# State-space Models for Ecological Time Series Data: Practical Model-fitting

Ken Newman[1,2], Ruth King[1], Víctor Elvira[1], Perry de Valpine[3], Rachel S. McCrea[4], and Byron J. T. Morgan[4]

[1]School of Mathematics, University of Edinburgh, Edinburgh, UK.

[2]Biomathematics and Statistics Scotland, Edinburgh, UK.

[3]Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA, USA.

[4]School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, UK.

October 6, 2025

**Abstract**

1. State-space models are an increasingly common and important tool in the quantitative ecologists' armoury, particularly for the analysis of time series data. This is due to both their flexibility and intuitive structure, describing the different individual processes of a complex system, thus simplifying the model specification step.
2. State-space models are composed of two processes (i) the system (or state) process that describes the dynamics of the true underlying state of the system over time; and (ii) the observation process that links the observed data with the current true state of the system at that time. Specification of the general model structure consists of considering each distinct ecological process within the system and observation processes, which are then automatically combined within the state-space structure.
3. There is typically a trade-off between the complexity of the model and the associated model-fitting process. Simpler model specifications permit the application of simpler model-fitting tools; whereas more complex model specifications, with non-linear dynamics and/or non-Gaussian stochasticity often require more sophisticated model-fitting algorithms to be applied.
4. We provide a brief overview of general state-space models before focusing on the different model-fitting tools available. In particular for different general state-space model structures we discuss established model-fitting tools that are available. We also offer practical guidance for choosing a specific fitting procedure.

**Keywords**: hidden Markov model; Kalman filter; Laplace approximation; likelihood-free methods; Markov chain Monte Carlo; Sampling-based methods; Sequential Monte Carlo.

# 1 Introduction

State-space models have become increasingly popular in the modeling of ecological time-series data due to their flexible and intuitive structure that includes multiple component processes. In particular, state-space models are composed from individual biological processes that drive the system's underlying temporal dynamics (i.e., the *state process*) and the observation processes that link the observed data to this state (or system) process (i.e., the *observation process*). The state process is generally constructed from biological knowledge that describes the dynamics of the ecological system over time as a function of directly interpretable model parameters. However, the underlying states are generally unobserved, and all inference on the model parameters comes from statistically disentangling state dynamics from the observed data. Ecological applications of state-space models include for example, the analysis of telemetry data (Hooten et al., 2017; Patterson et al., 2017), fisheries stock assessment (Millar and Meyer, 2000; de Valpine and Hilborn, 2005; Aeberhard et al., 2018), population dynamics (Newman et al., 2014; Buckland et al., 2004), capture-recapture-type data (King, 2014) and biodiversity (Kindsvater et al., 2018).

For further in-depth discussion of ecological state-space models, and additional examples, see Auger-Méthé et al. (2021) and Newman et al. (2009). The special case of discrete-valued states may lead to hidden Markov models (HMMs), which McClintock et al. (2021) recently reviewed with an emphasis on ecological applications. In this review we focus on the situation where the underlying states are continuous-valued.

In practice there is a trade-off between the complexity of the state-space model specification and the associated computational model-fitting challenges. Simpler biological models that reduce the system complexity and/or use Gaussian distributional assumptions that lead to an analytically tractable likelihood function are relatively easy to fit. Conversely, more biologically realistic models (e.g., with non-linear dynamics and/or non-Gaussian distributional assumptions) lead to a more complex and analytically intractable likelihood expression that require more sophisticated model-fitting algorithms, often with trade-offs between accuracy and efficiency. In this paper we provide a focused discussion of the model-fitting challenges of modern general ecological state-space models. In particular, we describe the

different model-fitting approaches and associated algorithms that can be applied to state-space models dependent on the properties of the specified model. In practice, there is no dominant algorithm that is universally the "best" for all ecological state-space models; instead the performance of different model-fitting tools depends on the particular form, and observed data, of the state-space model (Fasiolo et al., 2016).

## 2 State-space models

State-space models (SSMs) are a convenient and useful representation of the processes that generate time-series data as they separate these into (i) a *state model* that describes the dynamics of the true underlying (unobserved or partially observed) state of the system over time; and (ii) an *observation model* that describes how observed data are stochastic functions of the true underlying state. Different authors define SSMs more narrowly or broadly. Here we focus on SSMs for discrete-time, continuous-valued, first-order (Markov) states, a special case of hidden process models (Newman et al., 2006, 2014).

We start with a general form for SSMs. Let the observed (possibly multivariate) time-series data over the set of observation times $t = 1, \ldots, T$ be denoted by $\boldsymbol{y}_{1:T} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T\}$, where $\boldsymbol{y}_t$ is an $(K \times 1)$ *observation* vector. The data, $\boldsymbol{y}_{1:T}$, depend on latent (or unobservable) states, $\boldsymbol{\alpha}_{1:T} = \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_T\}$, where $\boldsymbol{\alpha}_t$ is an $(n \times 1)$ *state* vector. While the elements of $\boldsymbol{\alpha}_t$ are typically unknown, or partially observed, they are governed by a stochastic model that describes the ecological dynamics of the system and induce dependence (e.g., temporal correlation) among the data. The SSM is described by:

$$\text{Observation model} : f(\boldsymbol{y}_t|\boldsymbol{\alpha}_t, \boldsymbol{\theta}), \qquad t = 1, \ldots, T; \tag{1}$$

$$\text{State model} : g(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}, \boldsymbol{\theta}), \quad t = 1, \ldots, T, \tag{2}$$

where $f$ and $g$ are probability density functions, which in full generality could be time-varying. The observation and state models are parameterised by the parameters of both the ecological dynamics and the observation process, $\boldsymbol{\theta}$. An important feature that directly influences model-fitting algorithms relates to the conditional independence assumptions. First,

the latent states, $\boldsymbol{\alpha}_t$, are assumed to be first-order Markov; second, the observation at time $t$, $\boldsymbol{y}_t$, given $\boldsymbol{\alpha}_t$, is independent of previous observations and states. See Durbin and Koopman (2012) and Sarkka (2013) for further discussion of general SSMs.

The model specification is completed by an initial state distribution. This may be expressed for states at time 0 or time 1, $g(\boldsymbol{\alpha}_0|\boldsymbol{\theta})$ or $g(\boldsymbol{\alpha}_1|\boldsymbol{\theta})$. Typically, different rationales are used to choose the form of distribution. For example, it is often assumed to be simple and uninformative, with large variance(s); or if the state dynamics are ecologically stable, the equilibrium distribution of the system process may be used (de Valpine and Hastings, 2002). For structured population models, the stable age- or stage-distribution may be used with unknown population size (Besbeas and Morgan, 2012).

The associated likelihood for $\boldsymbol{\theta}$ results by expressing the joint density of latent states and observations, and then integrating out (marginalising over) the states. For the SSM above, due to the two independence assumptions, the likelihood can be expressed in sequential form:

$$\begin{aligned} L(\boldsymbol{\theta}|\boldsymbol{y}_{1:T}) &= \int p(\boldsymbol{y}_{1:T}, \boldsymbol{\alpha}_{0:T}|\boldsymbol{\theta}) d\boldsymbol{\alpha}_{0:T} \\ &= \int g(\boldsymbol{\alpha}_0|\boldsymbol{\theta}) \times \left[ \prod_{t=1}^{T} f(\boldsymbol{y}_t|\boldsymbol{\alpha}_t, \boldsymbol{\theta}) g(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}, \boldsymbol{\theta}) \right] d\boldsymbol{\alpha}_{0:T}. \end{aligned} \tag{3}$$

The term $p(\boldsymbol{y}_{1:T}, \boldsymbol{\alpha}_{0:T}|\boldsymbol{\theta})$ denotes the joint likelihood (probability distribution) of the data and latent states, given the parameters $\boldsymbol{\theta}$. The term "complete-data likelihood" is often used for $p(\boldsymbol{y}_{1:T}, \boldsymbol{\alpha}_{0:T}|\boldsymbol{\theta})$, and "missing data" for $\boldsymbol{\alpha}_{0:T}$, even when they are not missing but rather unmeasurable. In contrast, the likelihood, $L(\boldsymbol{\theta}|\boldsymbol{y}_{1:T})$, is often referred to as the "observed-data likelihood". The integral in the likelihood is analytically intractable for general SSMs, with no closed form expression. Closed form likelihood expressions do exist for two special cases: (i) linear and Gaussian (LG-)SSMs (see Section 3.1); and (ii) for a finite set of discrete-valued latent states that lead to HMMs (see Section 3.2.2).

We note that SSMs are a special case of a more general class of models called hierarchical or multi-level models. The main idea is that the observed data, which are random variables at one level, depend upon another set of random variables, generally not observed directly,

the latent variables, at a higher level. Classical random or fixed effects models are special cases of hierarchical models with two levels. Hierarchical models are often analysed in a Bayesian framework (Berliner, 1996; Cressie et al., 2009) that adds a third (or higher) level, namely prior probability distributions for the parameters of the distributions for the second level of the latent variables. Hierarchical models can characterise spatial data (Cressie, 2015) and spatio-temporal data (Wikle et al., 2019). Our focus here is narrower as we are only considering time series data within a hierarchical framework.

One of the most desirable features on hierarchical models is the divide-and-conquer approach to analysing potentially complex processes whereby attention can be focused on individual levels. With SSMs, for example, scientists and subject matter specialists can bring their expertise to bear in the formulation of the state process model while statisticians can focus on formulation of the observation process model linked to the state. In this paper we are admittedly just sketching the basic structures of state and observation models with little discussion of particular complex formulations that are possible—many such examples can be found in Auger-Méthé et al. (2021).

# 3 Model-fitting

The goals of model-fitting for SSMs vary depending on objectives and include estimating model parameters ($\boldsymbol{\theta}$), estimating latent states ($\boldsymbol{\alpha}_{0:T}$), and/or forecasting of future latent states/observations ($\boldsymbol{\alpha}_{T+k}/\boldsymbol{y}_{T+k}$). Estimates of $\boldsymbol{\theta}$ provide information about the mechanisms driving the state process dynamics; whereas estimates of $\boldsymbol{\alpha}_{0:T}$ provide information about the true state of nature.

Our focus is on estimation of $\boldsymbol{\theta}$ and we present a variety of classical and Bayesian inferential procedures. The primary inferential challenges, with the exception of LG-SSMs and HMMs, are intractable likelihoods, the calculation of which typically involves integration over $T+1$ dimensions. The likelihood, $L(\boldsymbol{\theta}|\boldsymbol{y}_{1:T})$, is central to classical and Bayesian inference. Classical inference via maximum likelihood estimation seeks $\boldsymbol{\theta}$ that maximises $L(\boldsymbol{\theta}|\boldsymbol{y}_{1:T})$.

