
$BF_3.$	2	0	12	1	1
$BF_4.$	4	0	21	2	1

5. Discussion

This paper presents a novel technical development: a sophisticated sampling methodology based on Adaptive MCMC [Roberts and Rosenthal, 2009] and Particle MCMC [Andrieu et al., 2010]; that enables a novel application: an efficient and robust approach to jointly estimate observation and process noise with the latent states and static parameters in Bayesian nonlinear state space models of population dynamics. We developed a novel adaptive version of the PMCMC algorithm and demonstrated its performance on several challenging high dimensional posterior models. These models were based on population dynamic models that are cited widely in the literature. We have also placed the problem of statistical inference and model selection in a more realistic setting by accounting for, and estimating, process error and observation noise jointly with the model parameters and latent population process, without the need for time-consuming tuning and algorithm design. We believe that the novel sampling framework presented here can be easily generalised to different ecological state space models.

The AdPMCMC sampling methodology we developed is more sophisticated than the standard MCMC algorithms currently employed in ecological settings [Millar and Meyer, 2000, Clark and Bjornstad, 2004, Ward, 2006] in which the equations describing the latent population dynamics are often highly non-linear. In addition, there is typically a strong correlation between the model parameters and the latent process. These two factors will make the mixing properties of basic MCMC or block-Gibbs sampling algorithms, highly inefficient in a time series setting, see [Andrieu et al., 2010] and [Nevat et al., 2010] for discussion.

To understand this, consider the simple sampling framework in which the posterior distribution for model M_i , denoted $p(\boldsymbol{\theta}, n_{1:T} | y_{1:T}, M_i)$ is sampled in the following block Metropolis-Hastings within Gibbs

framework, where the vector of latent states is split into k block of length $k\tau = T$,

$$\begin{aligned} \text{iteration j: } \boldsymbol{\Theta} &\sim p(\boldsymbol{\theta}|n_{1:T}, y_{1:T}, M_i) \\ \text{iteration j: } N_{1:\tau} &\sim p(n_{1:\tau}|\boldsymbol{\theta}, n_{\tau+1:T}, y_{1:T}, M_i) \\ &\vdots \\ \text{iteration j: } N_{(k-1)\tau+1:T} &\sim p(n_{(k-1)\tau+1:T}|\boldsymbol{\theta}, n_{1:(k-1)\tau}, y_{1:T}, M_i). \end{aligned}$$

Such a block design is highly inefficient if the parameters of the posterior distribution are correlated, as in the case of the *A. nisu*s data set (Figure 9). For moderate sized values of τ this sampling framework will mix poorly because the Metropolis-Hastings acceptance probabilities will be low irrespective of the proposals that are used. The simplest solution to this problem would be to sample directly from the full conditional distributions in a univariate Gibbs framework. However, even in the case where a single component Gibbs sampler is possible via inversion sampling from the full conditional cdf's for all parameters, the Markov chain will mix very slowly around the support of the posterior. This is especially problematic in high dimensional target posterior distributions, since it requires excessively long Markov chains to achieve samples from the stationary regime. It can also lead to very high autocorrelations in the Markov chain states, with a concomitant impact on the variance of the estimators in Equation 9. To avoid slow mixing Markov chains, one must sample from larger blocks of parameters, - i.e. larger τ . However, the design of an optimal proposal distribution for large blocks of parameters is very complicated. The Particle MCMC methodology steps around this problem by approximating the optimal proposal distribution for a large number of parameters via a Sequential Monte Carlo proposal distribution.

Another positive property of the AdPMCMC methodology is evident in the Markov chain sample paths, shown in Figure 3. As the AdPMCMC sampler mixes over different modes , the estimated adapted covariance in the static parameter proposal “adapts” or “learns” on-line to account for the additional mode and allows for more efficient mixing between the two modes of the posterior. This demonstrates the power of the adaptive MCMC methodology when combined within the PMCMC sampler.

Previous meta-analyses of ecological time-series have used model selection techniques to conclude that Allee effects are rare without explicitly accounting for the confounding effects of observation error [Saether et al., 1996, Sibly et al., 2005, Gregory et al., 2010]. Here we used a synthetic dataset to show that moderate process and observation error will lead to ambiguous model selection results in ecological time series that include strong Allee effects. This occurs because the noise masks the full signal of the deterministic density-dependent processes. Our results suggest, however, that negative density dependence is much easier to detect than the Allee effect. Indeed, a parsimonious model that includes negative dependence but omits the Allee effect will often be selected in the presence of observation error. The chance of correctly choosing a model with an Allee effect can be increased by decreasing the observation error.