Bayesian inference centres on the posterior distribution for $\boldsymbol{\theta}$:

$$p(\boldsymbol{\theta}|\boldsymbol{y}_{1:T}) = \frac{\pi(\boldsymbol{\theta})L(\boldsymbol{\theta}|\boldsymbol{y}_{1:T})}{m(\boldsymbol{y}_{1:T})}, \tag{4}$$

where $\pi(\boldsymbol{\theta})$ denotes the prior distribution and $m(\boldsymbol{y}_{1:T})$ the (unconditional) marginal distribution, where the latter involves additional integration over $\boldsymbol{\theta}$. For a general overview see for example Reich and Ghosh (2019); van de Schoot et al. (2021) and Appendix A for further details.

Before proceeding, we make remarks about two inferential procedures for the states: filtering and smoothing, as they are fundamental to inference about $\boldsymbol{\theta}$. Inference for $\boldsymbol{\alpha}_t$, conditional only on the observations available up to time $t$, $\boldsymbol{y}_{1:t}$, is called *filtering*, with all information contained in the filtering distribution, $p(\boldsymbol{\alpha}_t|\boldsymbol{y}_{1:t}, \boldsymbol{\theta})$. Inference for $\boldsymbol{\alpha}_t$ based on all available observations, $\boldsymbol{y}_{1:T}$, is called *smoothing*, with associated smoothing distribution, $p(\boldsymbol{\alpha}_{0:T}|\boldsymbol{y}_{1:T}, \boldsymbol{\theta})$. See Appendices D and E for further details. We now describe a range of different SSM fitting algorithms and discuss their usage.

## 3.1 Kalman filter and extensions

For ecological time-series data that are of LG-SSM form the likelihood can be easily calculated by the Kalman filter (KF; Kalman (1960); Welch and Bishop (1995)). MLEs of the parameters can be obtained via numerical optimisation, or a Bayesian posterior distribution can be explored. However, the LG assumptions are restrictive, necessitating the use of approximations or more advanced methods described later.

### 3.1.1 Kalman filter

We use univariate states and observations for simplicity. The LG-SSM assumptions hold if:

$$\begin{aligned} y_t|\alpha_t, \boldsymbol{\theta} &\sim \text{Normal}(c\alpha_t, \sigma^2); \\ \alpha_{t+1}|\alpha_t, \boldsymbol{\theta} &\sim \text{Normal}(a + b\alpha_t, \tau^2), \end{aligned}$$

where $a$, $b$, $c$, $\sigma$ and $\tau$ are parameters. For the mathematical details of the KF, including how to calculate the likelihood function, see Box 1. An example of the LG-SSM is the (stochastic) Gompertz population model where $\alpha_t$ is log population size, $y_t$ are noisy observed data of $\alpha_t$ (with $c = 1$), and $b$ incorporates density-dependence (Dennis et al., 2006).

**Box 1. LG-SSM & Kalman Filter.** The linear-Gaussian SSM (LG-SSM) is arguably the most celebrated SSM. Its success is mostly due to the tractability of the inferential methods (see Section 3). The observation model is expressed as

$$\boldsymbol{y}_t = \boldsymbol{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t, \tag{5}$$

for $t = 1, \ldots, T$, where $\boldsymbol{Z}_t$ denotes the $(K \times m)$ observation process matrix (and a function of the model parameters) and $\boldsymbol{\epsilon}_t$ the $(K \times 1)$ vector corresponding to observational noise. We typically assume that $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{H}_t)$, independently of each other, where $\mathbf{0}$ denotes the column vector with each element equal to 0 and $\boldsymbol{H}_t$ is a (potentially time varying) covariance matrix.

The process model is given by

$$\boldsymbol{\alpha}_t = \boldsymbol{T}_t \boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \tag{6}$$

for $t = 1, \ldots, T$, where $\boldsymbol{T}_t$ denotes the $(m \times m)$ transition matrix that governs the changes of the state-vector from occasion $t$ to occasion $t + 1$ (and is a function of the model parameters); and $\boldsymbol{\eta}_t$ an $(m \times 1)$ vector, such that, $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{F}_t)$, corresponding to the stochastic (e.g., environmental) variability. For many ecological applications the state vector comprises the abundances of individuals in several age (and or state) classes. In such a situation, the formation of $\boldsymbol{T}_t$ can be fairly straightforward for state vectors of small dimension, however it is sometimes of interest, and simpler, to decompose the formation of $\boldsymbol{T}_t$ into intermediate sub-processes, such as survival, ageing, reproducing, movement, etc. (Buckland et al., 2004).

**Box 1 (cont).** For the special case of the linear and Gaussian model, the log of the observed data likelihood can be calculated explicitly as follows:

$$\log p(\boldsymbol{y}_{1:T}|\boldsymbol{\theta}) = -\frac{T}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\left(\log|\boldsymbol{F}_t| + \boldsymbol{v}_t\boldsymbol{F}_t^{-1}\boldsymbol{v}_t\right),$$

where $\boldsymbol{v}_t$ and $\boldsymbol{F}_t$ are quantities calculated routinely from application of the Kalman filter, which is defined by the following set of equations:

$$\begin{aligned}
\boldsymbol{v}_t &= \boldsymbol{y}_t - \boldsymbol{Z}_t\boldsymbol{a}_t \\
\boldsymbol{F}_t &= \boldsymbol{Z}_t\boldsymbol{P}_t\boldsymbol{Z}_t^\top + \boldsymbol{H}_t \\
\boldsymbol{K}_t &= \boldsymbol{T}_t\boldsymbol{P}_t\boldsymbol{Z}_t^\top\boldsymbol{F}_t^{-1} \\
\boldsymbol{a}_{t+1} &= \boldsymbol{T}_t\boldsymbol{a}_t + \boldsymbol{K}_t\boldsymbol{v}_t \\
\boldsymbol{L}_t &= \boldsymbol{T}_t - \boldsymbol{K}_t\boldsymbol{Z}_t \\
\boldsymbol{P}_{t+1} &= \boldsymbol{T}_t\boldsymbol{P}_t\boldsymbol{L}_t^\top + \boldsymbol{Q}_t
\end{aligned}$$

where $\boldsymbol{a}_t = \mathbb{E}(\boldsymbol{\alpha}_t)$. This form of the likelihood is sometimes referred to as the prediction error decomposition likelihood owing to the interpretations of $\boldsymbol{v}_t$ and $\boldsymbol{F}_t$. Since the likelihood is available in closed form, the parameters can be estimated, classically, via maximum likelihood using a numerical optimisation algorithm or the posterior distribution of the parameters formed within a Bayesian analysis.

In order to obtain the mean of the distribution of $\boldsymbol{\alpha}_t$ conditional on the whole sample $\boldsymbol{y}_{1:T}$ fixed-interval smoothing can be implemented using the following recursive equations:

$$\begin{aligned}
\boldsymbol{a}_{t|T} &= \boldsymbol{a}_t + \boldsymbol{P}_t^*(\boldsymbol{a}_{t+1|T} - \boldsymbol{T}_{t+1}\boldsymbol{a}_t) \\
\boldsymbol{P}_{t|T} &= \boldsymbol{P}_t + \boldsymbol{P}_t^*(\boldsymbol{P}_{t+1|T} - \boldsymbol{P}_{t+1})\boldsymbol{P}_t^{*\top}
\end{aligned}$$

where $\boldsymbol{a}_{T|T} = \boldsymbol{a}_T$ and $\boldsymbol{P}_{T|T} = \boldsymbol{P}_T$ are values from the Kalman filter recursions above and $\boldsymbol{P}_t^* = \boldsymbol{P}_t\boldsymbol{T}_{t+1}^\top\mathbf{P}_{\mathbf{t+1}}^{\mathbf{-1}}$.

### 3.1.2 "Approximate linear" Kalman filter

Many models are non-linear but preserve additive Gaussian residuals. Extending the Gompertz example, different density-dependence assumptions lead to non-linear dynamics (de Valpine and Hastings, 2002; Wang, 2007). Similarly, within the observation process, specifying the additive Gaussian error to be on the (non-log) population scale leads to a non-linear component, such that $y_t|\alpha_t, \sigma^2 \sim \text{Normal}(\exp(\alpha_t), \sigma^2)$. See Knape et al. (2011) and Einarsson et al. (2016) for additional examples.

An approximation called the extended KF (EKF) applies a first-order Taylor expansion to non-linear model components within the KF algorithm (Einarsson et al., 2016; Sarkka, 2013, Chapter 5). However, its accuracy depends on the degree of non-linearity and will yield biased estimates if inaccurate. An alternative is the unscented KF (UKF), which propagates weights for a small set of state values (Wan et al., 2001). Wang (2007) found that UKF consistently outperformed EKF for population dynamics models, and moreso for stronger non-linearity, making it the preferred of these methods. The ensemble Kalman filter (EnKF), originally proposed as a computationally efficient approach for high dimensional system states, can be more generally applied to non-linear SSMs and involves simulating states to estimate covariance matrices (see Katzfuss et al. (2016) and Michaud et al. (2021) for more details). More recently the EnKF has been applied to non-Gaussian distributions and continues to perform well (Katzfuss et al., 2020; Michaud et al., 2021).

### 3.1.3 "Approximate Gaussian" Kalman filter

Some models that do not have Gaussian distributions may be approximated as Gaussian based on mean(s) and variance(s). For example, Poisson and binomial distributions can be approximated as Gaussian distributions if the number of individuals is "large" and the binomial probability is not close to 0 or 1. Besbeas et al. (2002) proposed this approach for a population dynamics model for juvenile and adult birds where the binomial assumption for survival and Poisson assumption for reproduction rendered the model non-Gaussian. Approximating these as Gaussian distributions performed well for reasonable population sizes (King et al., 2004). With the model "shoe-horned" into a LG-SSM, the KF likelihood

can be used for maximum likelihood or Bayesian methods.

## 3.2 Likelihood approximation and beyond

The above approximations may fail if models are too far from being linear and/or Gaussian. Here we discuss more general methods for handling the likelihood, approximately or exactly.

### 3.2.1 Laplace approximation

The Laplace method is a general and efficient approach for approximating some integrals (see Appendix B for mathematical details). In particular, the Laplace method uses a multivariate Gaussian approximation for the likelihood via a second-order Taylor expansion of the integrand over all time steps. Applying the Laplace approximation to SSMs, we can estimate the likelihood using,

$$L(\boldsymbol{\theta}|\boldsymbol{y}_{1:T}) \approx p(\boldsymbol{y}_{1:T}, \hat{\boldsymbol{\alpha}}_{0:T}|\boldsymbol{\theta}) \, (2\pi)^{\frac{T+1}{2}} \left|\mathbf{H}^{-1}_{\hat{\boldsymbol{\alpha}}_{0:T}}\right|^{\frac{1}{2}},$$

where, for given $\boldsymbol{\theta}$, $\hat{\boldsymbol{\alpha}}_{0:T}$ maximises the complete-data likelihood, $p(\boldsymbol{y}_{1:T}, \boldsymbol{\alpha}_{0:T}|\boldsymbol{\theta})$, and $H_{\hat{\boldsymbol{\alpha}}_{0:T}}$ denotes the Hessian (matrix of second derivatives) of $\ln p(\boldsymbol{y}_{1:T}, \boldsymbol{\alpha}_{0:T}|\boldsymbol{\theta})$ evaluated at $\hat{\boldsymbol{\alpha}}_{0:T}$.

We note that using the Laplace approximation for obtaining an estimate of the MLEs of $\boldsymbol{\theta}$ involves two stages of optimisation: the *inner optimisation* over $\hat{\boldsymbol{\alpha}}$ for given $\boldsymbol{\theta}$; and the *outer optimisation* over $\boldsymbol{\theta}$. Laplace approximation has been widely applied to SSMs for fishery stock assessment (e.g. Kristensen et al., 2016) and for marine animal movement (Auger-Méthé et al., 2017; Albertsen et al., 2015). While most applications of the Laplace approximation to SSMs have been in a classical framework, Monnahan and Kristensen (2018) recently carried out a Bayesian analysis and also commented on inaccuracies of the approximation for a specific case study.

Laplace approximation can also serve as the first step for more accurate methods. Importance sampling uses simulations from some distribution that can roughly approximate $p(\boldsymbol{\alpha}_{0:T}|\boldsymbol{y}_{1:T}, \boldsymbol{\theta}) \propto p(\boldsymbol{y}_{1:T}, \boldsymbol{\alpha}_{0:T}|\boldsymbol{\theta})$. Weighted calculations can then approximate the likelihood. The optimum $\hat{\boldsymbol{\alpha}}_{0:T}$ and corresponding Hessian provide a natural choice of Gaussian

approximating distribution (Skaug, 2002). A further extension is the Laplace importance Gauss-Hermite algorithm (Elvira et al., 2020). In both cases, asymptotically exact inference can be achieved with large simulation sizes. However all of these methods can struggle to work well for highly nonlinear dynamics (e.g., when the state distribution is bimodal) or for long time series.

### 3.2.2 Discrete approximation

If the dimensionality of states, $\boldsymbol{\alpha}_t$, is low (i.e. at most 2 or 3), very good approximations to the sequential form of the likelihood, given in Equation (3), can be obtained by discretising (making a grid for) values of $\boldsymbol{\alpha}_t$ (de Valpine and Hastings, 2002; Langrock and King, 2013; Besbeas and Morgan, 2020). A first-order approximation uses the probability of transitioning from each cell at time $t$ to each other cell at $t+1$, for example using a mid-point rule. This results in a HMM, reducing the integral to a set of closed form matrix operations; see Box 2. Higher accuracy results can be obtained using, for example, the "trapezoid method" of assuming line segments between discretised states, and/or using the fast Fourier transform to propagate additive Gaussian noise efficiently (de Valpine and Hastings, 2002). In general, there is a trade-off between accuracy and computational efficiency. Smaller grid cells will give more accurate results, but the method suffers from a "curse of dimensionality": if $\boldsymbol{\alpha}_t$ has too many dimensions, taking the Cartesian product of these states to define the overall combined state may lead to a computationally infeasible number of potential state transitions. In principle it is straightforward to repeat the analyses with decreasing grid sizes, adaptive grid cells and/or increasing complex estimates of the transition probabilities to investigate the robustness of the approach and the associated computational time trade-offs (Besbeas and Morgan (2019, 2020); Borowska and King, unpublished data). For low dimensions the approximation can be made arbitrarily accurate and is thus an attractive (and in our view under-used) model-fitting approach. The approach has been used within numerous ecological applications including accounting for observation error over space for telemetry models (Pedersen et al., 2011), population dynamics models (Besbeas and Morgan, 2019, 2020) and capture-recapture-type models with continuous individual time-varying covariates (Langrock and King, 2013).

**Box 2. HMM & Likelihood Inference.** If the values taken by the underlying system states belong to a finite set of discrete values, denoted $\{1, \ldots, M\}$, then the corresponding SSM reduces to a hidden Markov model (HMM) with an explicit likelihood expression. However we note that this distinction of terminology of SSM referring to continuous-valued states and HMM for discrete-valued states is not universal. The HMM has analogous model components as for the SSM, but now defined for discrete system states. Note that for notational simplicity we assume that the latent states are univariate (though if this is not the case the set of possible states can be specified in such a manner by taking the Cartesian product over possible state combinations). The HMM is then defined via the following components:

1. The initial state distribution: $\boldsymbol{\delta}$ of dimension $(1 \times M)$ such that $\delta_j = \mathbb{P}(\alpha_1 = j)$ (assuming that the initial state distribution is specified at time 1)
   [equivalent to $g(\alpha_1|\boldsymbol{\theta})$];

2. The state model: defined via the $(M \times M)$ transition probability matrix, $\boldsymbol{\Gamma}_{t-1} = (\gamma_{ij})_{t-1}$, such that $\gamma_{ijt-1} = \mathbb{P}(\alpha_t = j|\alpha_{t-1} = i)$ for $i, j = 1, \ldots, M$
   [equivalent to $g(\alpha_t|\alpha_{t-1}, \boldsymbol{\theta})$];

3. Observation model: define via the $(M \times M)$ observation process matrix, $\boldsymbol{P}(\boldsymbol{y}_t) = \text{diag}(f(\boldsymbol{y}_t|\alpha_t = 1, \boldsymbol{\theta}_O), \ldots, f(\boldsymbol{y}_t|\alpha_t = M, \boldsymbol{\theta}_O))$, where $\boldsymbol{\theta}_O$ denotes the parameters associated with the observation process only.
   [equivalent to $f(\boldsymbol{y}_t|\alpha_t, \boldsymbol{\theta})$].

The likelihood of the HMM is expressible in closed form, replacing the integral in Equation (3) by a finite sum, so that,

$$L(\boldsymbol{\theta}|\boldsymbol{y}_{1:T}) = \boldsymbol{\delta}\boldsymbol{P}(\boldsymbol{y}_1) \prod_{t=1}^{T-1} \boldsymbol{\Gamma}_t \boldsymbol{P}(\boldsymbol{y}_{t+1})\mathbf{1},$$

where, $\mathbf{1}$ denotes the $(M \times 1)$ row vector with each element equal to 1. For further discussion of HMMs, see for example, Zucchini et al. (2016).

For the discrete HMM approximation to the SSM the states $1, \ldots, M$ will typically correspond to a set of intervals (in the univariate case) or higher-dimensional boxes (in the multivariate case). Transitions between these states are a function of the general state model $g(\alpha_t|\alpha_{t-1}, \boldsymbol{\theta})$, approximated using the mid-point rule or higher dimensional approximations (Langrock, 2011; Langrock and King, 2013). A similar approximation is made in relation to the observation process.

### 3.2.3 Sequential Monte Carlo

Sequential Monte Carlo (SMC; Chopin and Papaspiliopoulos (2020)) methods, or particle algorithms, are essentially iterative importance sampling procedures (see Appendix C) that simulate values of the states, $\boldsymbol{\alpha}_{0:T}$, often referred to as particles, combined with a resampling procedure. The simulated states are used to construct stochastic approximations of filtering and smoothing distributions, as well as the likelihood. The terminology particle filtering (PF) and particle smoothing is often used for these approximations. SMC procedures typically include a resampling schedule, where particles with higher weights are retained and replicated, and particles with lower weights are discarded. Details on general recursive filtering and smoothing algorithms are given in Appendices D and E; and an example is described in Appendix F.

Two features that distinguish different SMC algorithms are (i) the importance sampler and (ii) the resampling schedule. These affect both the variance and bias of the approximated distributions/likelihood as well as computational expense. The primary issue with SMC methods is *particle depletion*, which arises when there are very few non-negligible weights associated with the simulated states. The weights are functions of the observation distribution and simulated state values. If the simulated states are "at odds" with the observed data, the weights will be small or negligible. An importance sampler that is closer to the target distribution (e.g., the filtering distribution) will yield simulated states more consistent with observations. Resampling is a delicate balancing act between removing particles with low weights whilst ensuring a large enough number of unique particles. Ironically, particle depletion worsens with more precise observations, as this limits the range of state values that are reasonable. Particle depletion is further exacerbated by mis-specified state models. A range of SMC algorithms have been proposed to reduce particle depletion. Notably, the auxiliary PF (APF; Pitt and Shephard (1999)) uses information of the new observation at the prediction step, generating more likely states (for example, see Thomas et al. (2005) for an application to a stage-structured population of grey seals). However, more complex algorithms can be more challenging to understand and/or fit so that understanding when such algorithms may fail can be particularly useful. For example, the APF generally performs poorly when the variability of the system process, $g(\boldsymbol{\alpha}_t|\boldsymbol{\theta}, \boldsymbol{\alpha}_{t-1})$, is large relative to

the observation process, $f(\boldsymbol{y}_t|\boldsymbol{\theta}, \boldsymbol{\alpha}_t)$, (Elvira et al., 2019).

SMC methods are particularly useful for complicated models, including, for example, when there are constraints on the processes, or the state-dimensions change stochastically. Further, many SMC methods only require simulating from the state process model rather than calculating its distribution (the "plug-and-play" feature; Ionides et al., 2015). This is advantageous for process models that are a sequence of random sub-processes, e.g., survival and reproduction, where the underlying probability models can be quite complex, including convolutions (for further discussion see Buckland et al. (2004)). In the remainder of this section we discuss classical and Bayesian approaches that use SMC with an emphasis on "off-line" procedures, where inferences are made from the complete time series, $\boldsymbol{y}_{1:T}$. More detailed description of SMC methods for estimating $\boldsymbol{\theta}$ in SSMs are found in Kantas et al. (2015) and Chopin and Papaspiliopoulos (2020); see also Michaud et al. (2021) for concise pseudocode.

**Classical inference**

Classical approaches based on SMC methods typically produce a stochastic estimate of the likelihood and then apply some optimisation algorithm. A difficulty is that the likelihood to maximise is now stochastic rather than smooth (de Valpine, 2004, Figure 4). To address this issue an importance sampling-type approach was proposed by Hürzeler and Künsch (2001). The idea is to get a smooth but local (in $\boldsymbol{\theta}$) approximation to the likelihood surface around a user-specified $\boldsymbol{\theta}_0$ based on a single sample, and reuse that sample repeatedly for different $\boldsymbol{\theta}$ values to move towards the MLE. If the approximation breaks down, i.e., the weights for a few particles are dominating, a new value of $\boldsymbol{\theta}_0$ is chosen and a new sample is created. In other words use SMC to generate $N$ sets of particles, and calculate $N$ associated estimates of the likelihood, $\hat{L}^i(\boldsymbol{\theta}_0|\boldsymbol{y}_{1:T})$, $i$=1,...,$N$. For a different value of $\boldsymbol{\theta}$, say $\boldsymbol{\theta}'$, calculate $N$ likelihoods using those same $N$ sets of particles and the value $\boldsymbol{\theta}'$, which are denoted $\hat{L}^i_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}'|\boldsymbol{y}_{1:T})$, with the subscripting of $L$ emphasising the dependence on $\boldsymbol{\theta}_0$. The average of

the ratios of these likelihoods given by,

$$\hat{L}(\boldsymbol{\theta}'|\boldsymbol{y}_{1:T}) = \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{L}^i_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}'|\boldsymbol{y}_{1:T})}{\hat{L}^i(\boldsymbol{\theta}_0|\boldsymbol{y}_{1:T})},$$

is then an importance sampling estimate of the likelihood of interest, $L(\boldsymbol{\theta}|\boldsymbol{y}_{1:T})$. The practical problem with this is the difficulty in choosing $\boldsymbol{\theta}_0$, as the variance of the estimate will often be high if $\boldsymbol{\theta}'$ is far from $\boldsymbol{\theta}_0$. In brief, importance sampling across high dimensions is extremely difficult.