Nevertheless, we did find evidence for a weak Allee effect in the nutria data set studied by [Drake, 2005], who reported the AIC and BIC model selection in the absence of observation noise. We included observation noise and extended the class of models to include M_4 , which allows for both strong and weak Allee effects,

and we found posterior evidence for model M_4 over all other models. Furthermore, we found that the second most likely model to explain the data was M_1 , which lacks an Allee effect, over model M_3 , which only has a strong Allee effect. This agrees with the BIC results in the analysis without observation noise of [Drake, 2005].

In the *A. nisu* data set, [Polansky et al., 2009] use a numerical approach inspired by [Kitigawa, 1987] to find multiple modes with local optima that suggests a nonlinear relationship between population density and per capita growth rate. The AdPMCMC sampling methodology demonstrates efficient mixing between these multiple modes. [Kitigawa, 1987] acknowledges that a grid-based approach will not scale efficiently with increasing dimensions in the latent path space or the static parameters; for instance, multispecies analyses will be precluded. On the other hand, our AdPMCMC approach will scale up to much larger dimensions. If an optimization algorithm is used instead, as in [Polansky et al., 2009], then it becomes difficult to calculate the uncertainty associated with the maximum likelihood estimates. In particular, it is difficult to determine the joint uncertainty of the static parameters and the latent path space estimates. In contrast, the AdPMCMC approach directly approximates these joint densities.

It is important to note that model parameterisation has important implications for the choice of static parameter priors in the population models used here. In this paper we have followed a common practice in the ecological literature, e.g., [Clark, 2007], wherein parameters such as the intrinsic rate of growth r and carrying capacity K are combined into new parameters b_i such that the model on the log scale is linear in all (e.g., $M_{0,1,4}$) or some (e.g., $M_{2,3}$) of the parameters. The b_i for these models are often nonlinear functions of ecological parameters. We have observed that independent Gaussian priors on the b_i induce strong prior dependence among the ecological parameters. The AdPMCMC approach does not require linearity in the parameters and can fit the underlying ecological parameters directly and efficiently. In other words our methodology is general and can be applied to any parameterization either linear or non-linear, under any chosen prior structure with or without dependence.

On the other hand, the more general b_i formulation allows researchers to question how much support the data gives to parameter spaces consistent with the traditional ecological formulations. For example, Figure 9 shows that it is unlikely for both $b_0, b_2 > 0$. This corresponds to unchecked population growth in the latent model; this unlikely but not impossible result is precluded in the traditional formulation of the theta-logistic in terms of r, K, θ [Morris and Doak, 2002]. The more general approach is useful because there are many ways to parameterize a population model's structure, and different ecological considerations may lead to quite different constraints. For instance, previous studies have proposed the following constraints in the theta-logistic model M_2 : $b_3 > 0$ [Ross, 2006, Ward, 2006], $b_3 > -1$ [Sæther et al., 2008, Ross, 2009], $\text{sgn } b_3 = \text{sgn } b_0$ [Polansky et al., 2009]. Figure 9 shows that these constraints may be unrealistic or unduly restrictive in the presence of both process error and observation noise. We argue that, wherever possible, it is better to estimate the probability of constraints rather than impose them, and the more general formulation provides a way to achieve this. Finally, we note that hyperpriors on the prior parameter values can also be considered in the AdPMCMC framework to address the usual concerns about prior sensitivity.

In our opinion the most important results of this study: a) are the confounding effects of observation and process errors on model selection; and, b) the ability of the AdPMCMC to mix efficiently over the complex ridges and multiple modes of the likelihood surfaces associated with at least some population dynamic models. In particular, it is clear that the development of adaption in the MCMC proposal used for the static parameters clearly improves mixing. Through the use of Bayes factors we have emphasised how important it is to consider a range of alternative models when seeking to understand the nonlinear and density dependent effects that drive population dynamics. It is important to recognise, however, that our limited understanding of these effects, reflected in the process error of a population model, coupled with our imperfect measuring devices, reflected in the variance of a observation model, may hamper our ability to distinguish between different plausible models. It is important that ecologists recognise this, particularly if a “best fitting” model to make predictions of population trajectories beyond the observations.