An alternative approach is the iterated filtering (IF) method (Ionides et al., 2006). IF uses $\boldsymbol{\theta}$ particles that are randomly perturbed at each time step of each iteration of the filter. The perturbation scale (e.g. standard deviation) follows a schedule of decreasing magnitude such that the average of the $\boldsymbol{\theta}$ particles will converge to the MLE (Ionides et al., 2006, 2011). To improve the performance of the algorithm, a variety of tuning parameters are available including the convergence schedule, initial distribution of $\boldsymbol{\theta}$ particles, perturbation scale, and number of particles (Ionides et al., 2015). Too fast a schedule and/or too few particles can give false convergence (Michaud et al. (2021), Figure 4), so that some care is required in its implementation. For ecological applications of iterated filtering, see for example, Breed et al. (2012); Dowd and Joy (2011); Fasiolo et al. (2016).

**Bayesian inference**

SMC is used in two ways to achieve Bayesian inference. One simulates the states (to obtain an estimate of the marginal likelihood); the other simulates both the states and the model parameters $\boldsymbol{\theta}$, which permits estimation of the parameters via the associated weights of the generated particles. The Liu-West filter (LWF Liu and West, 2001) is an example of the latter. The LWF introduces artificial "state dynamics" for the parameters and simple random perturbations to alleviate particle depletion. Thomas et al. (2005) applied this method to gray seal metapopulation dynamics, while Newman et al. (2006) used it for salmon population dynamics. However, a drawback is bias that is introduced by these artificial perturbations of $\boldsymbol{\theta}$ (Kantas et al., 2015), with Chopin et al. (2013) demonstrating relatively severe biases in an estimate of mean volatility in a stochastic volatility model. Thus we advise caution if

using this method.

Particle MCMC (PMCMC; Andrieu et al., 2010) uses MCMC to sample from the posterior distribution of the parameters, $\boldsymbol{\theta}$, where the likelihood function is replaced by an SMC approximation. PMCMC is "exact approximate" in that the posterior distribution for $\boldsymbol{\theta}$ is correct even though the SMC used to obtain the likelihood is only an unbiased estimate. However, the algorithm is generally computationally expensive. In practice there is a trade-off between the number of particles used and the mixing of the Markov chain (Michaud et al., 2021, Figures 2-3), thus some pilot-tuning is generally advisable. Examples of ecological applications of PMCMC are given by Knape and De Valpine (2012); Hosack et al. (2012); Osada et al. (2019); Finke et al. (2019). See Appendix G for more details. Further, we note that, this approach can also be used to obtain estimates of posterior model probabilities among competing models (Finke et al., 2019).

The SMC$^2$ method (Chopin et al., 2013) is an amalgamation of one SMC algorithm for states, conditional on $\boldsymbol{\theta}$, and another SMC algorithm, Iterated Batch Importance Sampling (Chopin, 2002, IBIS) for parameters. There are two distinct features of SMC$^2$: (i) for given $\boldsymbol{\theta}$, an SMC algorithm is carried out to produce an estimate of the likelihood $L(\boldsymbol{\theta}|\boldsymbol{y}_{1:T})$; (ii) the sample of $\boldsymbol{\theta}$ values are periodically replenished, using a Metropolis-Hastings procedure for example, to produce new values of $\boldsymbol{\theta}$. SMC$^2$ is an accurate method but requires significant computational resources, particularly for long time series; a more computationally efficient alternative is the nested particle filters (Crisan and Miguez, 2018). Application of SMC$^2$ to ecological models appears to be relatively limited, one example being a toy application to the Lotka-Volterra model by Jacob (2015). Appendix H provides additional details.

**Advanced SMC**

We discuss some recent approaches that as yet do not appear to be well known/used within the ecological literature. In particular, fast variational methods have been developed to approximate target distributions of interest with simpler parametric forms and decrease computational effort. The approximating distribution is a parametric distribution, denoted $q(\cdot|\boldsymbol{\lambda})$, that is expressed as a function of some parameters, $\boldsymbol{\lambda}$. The value of $\boldsymbol{\lambda}$ is optimised via a given criterion, such as the Kullback-Leibler divergence between the target and $q(\cdot|\boldsymbol{\lambda})$

(Blei et al., 2017). Such variational methods have been combined with SMC for the filtering distribution. For example, Naesseth et al. (2018) specify the approximate parametric distribution $q(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}, \boldsymbol{\lambda})$ with $\boldsymbol{\lambda}$ iteratively improved via variational inference; and Courts *et al* (unpublished data) propose an MLE of $\boldsymbol{\theta}$ via a variational approach that iteratively improves the lower bound of the likelihood. Alternatively, there have been proposed advanced particle filters that require fewer particles (for example, the improved (I)APF in Elvira et al. (2018) and the optimised (O)APF in Branchini and Elvira (2021)) and approaches for reducing particle depletion that include the use of non-linear transformations of the importance weight, improving the sample diversity following resampling at the expense of introducing additional bias (Ionides, 2008; Martino et al., 2018). Recent works have built on these ideas via the investigation of the effective sample size and proposed novel alternative metrics for adaptive resampling with SMC (Martino et al., 2017)

### 3.3 Markov chain Monte Carlo using Bayesian data augmentation

While the likelihood is analytically intractable due to the unknown latent states that need to be integrated out, the complete data likelihood, $p(\boldsymbol{y}_{1:T}, \boldsymbol{\alpha}_{0:T}|\boldsymbol{\theta})$, is generally trivial to calculate. It is simply the product of the simpler observation and state model distributions (see Equation (3)), and this suggests a Bayesian data augmentation approach for general SSMs, where the latent states are treated as parameters (or auxiliary variables). The joint posterior distribution of the model parameters and latent states can be expressed as,

$$p(\boldsymbol{\theta}, \boldsymbol{\alpha}_{0:T}|\boldsymbol{y}_{1:T}) \propto g(\boldsymbol{\alpha}_0|\boldsymbol{\theta}) \prod_{t=1}^{T} f(\boldsymbol{y}_t|\boldsymbol{\alpha}_t, \boldsymbol{\theta}) g(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}\boldsymbol{\theta}) \pi(\boldsymbol{\theta}),$$

where $\pi(\boldsymbol{\theta})$ denotes the prior for $\boldsymbol{\theta}$. Specifying the latent states as additional parameters means that the joint posterior distribution is easily evaluated (up to proportionality) and standard MCMC algorithms can be used to sample from $p(\boldsymbol{\theta}, \boldsymbol{\alpha}_{0:T}|\boldsymbol{y}_{1:T})$. The marginal posterior distribution for $\boldsymbol{\theta}$ is obtained by integrating over the auxiliary variables, $p(\boldsymbol{\theta}|\boldsymbol{y}_{0:T}) = \int p(\boldsymbol{\theta}, \boldsymbol{\alpha}_{0:T}|\boldsymbol{y}_{1:T}) d\boldsymbol{\alpha}_{0:T}$. This is typically performed indirectly within an MCMC sampling algorithm: given a set of sampled values from the joint posterior distribution, $p(\boldsymbol{\theta}, \boldsymbol{\alpha}_{0:T}|\boldsymbol{y}_{1:T})$, considering only the sampled $\boldsymbol{\theta}$ values integrates out the $\boldsymbol{\alpha}_{0:T}$ values and can be regarded as

a sample from $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:T})$. For some applications, posterior distributions of $\boldsymbol{\alpha}_{0:T}$ are also of interest, and can be similarly obtained.

The performance of the MCMC algorithms in this context ranges from fast and efficient to prohibitively slow and inefficient, depending on model structure, length of time series and observed data. See Appendix A and references therein for further discussion. The sequential Markov structure simplifies single-updates for latent states since the conditional distribution at time $t$ only requires the system processes at times $t-1$ and $t$, and the observation process at time $t$; whereas updating the parameters typically requires the calculations over all time steps (and hence is computationally slower). However, this Markov structure also leads to high posterior correlations between the latent states (posterior correlation with parameters can also be high), leading to high autocorrelation in the MCMC chains, and slow exploration of the parameter space (see, for example, King (2011)). More advanced strategies include "semi-complete" data likelihood approaches, where a selectively chosen subset of unknown states are not treated as auxiliary variables and imputed but are instead integrated out numerically such that the correlation between the auxiliary variables is reduced (Borowska and King; unpublished data), and blocked updates, simultaneously updating multiple states within a single update (Fearnhead, 2011). Hamiltonian Monte Carlo (HMC; Neal (2011)) represents a distinct blocking approach for (necessarily continuous-valued) parameters. HMC uses the shape of the posterior surface to determine the proposal distribution at each iteration, increasing the computational cost in terms of calculating the necessary gradients but having the advantage of traversing the posterior space more quickly. In the case of SSMs HMC is typically applied by simultaneously updating both the model parameters, $\boldsymbol{\theta}$, and auxiliary variables, $\boldsymbol{\alpha}_{0:T}$. For a wide range of ecological SSM examples of Bayesian data augmentation using a variety of different MCMC algorithms, see for instance, Millar and Meyer (2000); Jamieson and Brooks (2004); Brooks et al. (2004); King et al. (2008); King (2011); McClintock et al. (2012, 2013); Monnahan et al. (2017); Best and Punt (2020).

### 3.4 Other sampling based methods for classical inference

For general SSMs there are a variety of methods that permit the calculation of MLEs of the parameters using ideas/techniques often associated with sampling and/or data augmentation.

### 3.4.1 Monte Carlo EM

The EM (Expectation-Maximisation) algorithm (Dempster et al., 1977) is often applied to missing data problems to obtain the MLEs of the model parameters via an iterative two-step process. Step 1 (the "E"-step) involves calculating the expectation of the log-likelihood of the data given the model parameters; Step 2 (the "M"-step) calculates the values of the model parameters that maximises this expectation. Iterating these steps leads to parameter values that converge to their associated MLEs. See Morgan (2000) for further discussion and detailed case studies. For general SSMs the expectation in the E-step is analytically intractable. In this case, Monte Carlo (MC)EM can be applied, which uses a MC estimate of the expectation by simulating a set of $N$ latent states from the smoothing distribution $p(\boldsymbol{\alpha}_{0:T}|\boldsymbol{y}_{1:T}, \boldsymbol{\theta})$, denoted $\boldsymbol{\alpha}^1, \ldots, \boldsymbol{\alpha}^N$ (for example, by using SMC or MCMC). The M-step updates $\boldsymbol{\theta}$ by maximising the MC estimate of the log-likelihood function given by,

$$\mathbb{E}[\ln p(\boldsymbol{y}_{1:T}, \boldsymbol{\alpha}_{0:T}|\boldsymbol{\theta})] \approx \frac{1}{N}\sum_{j=1}^{N} \ln\left[g(\boldsymbol{\alpha}_0^j|\boldsymbol{\theta})\prod_{t=1}^{T} f(\boldsymbol{y}_t|\boldsymbol{\alpha}_t^j, \boldsymbol{\theta})g(\boldsymbol{\alpha}_t^j|\boldsymbol{\alpha}_{t-1}^j, \boldsymbol{\theta})\right].$$

Despite the apparent relative simplicity, the EM algorithm can be slow to converge and there is a computational trade-off between the number of samples generated, $N$, and computational time. Large $N$ is typically required due to the fine-scale maximisation issues owing to the stochastic approximation of the E-step; but leads to increased computational cost. For further discussion and a range of ecological applications, see for example, de Valpine (2012); Halladay (2007); Stoklosa et al. (2015); Picchini and Samson (2018).

### 3.4.2 Data cloning

The idea of data cloning (DC) is to obtain the MLEs of the parameters, and their variances, using a Bayesian framework where the likelihood is artificially amplified so that the corresponding posterior distribution is dominated by the data and the distribution approaches a multiple of the large-sample distribution of the maximum likelihood estimates. This is achieved by creating a new data set consisting of $M$ exact "clones" of the data, denoted $\mathbf{Y}_{1:T}^M = (\boldsymbol{y}_{1:T}^1, \ldots, \boldsymbol{y}_{1:T}^M)$, where $\boldsymbol{y}_{1:T}^j = \boldsymbol{y}_{1:T}$ for $j = 1, \ldots, M$. The posterior distribution is

formed, $p(\boldsymbol{\theta}|\mathbf{Y}_{1:T}^{M})$, assuming some arbitrary prior, and explored using standard techniques (Lele et al. (2007) advocate an MCMC data augmentation approach). The MLE of the parameters is estimated by the posterior mean, and the associated standard error the posterior standard deviation multiplied by $\sqrt{M}$. See Lele et al. (2007) for the associated justification and application to Gompertz and Ricker models. DC is more computationally demanding than a standard Bayesian approach (with $M = 1$) because every clone requires separate latent states. In general, there is a trade-off between accuracy in terms of the number of clones, $M$, and computational expense. Using an MCMC data augmentation approach, the number of additional imputed latent states will be $M$ times larger than the observed data. See Lele et al. (2010) for further discussion regarding determining a suitable value for $M$ and additional inference on the latent states; and for ecological applications, see for example, Ponciano et al. (2009); Schlägel and Lewis (2014).

### 3.4.3 Monte Carlo kernel likelihood

The Monte Carlo kernel likelihood (MCKL) approach uses kernel smoothing of the posterior distribution of $\boldsymbol{\theta}$ to approximate its density (de Valpine, 2004, 2012). In particular, de Valpine (2004) proposed the use of a Bayesian approach to obtain a sample of $\boldsymbol{\theta}$ values from its associated posterior distribution before the impact of the prior is "divided out", yielding a likelihood approximation. This is, at first glance, a potentially hazardous idea because kernel density estimation suffers from a curse of dimensionality, in this case in relation to the number of model parameters i.e. the dimension of $\boldsymbol{\theta}$. However, it is at its most accurate for estimating a mode (maximum), the sample sizes can be very large due to MCMC sampling, corrections from distribution theory can be applied, and the whole procedure can be iterated to improve accuracy. de Valpine (2004) gave highly skewed posterior examples where the approach was accurate for even up to 20 parameter dimensions. This approach has been seldom applied, but remains potentially useful (see Karban and de Valpine, 2010).

### 3.5 Likelihood-free methods

For SSMs that are highly non-linear, model-fitting can be particularly difficult. In such instances, likelihood-free (or information-reduction) methods (Hartig et al., 2011), namely, Approximate Bayesian Computation (ABC; Sisson et al. (2018); Martin et al. (2019)) and synthetic likelihood (SL; Wood (2010); Drovandi et al. (2018)), are alternative approaches. These approaches rely only on the ability to simulate data from the SSM, given the model parameters. This involves simulating the underlying states ($\boldsymbol{\alpha}_{1:T}$) from the state model, and then, conditional on these, simulating data ($\boldsymbol{y}_{1:T}$) from the observation model, another example of "plug-and-play". Model parameter values (and potentially system states) are deemed plausible if they generate data that are similar to the observed data. Similarity is typically defined by transforming the data into a vector of lower dimensional summary statistics and using a distance measure, such as the squared Mahalanobis distance function. For ABC there is a further specification to be made in terms of the tolerance (or threshold), such that parameter values are retained if this distance measure is within the tolerance. In practice a range of tolerances may be used to assess the sensitivity of the approximate posterior samples to this level, and to determine an appropriate tolerance to use. For the SL approach, a parametric (multivariate normal) distribution is assumed for the summary statistics, which implicitly defines the given distance metric and removes the need for any additional tolerance level to be specified. See Fasiolo and Wood (2018) for further discussion and comparison of the ABC and SL approaches.

ABC and SL have been successfully applied to several ecological state-space models, (Wood, 2010; Scranton et al., 2014; Fasiolo and Wood, 2018; Ruiz-Suarez et al., 2020). The principle challenge lies in the specification of the reduced summary statistics. For the ecological applications considered, summary statistics used range from the raw data values themselves to combinations of mean(s)/standard deviation(s), regression models fitted to the values (or their differences) and/or auto-covariance functions. In practice, biological knowledge may be useful to identify the most important or useful summary statistics, which may be weighted accordingly. Similarly it may be possible to identify particular areas of lack of fit of the model to the data via such metrics. For increasingly complex SSMs, for example, higher-order Markov system processes, or highly nonlinear dynamics, it is often still trivial

to simulate data, and in these circumstances ABC and SL may provide a very useful tool.

# 4 Discussion

Regardless of the model-fitting method, checks should always be conducted to assess the fit of the SSM to the data, which can help diagnose poor-fitting models or even structural estimation problems. We discuss both relative and absolute goodness-of-fit measures before describing available SSM software and offering some general practical recommendations.

## 4.1 Model assessment

### Parameter redundancy

# 5 Kalman filter

The Kalman filtering algorithm evaluates $\mathbb{E}(S_t|y_1,\ldots,y_t)$. Let $\mu_t = \mathbb{E}(S_t|y_1,\ldots,y_t)$ and $P_t = \mathrm{Var}(S_t|y_1,\ldots,y_t)$, then using the general matrix representations above, the Kalman filter recursions are defined as

$$\begin{aligned}
\mathbf{v}_t &= \mathbf{y}_t - \mathbf{Z}_{tt} \\
\mathbf{F}_t &= \mathbf{Z}_t\mathbf{P}_t\mathbf{Z}_t' + \mathbf{H}_t \\
\mathbf{G}_t &= \mathbf{\Lambda}_t\mathbf{P}_t\mathbf{Z}_t'\mathbf{F}_t^{-1} \\
_{t+1} &= \mathbf{\Lambda}_{tt} + \mathbf{G}_t\mathbf{v}_t \\
\mathbf{L}_t &= \mathbf{\Lambda}_t - \mathbf{G}_t\mathbf{Z}_t \\
\mathbf{P}_{t+1} &= \mathbf{\Lambda}_t\mathbf{P}_t\mathbf{L}_t' + \mathbf{Q}_t.
\end{aligned} \tag{7}$$

Following the implementation of the Kalman filter, the likelihood for the unknown parameters based on all of the observed data is defined by

$$\log L(;y_1,\ldots,y_T) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\log|\mathbf{F}_t| + \mathbf{v}_t'\mathbf{F}_t^{-1}\mathbf{v}_t\right), \tag{8}$$

where  denotes all the model parameters. The values of $\mathbf{v}_t$ and $\mathbf{F}_t$ are calculated automatically when the Kalman filter is implemented. An attraction of the Kalman filter is that the recursions only require knowledge of values at the previous time, resulting in a fast algorithm.

The smoothing stage of the Kalman filter determines $\mathbb{E}(S_t|y_1,\ldots,y_T)$, and we give below details of what is known as fixed-interval smoothing. The algorithm starts with the final estimates $_T$ and $\mathbf{P}_T$ and iterates backwards. The recursions are defined by:

$$_{t|T} = _t + \mathbf{P}_t^*(_{t+1|T} - _{t+1t})$$

and

$$\mathbf{P}_{t|T} = \mathbf{P}_t + \mathbf{P}_t^*(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t})\mathbf{P}_t^{'*}$$

where $\mathbf{P}_t^* = \mathbf{P}'_{tt+1}\mathbf{P}_{t+1|t}^{-1}$ for $t = T-1,\ldots,1$ with $_{T|T} = _T$ and $\mathbf{P}_{T|T} = \mathbf{P}_T$.

There are several different ways in which one might initialise the Kalman filter iterations, and these are compared and contrasted in **?**.

SSMs can encounter issues of parameter redundancy such that parameters are not uniquely estimable given the observed data (Knape, 2008; Auger-Méthé et al., 2016; Cole, 2020). Near parameter redundancy can also lead to poor precision and poor model-fitting performance. This has been observed for the process and observation errors, particularly for short time series, and ecological data sets are often of short duration, e.g., 10 years. Cole and McCrea (2016) discuss analytical approaches for detecting redundancy problems in SSMs, while Lele et al. (2010) describe the use of data cloning to do so, where the behaviour of the eigenvalues of the posterior variance-covariance matrix provide a measure of estimability. Repeated data collection within the study design can aid with disentangling these errors (Dennis et al., 2010; Knape et al., 2013; Besbeas and Morgan, 2020); while Besbeas and Morgan (2017) explored the use of penalised likelihood and pseudo-replicated series (involving replacing time-series values by neighbouring values).

Constructing models that are (near) parameter redundant is, unfortunately, all too easy, particularly when considering multiple biological mechanisms acting on the system process and/or where the observed data are limited to only a subset of the latent states (e.g. only young/adults for age-structured populations; or where multiple processes occur between observations). Even for very simple univariate LG-SSMs, problems can arise in separating the effects of state process variation and the observation noise. As both Dennis et al. (2006) and Auger-Méthé et al. (2016) have shown there can be a ridge in the joint likelihood function for the state and observation variances whereby nearly equivalent fits can result from combinations of low process variation-high observation noise and high process variation-low observation noise. This non-identifiability or weak identifiability can easily arise in other seemingly simple SSMs and hierarchical models in general and the unintended effects on statistical inference can be striking—see Staples et al. (2009) for the effects on estimates of the probability of quasi-extinction and see Lele (2020) for potential conflicts in Bayesian inference arising from seemingly harmless alternative parameterizations. The bottom line cautionary note is that the available data necessarily constrains the complexity of the SSM that can be fit.