The AdPMCMC algorithm presented here is capable of efficient SSM inference in non-linear population dynamic models with complex likelihood surfaces. It thereby frees practitioners from the potentially slowly mixing constraints of the SSM algorithms, particularly Metropolis Hastings within Gibbs, that are currently available. The AdPMCMC strategy can be readily generalised to other equivalent problems, and moreover, readily extends to include more complex SMC methods that also incorporate adaption and/or more realistic error variance structures. For these reasons, we believe the algorithm holds great promise in applied contexts.

6. Acknowledgments

We thank Prof. Arnaud Doucet for comments and suggestions during his research visit to CSIRO in Tasmania. The contributions of GRH and KRH were funded by the CMIS Capability Development Fund (CDF). GWP gratefully acknowledges the support of CDF during this research collaboration.

References

- C. Andrieu and YF Atchade. On the efficiency of adaptive MCMC algorithms. page 43, 2006.
- C. Andrieu and É. Moulines. On the ergodicity properties of some adaptive MCMC algorithms. *Annals of Applied Probability*, 16(3):1462–1505, 2006.
- C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *J. R. Statist. Soc. B*, 72(2):1–33, 2010.
- Y.F. Atchadé and J.S. Rosenthal. On adaptive markov chain monte carlo algorithms. *Bernoulli*, 11(5):815–828, 2005.
- C.M. Bishop et al. *Pattern recognition and machine learning*. Springer New York:, 2006.
- D.S. Boukal and L. Berec. Single-species models of the Allee effect: Extinction boundaries, sex ratios and mate encounters. *Journal of Theoretical Biology*, 218:375–394, 2002.

- C. Calder, M. Lavine, P. Muller, and J. S. Clark. Incorporating multiple sources of stochasticity into dynamic population models. *Ecology*, 84:1395–1402, 2003.
- B.P. Carlin and T.A. Louis. Bayes and empirical Bayes methods for data analysis. *Statistics and Computing*, 7(2):153–154, 1997.
- R.J. Carroll, D. Ruppert, and L.A. Stefanski. *Measurement error in nonlinear models*. Chapman and Hall, 1995.
- CK Carter and R. Kohn. On Gibbs sampling for state space models. *Biometrika*, 81(3):541, 1994.
- J. S. Clark. *Models for Ecological Data: An Introduction*. Princeton University Press, Cambridge, England, 2007.
- J. S. Clark and O. N. Bjornstad. Population inference from messy data: errors, missing and hidden states, and lagged responses. *Ecology*, 85:3140–3150, 2004.
- J. Cornebise, E. Moulines, and J. Olsson. Adaptive methods for sequential importance sampling with application to state space models. *Statistics and Computing*, 18(4):461–480, 2008.
- Dennis. Allee effects in stochastic populations. *Oikos*, 96:389–401, 2002.
- A. Doucet and A.M. Johansen. A tutorial on particle filtering and smoothing: Fifteen years later. *The Oxford Handbook of Nonlinear Filtering, Oxford University Press. To appear*, 2009.
- J. M. Drake. Risk analysis for invasive species and emerging infectious diseases: Concepts and applications. *American Midland Naturalist*, 153:4–19, 2005.
- R. P. Freckleton, A. R. Watkinson, R. E. Green, and W. J. Sutherland. Census error and the detection of density dependence. *Journal of Animal Ecology*, 75:837–851, 2006.
- J. Geweke, JO Berger, and AP Dawid. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *In Bayesian Statistics 4*, 1992.
- N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. 140(2):107–113, 1993.
- LM Gosling. The dynamics and control of a feral coypu population. pages 1553–2056, 1981.
- R. Gramacy, R. Samworth, and R. King. Importance tempering. *Statistics and Computing*, 20(1):1–7, 2010.
- S. D. Gregory, C. J. A. Bradshaw, B. W. Brook, and F. Courchamp. Limited evidence for the demographic allee effect from numerous species across taxa. *Ecology*, In press, 2010.
- H. Haario, E. Saksman, and J. Tamminen. Componentwise adaptation for high dimensional MCMC. *Computational Statistics*, 20(2):265–273, 2005.