To improve estimability and to thus fit more complex SSMs, additional external data may be needed (Cole and McCrea, 2016). Incorporating different kinds of data from different surveys or data collection procedures can lead to integrated population models. For example, count data have been combined with other data such as ring-recovery (Besbeas et al., 2002, 2005; King et al., 2008) or nest productivity data (Abadi et al., 2010; Besbeas and Morgan, 2019) to separate survival from reproduction. Also, the inclusion of covariates, such as environmental conditions and/or density dependence, for sub-process models in the state model has been shown to improve identifiability (Besbeas et al., 2002; Finke et al., 2019; Polansky et al., 2021).

### Model assessment

Once a model has been fitted to data it is useful to investigate how well the model describes the data, i.e., the *absolute goodness-of-fit*. A well-fitting model provides confidence that the main biological drivers have been identified and the estimated parameters are mean-

ingful. In contrast a poor-fitting model would lead to a more cautious interpretation and ideally identify where the lack-of-fit is most prevalent, potentially leading to alternative better fitting models. Newman et al. (2014, §5.6) outline the challenges for absolute goodness-of-fit procedures for SSMs, and provide a range of alternative procedures, including plotting appropriate residuals, measures of discrepancy (also called innovations), Bayesian $p$-values and cross-validation. Additional discussion is provided by Auger-Méthé et al. (2021). For a general guide to Bayesian model checking for ecologists see Conn et al. (2018), who emphasised the conservative nature of Bayesian $p$-values, as while they detect extreme model inadequacy, they can fail to detect more subtle misfit. Recent work has also focused on the one-step ahead predictions of the observations using different statistics to assess potential model mismatch (Karban and de Valpine, 2010; Djuric and Míguez, 2010; Thygesen et al., 2017; Elvira et al., 2017, 2021). Extending SSMs to integrated population models provides further goodness-of-fit challenges with Besbeas and Morgan (2014) proposing a calibrated simulation approach.

### Model selection

Often there are multiple possible models that represent competing biological hypotheses and one may want to compare the models in terms of *relative goodness-of-fit* and to select a single model. Hooten and Hobbs (2015) provide a guide to model selection for ecologists; and for the specific application to SSMs see Newman et al. (2014, §5.4) and Auger-Méthé et al. (2021). A common model selection approach for ecological models is to use information criterion, notably Akaike Information Criterion (AIC) and its variants (including the method of Bengtsson and Cavanaugh (2006) for SSMs, though this is relatively complex and computationally demanding) within classical analyses. The Watanabe-Akaike Information Criterion (WAIC; Watanabe and Opper (2010); Watanabe (2013, 2021)) is the preferred criterion for Bayesian analyses. Alternatively, within the Bayesian framework, posterior model probabilities (and the related idea of *Evidence*; Finke et al. (2019)) may be calculated to provide a quantitative comparison of competing models, and permit model-averaged estimates. These can be calculated using reversible jump (RJ)MCMC (Jamieson and Brooks, 2004; King et al., 2009; Gimenez et al., 2009; King, 2011), SMC methods (Urteaga et al.,

2016; Martino et al., 2017; Finke et al., 2019), or general "normalising constant" methods (de Valpine, 2012). Auger-Méthé et al. (2021) discuss these criteria further and emphasise that for information criteria the system processes need to be of the same dimension. To address this issue Besbeas, McCrea and Morgan (unpublished data) present ways to ensure that such model comparisons do not affect the asymptotic distribution of a likelihood-ratio test.

## 5.1 Software

Many software packages provide one or more methods for some assortment of SSMs, and these range from application-specific domains to general model specification. Given the huge variety of packages, we classify some of the most widely used general packages in Table 1, excluding domain-specific applications (for example, packages for animal movement models, etc). However we do include some `R` packages that focus on types of models. For example, of particular interest to ecologists working with ecological communities and multiple species is the `R` package `MARSS` for "multivariate autoregressive state-space" modelling (Holmes et al., 2012) which is an implementation of probabilistic models for community stability, among other things (Ives et al., 2003). The underlying of assumptions of MARSS are normality and linearity; for an application see Hampton et al. (2013).

Our list is non-exhaustive and the categories are coarse, as there are variants of each method and different ranges of models for each package, but will allow researchers to explore some of the methods summarised above. We would also recommend further investigation for additional domain-specific software/packages. Comparisons in terms of efficiency or range of supported models are beyond our scope. Most of the covered tools are, or can be interfaced from, `R` packages. Of course, individuals who are reasonably fluent writing `R` code can implement their own customised algorithms. Tools in Python, Julia, standalone `C++`, or other languages are not listed.

A few important and interesting themes emerge from Table 1. MCMC "black-box" software packages such as `nimble` (de Valpine et al., 2017), `JAGS` (Plummer, 2016), `Open/WinBUGS` (Lunn et al., 2012), and `Stan` (Stan Development Team, 2021) are not specific to SSMs and

| | KF packages | TMB | HMM packages | LibBi | pomp | nimble | JAGS | Open/WinBUGS | Stan | dclone |
|---|---|---|---|---|---|---|---|---|---|---|
| KF | ✓ | | | | | 1 | | | | |
| EKF | | | | ✓ | | 1 | | | | |
| UKF | | | | | | 1 | | | | |
| EnKF | | | | | ✓ | ✓ | | | | |
| Laplace | | ✓ | | | | 2 | | | | |
| Discrete approx (HMM) | | | ✓ | | | 4 | 4 | 4 | 4 | |
| SMC/PMCMC | | | | ✓ | ✓ | ✓ | | | | |
| IF | | | | | ✓ | ✓ | | | | |
| LWF | | | | | ✓ | 1 | | | | |
| SMC$^2$ | | | | ✓ | | 1 | | | | |
| MCMC (data augmentation) | | | | | | ✓ | ✓ | ✓ | ✓ | |
| MCEM | | | | | | ✓ | | | | |
| MCKL | | | | | | 3 | 3 | 3 | 3 | |
| DC | | | | | | 3 | 3 | 3 | 3 | ✓ |
| ABC | | | | | ✓ | 1 | | | | |
| SL | | | | | ✓ | | | | | |
| VI | | | | | | | | | ✓ | |

Table 1: Software packages (columns) against methods (rows). KF `R` packages include `dlm`, `FKF`, `KFAS`, `MARSS`; HMM `R` packages include `HiddenMarkov`, `depmixS4`, `msm`. `TMB` = Template Model Builder (Kristensen et al., 2016); `dclone` (Sólymos, 2010) is an `R` package. 1: Method could be supported in `nimble`'s algorithm language but is not part of the package. 2: Method is in development. 3: Any MCMC engine could be used for MCKL or DC. 4: HMMs must be marginalised in `Stan` and may be handled in different ways in `nimble`, `JAGS` or `Open/WinBUGS`.

support general model structures. `LibBi` (Murray, 2015) (and associated `R` package, `rbi`) is an SMC-specific package that supports a range of these methods. The packages `pomp` (King et al., 2016) and `nimble` stand out as the most general, making them useful for methodological research and comparisons *per se*, with `nimble` the most general due to its combination of extensible models and embedded algorithm language. Using `nimble`, the `nimbleSMC` package supports SMC methods, including some more advanced techniques such as APF, with the full model flexibility of the `BUGS` model language.

### 5.2 Practical guidance for implementation

Here we offer guidance for fitting SSMs based on our experiences, admittedly subjective, with the different model-fitting approaches. There is no universal "best" approach, as they nearly

all try to achieve valid classical or Bayesian inference but differ in difficulty of implementation, delicacy of tuning, and computational cost. In Table 2, we have compressed the pros and cons of each method into an overall score from 1 (hardest / slowest) to 4 (easiest / fastest), regardless of classical vs Bayesian analysis goals. We emphasise that the efficiency of different model-fitting algorithms depends on both the model structure and the observed data, so these are coarse scores. For example, high quality data with relatively small observation error can lead to high particle depletion within SMC approaches; conversely, small system process variation can lead to very high correlation and poor mixing within MCMC approaches.

General advice is to start with a relatively simple model, potentially oversimplified, and correspondingly simple model-fitting procedures, so as to get initial parameter estimates and predictions of observations. Goodness-of-fit assessment may indicate unacceptable results, and then one formulates a more complex SSM structure and applies a more sophisticated fitting procedure. And more iterations may ensue.

For many problems, the Kalman Filter and its approximations, and MCMC with data augmentation of latent states, are the first stop for practical and effective methods. Maximum likelihood estimation with one of the KF-based methods can be fast and thus amenable to resampling-based model-checking and model-selection methods such as the moving-block bootstrap. MCMC, on the other hand, gives Bayesian accounting of uncertainty and use of prior information when desired. For relatively short time series, e.g., $T \leq 20$, an MCMC fit may be fine even with relatively high correlations, so long as a long enough run can be made in an acceptable amount of time.

When classical inference is desired but KF-based methods are inadequate, one can turn to Monto Carlo approaches, Laplace approximation, or discretisation. If the dimension of states ($n$) at each time step is small, discretisation into an HMM is particularly attractive and can give efficient and accurate results. While one needs to check the effects of different levels of discretisation of the state vector components, depending on the dimension of the state and model complexity, this is a relatively easy procedure. When MCMC is practical, the general Monte Carlo MLE methods of MCEM, MCKL and DC can be considered. Among these, DC can potentially be run from any MCMC engine and requires tuning the number of data clones. MCEM has simpler MCMC needs (only sampling latent states) and requires

tuning parameters for stochastic convergence assessment. Both of these can require some trial and error. MCKL can be run from any MCMC engine but currently lacks a general package. Laplace approximation can be simple, fast, and practical, but its accuracy should be assessed; e.g., by simulating data from the fitted model and seeing if the score function (the derivative of the log likelihood) equals zero at the MLEs (personal communication, Hans Skaug). For state models that have probability distributions that are difficult to evaluate but can be simulated, then the IF algorithm is attractive as a "plug-and-play" procedure.

When Bayesian inference is desired (and KF-based methods are inadequate), again Monte Carlo approaches, particularly MCMC or SMC methods, can be employed. In general for complex higher-dimensional systems, such sampling based approaches may become necessary, though these often require a greater amount of pilot-tuning and are computationally more expensive. SMC methods come into their own when MCMC does not work well, either because the likelihood is hard to evaluate, or the posterior is highly correlated, or specialised MCMC sampling methods would be needed. The latter case might arise to respect a constraint in the model, or to sample between different numbers of latent state dimensions, or to sample latent states for every individual in a population rather than for cohorts or whole populations. However, while SMC methods sound good in principle, in practice they can require large simulation sample sizes and can perform poorly for inadequate models, so they can take some effort to make work well. That said, there are many problems where MCMC does not work well (although standard MCMC implementations do indeed work well in many situations, too), or where online (real-time) forecasting is needed, so SMC is a rapidly developing field.

Synthetic likelihood and approximate Bayesian computation are most useful when none of the above work, likelihoods are beyond reach, and one is willing to formulate *ad-hoc* metrics of what "fitting the data" should mean. For example, one might use various correlations over time and among observation dimensions as metrics that should match between model and data, in a spirit similar to methods-of-moments approaches. Then SL and ABC offer a path to a rough form of estimation and inference, contingent on the chosen metrics. Variational methods, on the other hand, come into their own when the time-series is so long that above methods bog down computationally.

In summary, there is no single algorithm that dominates for general SSMs and model-fitting may require consideration of more than one technique to satisfactorily fit a model to data. However, there is an ever-growing set of computational algorithms and associated software (as discussed in Section 5.1) in the ecologists toolbox that can now be applied to a vast range of ecologically interesting and complex models.

| | LG | ΔNL | ΔNG | NL/NG | HNL |
|---|---|---|---|---|---|
| Likelihood evaluation/estimation | | | | | |
| KF | ✓✓✓✓ | | | | |
| EKF | =KF | ✓ ✓ | | | |
| UKF | =KF | ✓✓✓ | | | |
| EnKF | ✓✓✓ | ✓✓✓ | ✓✓✓ | ✓✓ | |
| Approx KF | =KF | | ✓✓✓ | | |
| Laplace | =KF | ✓✓ | ✓✓✓ | ✓✓ | |
| Discrete ($n = 1, 2$) | — | ✓✓✓✓ | ✓✓✓✓ | ✓✓✓✓ | ✓✓✓ |
| Discrete ($n$ = 3-5) | — | ✓✓ | ✓✓ | ✓✓ | ✓ |
| SMC (direct optimisation) | — | ✓ | ✓ | ✓ | ✓ |
| Classical approaches | | | | | |
| IF | — | ✓✓ | ✓✓ | ✓✓ | ✓ |
| MCEM | — | ✓✓ | ✓✓ | ✓✓ | ✓ |
| MCKL | — | ✓✓ | ✓✓ | ✓✓ | ✓ |
| DC | — | ✓✓ | ✓✓ | ✓✓ | ✓ |
| SL | — | ✓ | ✓ | ✓ | ✓✓ |
| Bayesian approaches | | | | | |
| SMC/PMCMC | — | ✓✓ | ✓✓ | ✓✓ | ✓ |
| MCMC (data augmentation) | — | ✓✓✓ | ✓✓✓ | ✓✓✓ | ✓ |
| $SMC^2$ | — | ✓✓ | ✓✓ | ✓✓ | ✓ |
| Variational | — | ✓✓ | ✓✓ | ✓✓ | ✓ |
| ABC | — | ✓ | ✓ | ✓ | ✓✓ |

Table 2: Overview of the general appropriateness of the different model-fitting tools for different categories of models: LG = linear and Gaussian; ΔNL/NG = mild non-linearity and/or non-Gaussian; NL/NG = non-linear and/or non-Gaussian; HNL = highly non-linear SSMs. ✓✓✓✓ = generally easily applied and efficient; ✓✓✓ = should be generally applicable but some mild additional considerations (such as further reasonable assumptions; increased computational effort); ✓✓ = may be applicable under certain conditions with additional considerations (such as stronger modelling assumptions; algorithm requires tuning; increased computation effort); ✓ = potentially applicable but greater considerations (such as strong approximations required; potential bias; information reduction; large computational effort); — denotes that although the approach can be applied it is not necessary due to the analytical calculation of the likelihood via the KF, $n$ for the discrete approaches is the dimension of the state vector.

## Acknowledgements

RK was supported by the Leverhulme research fellowship RF-2019-299. RM was supported by EPSRC grant EP/S020470/1. BJTM was supported by the Emeritus Leverhulme Fellowship EM-2017-061.

### Authors contribution statement

RK and RM conceived the review paper; EV, RK, and KN developed the focus and breadth of the paper. All authors contributed to the writing of the paper, contributed critically to the drafts and gave final approval for publication.

### Conflict of Interest Statement

All authors declare no conflicts of interest.

## References


Abadi, F., O. Gimenez, B. Ullrich, R. Arlettaz, and M. Schaub (2010). Estimation of immigration rate using integrated population models. *Journal of Applied Ecology 47*, 393–400.

Aeberhard, W. H., J. Mills Flemming, and A. Nielsen (2018). Review of state-space models for fisheries science. *Annual Review of Statistics and Its Application 5*, 215–235.

Albertsen, C. M., K. Whoriskey, D. Yurkowski, A. Nielsen, and J. Mills Flemming (2015). Fast fitting of non-Gaussian state-space models to animal movement data via Template Model Builder. *Ecology 96*, 2598–2604.

Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B 72*, 269–342.

Auger-Méthé, C. M., Albertsen, I. D. Jonsen, A. E. Derocher, D. C. Lidgard, K. R. Studholme, W. D. Bowen, G. T. Crossin, and J. Mills-Flemming (2017). Spatiotemporal modelling of marine movement data using Template Model Builder (TMB). *Marine Ecology Progress Series 565*, 237–249.

Auger-Méthé, M., C. Field, C. M. Albertsen, A. E. Derocher, M. A. Lewis, I. D. Jonsen, and J. M. Flemming (2016). State-space models' dirty little secrets: even simple linear Gaussian models can have estimation problems. *Scientific Reports 6*, 1–10.

Auger-Méthé, M., K. Newman, D. Cole, F. Empacher, R. Gryba, A. A. King, V. Leos-Barajas, J. Mills Flemming, A. Nielsen, G. Petris, and L. Thomas (2021). A guide to state–space modeling of ecological time series. *Ecological Monographs 91*, 1–38.

Bengtsson, T. and J. E. Cavanaugh (2006). An improved Akaike information criterion for state-space model selection. *Computational Statistics & Data Analysis 50*, 2635–2654.

Berliner, L. M. (1996). Hierarchical Bayesian time series models. In *Maximum Entropy and Bayesian Methods*, pp. 15–22. Springer.

Besbeas, P., S. N. Freeman, and B. J. T. Morgan (2005). The potential of integrated population modelling. *Australian & New Zealand Journal of Statistics 47*, 35–48.

Besbeas, P., S. N. Freeman, B. J. T. Morgan, and E. A. Catchpole (2002). Integrating mark–recapture–recovery and census data to estimate animal abundance and demographic parameters. *Biometrics 58*, 540–547.

Besbeas, P. and B. J. T. Morgan (2012). Kalman filter initialization for integrated population modelling. *Journal of the Royal Statistical Society: Series C 61*, 151–162.

Besbeas, P. and B. J. T. Morgan (2014). Goodness-of-fit of integrated population models using calibrated simulation. *Methods in Ecology and Evolution 5*, 1373–1382.

Besbeas, P. and B. J. T. Morgan (2017). Variance estimation for integrated population models. *AStA Advances in Statistical Analysis 101*, 439–460.

Besbeas, P. and B. J. T. Morgan (2019). Exact inference for integrated population modelling. *Biometrics 75*, 475–484.

Besbeas, P. and B. J. T. Morgan (2020). A general framework for modeling population abundance data. *Biometrics 76*, 281–292.

Best, J. K. and A. E. Punt (2020). Parameterizations for Bayesian state-space surplus production models. *Fisheries Research 222*, 105411.

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association 112*, 859–877.

Branchini, N. and V. Elvira (2021). Optimized auxiliary particle filters: adapting mixture proposals via convex optimization. *37th Conference on Uncertainty in Artificial Intelligence*.

Breed, G. A., D. Costa, I. Jonsen, P. Robinson, and J. Mills-Flemming (2012). State-space methods for more completely capturing behavioral dynamics from animal tracks. *Ecological Modelling 235*, 49–58.

Brooks, S. P., R. King, and B. J. T. Morgan (2004). A Bayesian approach to combining animal abundance and demographic data. *Animal Biodiversity and Conservation 27*, 515–529.

Buckland, S. T., K. B. Newman, L. Thomas, and N. B. Koesters (2004). State-space models for the dynamics of wild animal populations. *Ecological Modelling 171*, 157–175.

Chopin, N. (2002). A sequential particle filter method for static models. *Biometrika 89*, 539–552.

Chopin, N., P. E. Jacob, and O. Papaspiliopoulos (2013). SMC2: An efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society: Series B 75*, 397–426.

Chopin, N. and O. Papaspiliopoulos (2020). *An Introduction to Sequential Monte Carlo*. Springer.

Cole, D. J. (2020). *Parameter Redundancy and Identifiability*. Chapman and Hall/CRC Press.

Cole, D. J. and R. S. McCrea (2016). Parameter redundancy in discrete state-space and integrated models. *Biometrical Journal 58*, 1071–1090.

Conn, P. B., D. S. Johnson, P. J. Williams, S. R. Melin, and M. B. Hooten (2018). A guide to Bayesian model checking for ecologists. *Ecological Monographs 88*(4), 526–542.

Cressie, N. (2015). *Statistics for Spatial Data*. John Wiley & Sons.

Cressie, N., C. A. Calder, J. S. Clark, J. M. V. Hoef, and C. K. Wikle (2009). Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecological Applications 19*, 553–570.

Crisan, D. and J. Miguez (2018). Nested particle filters for online parameter estimation in discrete-time state-space Markov models. *Bernoulli 24*, 3039–3086.

de Valpine, P. (2004). Monte Carlo state-space likelihoods by weighted posterior kernel density estimation. *Journal of the American Statistical Association 99*, 523–536.

de Valpine, P. (2012). Frequentist analysis of hierarchical models for population dynamics and demographic data. *Journal of Ornithology 152*, 393–408.

de Valpine, P. and A. Hastings (2002). Fitting population models incorporating process noise and observation error. *Ecological Monographs 72*, 57–76.

de Valpine, P. and R. Hilborn (2005). State-space likelihoods for nonlinear fisheries time-series. *Canadian Journal of Fisheries and Aquatic Sciences 62*, 1937–1952.

de Valpine, P., D. Turek, C. Paciorek, C. Anderson-Bergman, D. Temple Lang, and R. Bodik (2017). Programming with models: writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics 26*, 403–417.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B 39*, 1–38.

Dennis, B., J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples (2006). Estimating density dependence, process noise, and observation error. *Ecological Monographs 76*, 323–341.

Dennis, B., J. M. Ponciano, and M. L. Taper (2010). Replicated sampling increases efficiency in monitoring biological populations. *Ecology 91*, 610–620.

Djuric, P. M. and J. Míguez (2010). Assessment of nonlinear dynamic models by Kolmogorov-Smirnov statistics. *IEEE Transactions on Signal Processing 58*, 5069–5079.

Dowd, M. and R. Joy (2011). Estimating behavioral parameters in animal movement models using a state-augmented particle filter. *Ecology 92*, 568–575.

Drovandi, C. C., C. Grazian, K. Mengersen, and C. Robert (2018). Approximating the likelihood in ABC. In *Handbook of Approximate Bayesian Computation*, pp. 321–368. CRC Press.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods*. Oxford University Press.

Einarsson, A., U. Hauptfleisch, P. Leavitt, and A. Ives (2016). Identifying consumer–resource population dynamics using paleoecological data. *Ecology 97*, 361–371.

Elvira, V., L. Martino, M. F. Bugallo, and P. M. Djurić (2018). In search for improved auxiliary particle filters. In *2018 26th European Signal Processing Conference*, pp. 1637–1641.