- A.C. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge Univ Pr, 1991.
- G. Kitigawa. Non-Gaussian state-space modelling of nonstationary time series. *Journal of the American Statistical Association*, 82:1032–1041, 1987.
- R. Lande, S. Engen, and B.-E. Sther. *Stochastic population dynamics in ecology and conservation*. Oxford University Press, Oxford, 2003.
- M. A. Lewis and P. Kareiva. Allee dynamics and the spread of invading organisms. *Theoretical Population Biology*, 43:141–158, 1993.
- A. Linden and J. Knape. Estimating environmental effects on population dynamics: consequences of observation error. *Oikos*, 118:675–680, 2009.
- S.T. Lindley. Estimation of population growth and extinction parameters from noisy data. *Ecology Applications*, 13:806–813, 2003.
- R. B. Millar and R. Meyer. Non-linear state space mode of fisheries biomass dynamics by using metropolis-hastings within gibbs sampling. *Applied Statistics*, 49:327–342, 2000.
- W. F. Morris and D. F. Doak. *Quantitative Conservation Biology*. Sinauer Associates, Inc., 2002.
- NERC Centre for Population Biology, Imperial College. The Global Population Dynamics Database. <http://www.sw.ic.ac.uk/cpb/cpb/gpdd.html>. 1999. URL <http://www.sw.ic.ac.uk/cpb/cpb/gpdd.html>.
- I. Nevat, G.W. Peters, A. Doucet, and Y. Jinhong. Channel tracking for relay networks via adaptive particle MCMC. *preprint*, 2010.
- I. Newton and M. Marquiss. Population regulation in sparrowhawks. *Journal of Animal Ecology*, 55:463–480, 1986.
- I. Newton and P. Rothery. Senescence and reproductive value in sparrowhawks. *Ecology*, 78:1000–1008, 1997.
- G.W. Peters, Y. Fan, and S.A. Sisson. Likelihood-free Bayesian inference for α -stable models. 2009.
- L. Polansky, P. de Valpine, J. O. Lloyd-Smith, and W. M. Getz. Likelihood ridges and multimodality in population growth models. *Ecology*, 90:2313–2320, 2009.
- C.P. Robert. *The bayesian choice*. Springer New York, 2001.
- G.O. Roberts and J.S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4):351–367, 2001.
- G.O. Roberts and J.S. Rosenthal. Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367, 2009.

- J.V. Ross. Comment on the regulation of populations of mammals, birds, fish, and insects II. *Science*, 311: 1100, 2006.
- J.V. Ross. A note on density dependence in population models. *Ecological Modelling*, 220:3472–3474, 2009.
- B.-E. Saether, T. H. Ringsby, and E. Roskaft. Life history variation, population processes and priorities in species conservation: Towards a reunion of research paradigms. *Oikos*, 77:217–226, 1996.
- B.-E. Sæther, M. Lillegård, V. Grøtan, F. Filli, and S. Engen. Predicting fluctuations of reintroduced ibex populations: the importance of density dependence, environmental stochasticity and uncertain population estimates. *Journal of Animal Ecology*, 76:326–336, 2007.
- B.-E. Sæther, S. Engen, V. Grøtan, T. Bregnballe, C. Both, P. Trjanowski, A. Leivits, J. Wright, A. P. Møller, M. E. Visser, and W. Winkel. Forms of density regulation and (quasi-) stationary distributions of population sizes in birds. *Oikos*, 117:1197–1208, 2008.
- T. M. Shenk, G. C. White, and K. P. Burnham. Sampling-variance effects on detecting density dependence from temporal trends in natural populations. *Ecological Monographs*, 68:445–463, 1998.
- R. M. Sibly, D. Barker, M. C. Denham, J. Hone, and M. Pagel. On the regulation of populations of mammals, birds, fish and insects. *Science*, 309:607–610, 2005.
- N. C. Stenseth, H. Viljugrein, T. Saitoh, T. F. Hansen, M. O. Kittilsen, E. Bølviken, and F. Glöckner. Seasonality, density dependence, and population cycles in Hokkaido voles. *Proceeding of the National Academy of Sciences USA*, 100:11478–11483, 2003.
- P. Tichavský, C.H. Muravchik, and A. Nehorai. Posterior Cramer-Rao bounds for discrete-time nonlinear filtering. *IEEE Transactions on Signal Processing*, 46(5):1386–1396, 1998.
- G. Wang. On the latent state estimation of nonlinear population dynamics using Bayesian and non-Bayesian state-space models. *Ecological Modelling*, 200(3-4):521–528, 2007.
- E. J. Ward. A new BEAST: Bayesian software tools for ecological trend analysis. *Wildlife Society Bulletin*, 34:1420–1424, 2006.
- G. Welch and G. Bishop. An introduction to the Kalman filter. *University of North Carolina at Chapel Hill, Chapel Hill, NC*, 1995.
- C. K. Williams, A. R. Ives, and R. D. Applegate. Population dynamics across geographical ranges: Time-series analyses of three small game species. *Ecology*, 84:2654–2667, 2003.
- C. A. Woods, L. Contreras, G. Willner-Chapman, and H. P. Whidden. *Myocastor coypus*. *Mammalian Species*, 398:1–8, 1992.
- Z. Zeng, R.M. Nowierski, M.L. Taper, B. Dennis, and W.P. Kemp. Complex population dynamics in the real world: modeling the influence of time-varying parameters and time lags. *Ecology*, 79:2193–2209, 1998.

Algorithm 1 Construction of optimal path space proposal for Model M_i given by $\hat{p}(n_{1:T}|y_{1:T}, [\boldsymbol{\theta}](j+1))$

Initialisation of SIR filter at iteration $(j+1)$ of the Markov chain

- 1: SIR particle filter for $N_{1:T}$: initialise L particles $\{[n_1](j+1, i)\}_{i=1:L}$ via sampling from the priors in Section 3.1

2: **for** $t = 2, \dots, T$ **do**

Perform mutation of the L particles at time $t-1$ to obtain new particles at t via state evolution.

- 3: Sample the i -th particle $[n_t](j+1, i)$ from particle filter proposal according to state equation given in relevant model M_i Section 2.2. Example M_3 sample

$$N_t \sim N(2 \log [N_{t-1}](j, i) - \log ([b_4](j+1, i) + [N_{t-1}](j, i)) + [b_0](j+1, i)) + N([b_1](j+1, i)[N_{t-1}](j, i), [\sigma_e^2](j+1, i)) \quad (17)$$

Incremental SIR importance sampling weight correction.

- 4: Evaluate the unnormalised importance sampling weights, $[\tilde{W}_t](j+1, i)$, for the L particles, with the i -th weight given by

$$\begin{aligned} [\tilde{W}_t](j+1, i) &\propto [W_{t-1}](j+1, i) [w_{t-1}](j+1, i) \\ &\propto [W_{t-1}](j+1, i) p(y_t | [n_t, \boldsymbol{\theta}](j+1, i)), \end{aligned} \quad (18)$$

- 5: Normalise the importance sampling weights $[W_t](j+1, i) = \frac{[\tilde{W}_t](j+1, i)}{\sum_{i=1}^L [\tilde{W}_t](j+1, i)}$

Evaluate the importance estimate and resample adaptively.

- 6: Calculate the Effective Sample size, $ESS = \frac{1}{\sum_{i=1}^L [W_t](j+1, i)^2}$
 - 7: If the Effective Sample size is less than 80% then resample the particles at time t using stratified resampling based on the empirical distribution constructed from the importance weights to obtain new particles with equal weight.
 - 8: **end for**
 - 9: Evaluate marginal likelihood $\hat{p}(y_{1:T} | [\boldsymbol{\theta}](j+1)) = \prod_{t=1}^T \left(\frac{1}{L} \sum_{i=1}^L [w_t](j+1, i) \right)$
-

Algorithm 2 Adaptive MCMC for static parameters Θ

1: Sample a realisation u_1 of random variable $U_1 \sim U[0, 1]$

Sample from the adaptive mixture proposal (13)

2: **if** $u_1 \geq w_1$ **then**

Sample $[\boldsymbol{\theta}](j+1)$ from the adaptive component of the mixture proposal of (13)

3: Estimate Σ_j , the empirical covariance of Θ , using samples $\{[\boldsymbol{\theta}](i)\}_{i=1:j}$.

4: Sample proposal $[\boldsymbol{\theta}](j+1) \sim N\left(\boldsymbol{\theta}; [\boldsymbol{\theta}](j), \frac{(2.38)^2}{d}\Sigma_j\right)$;

5: **else**

Sample $[\boldsymbol{\theta}](j+1)$ from the non-adaptive component of the mixture proposal of (13)

6: Sample proposal $[\boldsymbol{\theta}](j+1) \sim N\left(\boldsymbol{\theta}; [\boldsymbol{\theta}](j), \frac{(0.1)^2}{d}I_{d,d}\right)$

7: **end if**

This text file "unconstrainedInitT.txt.gz" is available in "gzipped" format from:

<http://arxiv.org/ps/1005.2238v1>