Elvira, V., L. Martino, M. F. Bugallo, and P. M. Djuric (2019). Elucidating the auxiliary particle filter via multiple importance sampling. *IEEE Signal Processing Magazine 36*, 145–152.

Elvira, V., L. Martino, and P. Closas (2020). Importance Gaussian quadrature. *IEEE Transactions on Signal Processing 69*, 474–488.

Elvira, V., J. Míguez, and P. Djurić (2017). Adapting the number of particles in sequential Monte Carlo methods through an online scheme for convergence assessment. *IEEE Transactions on Signal Processing 65*, 1781–1794.

Elvira, V., J. Míguez, and P. Djurić (2021). On the performance of particle filters with adaptive number of particles. *Statistics and Computing 31*(81), 1–18.

Fasiolo, M., N. Pya, and S. N. Wood (2016). A comparison of inferential methods for highly nonlinear state space models in ecology and epidemiology. *Statistical Science*, 96–118.

Fasiolo, M. and S. N. Wood (2018). ABC in ecological modelling. In *Handbook of Approximate Bayesian Computation*, pp. 597–622. CRC Press.

Fearnhead, P. (2011). MCMC for state-space models. In *Handbook of Markov Chain Monte Carlo: Methods and Applications*, pp. 513–529. CRC Press.

Finke, A., R. King, A. Beskos, and P. Dellaportas (2019). Efficient sequential Monte Carlo algorithms for integrated population models. *Journal of Agricultural, Biological and Environmental Statistics 24*, 204–224.

Gimenez, O., S. Bonner, R. King, R. A. Parker, S. P. Brooks, L. E. Jamieson, V. Grosbois, B. J. T. Morgan, and L. Thomas (2009). WinBUGS for population ecologists: Bayesian modelling using Markov chain Monte Carlo (MCMC) methods. In *Modeling Demographic Processes in Marked Populations*, pp. 885–918. Springer.

Halladay, A. (2007). Parameter estimation of a robust state-space model via Monte Carlo methods and the EM algorithm. Master's thesis, Dalhousie University.

Hampton, S. E., E. E. Holmes, L. P. Scheef, M. D. Scheuerell, S. L. Katz, D. E. Pendleton, and E. J. Ward (2013). Quantifying effects of abiotic and biotic drivers on community dynamics with multivariate autoregressive (MAR) models. *Ecology 94*, 2663–2669.

Hartig, F., J. Calabrese, B. Reineking, T. Wiegand, and A. Huth (2011). Statistical inference for stochastic simulation models–theory and application. *Ecology Letters 14*(8), 816–827.

Holmes, E. E., E. J. Ward, and K. Wills (2012). MARSS: multivariate autoregressive state-space models for analyzing time-series data. *R journal 4*.

Hooten, M. B. and N. T. Hobbs (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs 85*, 3–28.

Hooten, M. B., D. S. Johnson, B. T. McClintock, and J. M. Morales (2017). *Animal Movement: Statistical Models for Telemetry Data*. CRC Press.

Hosack, G. R., G. W. Peters, and K. R. Hayes (2012). Estimating density dependence and latent population trajectories with unknown observation error. *Methods in Ecology and Evolution 3*, 1028–1038.

Hürzeler, M. and H. Künsch (2001). Approximating and maximising the likelihood for a general state-space model. In *Sequential Monte Carlo Methods in Practice*, pp. 159–175. Springer.

Ionides, E. L. (2008). Truncated importance sampling. *Journal of Computational and Graphical Statistics 17*, 295–311.

Ionides, E. L., A. Bhadra, Y. Atchadé, and A. A. King (2011). Iterated filtering. *Annals of Statistics 39*, 1776–1802.

Ionides, E. L., C. Bretó, and A. A. King (2006). Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences 103*, 18438–18443.

Ionides, E. L., D. Nguyen, Y. Atchadé, S. Stoev, and A. A. King (2015). Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proceedings of the National Academy of Sciences 112*, 719–724.

Ives, A. R., B. Dennis, K. Cottingham, and S. R. Carpenter (2003). Estimating community stability and ecological interactions from time-series data. *Ecological monographs 73*, 301–330.

Jacob, P. E. (2015). Sequential Bayesian inference for implicit hidden Markov models and current limitations. *ESAIM: Proceedings and Surveys 51*, 24–48.

Jamieson, L. E. and S. P. Brooks (2004). Density dependence in North American ducks. *Animal Biodiversity and Conservation 27*, 113–128.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions ASMA, Series D 82*, 35–45.

Kantas, N., A. Doucet, S. S. Singh, J. Maciejowski, and N. Chopin (2015). On particle methods for parameter estimation in state-space models. *Statistical Science 30*, 328–351.

Karban, R. and P. de Valpine (2010). Population dynamics of an Arctiid caterpillar–tachinid parasitoid system using state-space models. *Journal of Animal Ecology 79*, 650–661.

Katzfuss, M., J. R. Stroud, and C. K. Wikle (2016). Understanding the ensemble Kalman filter. *The American Statistician 70*, 350–357.

Katzfuss, M., J. R. Stroud, and C. K. Wikle (2020). Ensemble Kalman methods for high-dimensional hierarchical dynamic space-time models. *Journal of the American Statistical Association 115*, 866–885.

Kindsvater, H. K., N. K. Dulvy, C. Horswill, M. Juan-Jordá, M. Mangel, and J. Matthiopoulos (2018). Overcoming the data crisis in biodiversity conservation. *Trends in Ecology & Evolution 33*, 676–688.

King, A. A., D. Nguyen, and E. L. Ionides (2016). Statistical inference for partially observed Markov processes via the `R` package `pomp`. *Journal of Statistical Software 69*.

King, R. (2011). Statistical ecology. In *Handbook of Markov Chain Monte Carlo: Methods and Applications*. CRC Press.

King, R. (2014). Statistical ecology. *Annual Review of Statistics and its Application 1*, 401–426.

King, R., S. P. Brooks, C. Mazzetta, S. N. Freeman, and B. J. T. Morgan (2008). Identifying and diagnosing population declines: a Bayesian assessment of lapwings in the UK. *Journal of the Royal Statistical Society: Series C 57*, 609–632.

King, R., S. P. Brooks, and B. J. T. Morgan (2004). A Bayesian approaching to combining animal abundance and demographic data. *Animal Biodiversity and Conservation 27*, 515–529.

King, R., B. J. T. Morgan, O. Gimenez, and S. P. Brooks (2009). *Bayesian Analysis for Population Ecology*. CRC Press.

Knape, J. (2008). Estimability of density dependence in models of time series data. *Ecology 89*(11), 2994–3000.

Knape, J., P. Besbeas, and P. de Valpine (2013). Using uncertainty estimates in analyses of population time series. *Ecology 94*, 2097–2107.

Knape, J. and P. De Valpine (2012). Fitting complex population models by combining particle filters with Markov chain Monte Carlo. *Ecology 93*, 256–263.

Knape, J., N. Jonzén, and M. Sköld (2011). On observation distributions for state space models of population survey data. *Journal of Animal Ecology 80*, 1269–1277.

Kristensen, K., A. Nielsen, C. W. Berg, H. Skaug, and B. Bell (2016). TMB: automatic differentiation and Laplace approximation. *Journal of Statistical Software 70*, 1–21.

Langrock, R. (2011). Some applications of nonlinear and non-Gaussian state–space modelling by means of hidden Markov models. *Journal of Applied Statistics 38*, 2955–2970.

Langrock, R. and R. King (2013). Maximum likelihood estimation of mark–recapture–recovery models in the presence of continuous covariates. *Annals of Applied Statistics 7*, 1709–1732.

Lele, S. R. (2020). Consequences of lack of parameterization invariance of non-informative Bayesian analysis for wildlife management: Survival of San Joaquin kit fox and declines in amphibian populations. *Frontiers in Ecology and Evolution 7*, 501.

Lele, S. R., B. Dennis, and F. Lutscher (2007). Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters 10*, 551–563.

Lele, S. R., K. Nadeem, and B. Schmuland (2010). Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association 105*, 1617–1625.

Liu, J. and M. West (2001). Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo Methods in Practice*, pp. 197–223. Springer.

Lunn, D., C. Jackson, N. Best, A. Thomas, and D. Spiegelhalter (2012). *The BUGS Book: A Practical Introduction to Bayesian Analysis*. CRC Press/Chapman and Hall.

Martin, G. M., B. P. M. McCabe, D. T. Frazier, W. Maneesoonthorn, and C. P. Robert (2019). Auxiliary likelihood-based approximate Bayesian computation in state space models. *Journal of Computational and Graphical Statistics 28*, 508–522.

Martino, L., V. Elvira, and F. Louzada (2017). Effective sample size for importance sampling based on discrepancy measures. *Signal Processing 131*, 386–401.

Martino, L., V. Elvira, J. Míguez, A. Artés-Rodríguez, and P. Djurić (2018). A comparison of clipping strategies for importance sampling. In *2018 IEEE Statistical Signal Processing Workshop*, pp. 558–562.

Martino, L., J. Read, V. Elvira, and F. Louzada (2017). Cooperative parallel particle filters for online model selection and applications to urban mobility. *Digital Signal Processing 60*, 172 – 185.

McClintock, B., R. King, L. Thomas, J. Matthiopoulos, B. McConnell, and J. Morales (2012). A general discrete-time modeling framework for animal movement and migration using multi-state random walks. *Ecological Monographs 82*, 335–349.

McClintock, B. T., R. Langrock, O. Gimenez, E. Cam, D. Borchers, R. Glennie, and T. A. Patterson (2021). Uncovering ecological state dynamics with hidden Markov models. *Ecology Letters 23*, 1878–1903.

McClintock, B. T., D. J. F. Russell, J. Matthiopoulos, and R. King (2013). Combining individual animal movement and ancillary biotelemetry data to investigate population-level activity budgets. *Ecology 94*, 838–849.

Michaud, N., P. de Valpine, D. Turek, C. J. Paciorek, and D. Nguyen (2021). Sequential Monte Carlo methods in the Nimble R package. *Journal of Statistical Software*. in press.

Millar, R. B. and R. Meyer (2000). Non-linear state space modelling of fisheries biomass dynamics by using Metropolis-Hastings within-Gibbs sampling. *Journal of the Royal Statistical Society: Series C 49*, 327–342.

Monnahan, C. and K. Kristensen (2018). No-U-turn sampling for fast Bayesian inference in ADMB and TMB: Introducing the `adnuts` and `tmbstan` `R` packages. *PloS one 13*(5), e0197954.

Monnahan, C. C., J. T. Thorson, and T. A. Branch (2017). Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. *Methods in Ecology and Evolution 8*, 339–348.

Murray, L. M. (2015). Bayesian state-space modelling on high-performance hardware using LibBi. *Journal of Statistical Software 67*, 1–36.

Naesseth, C., S. Linderman, R. Ranganath, and D. Blei (2018). Variational sequential Monte Carlo. In *International Conference on Artificial Intelligence and Statistics*, pp. 968–977.

Neal, R. (2011). MCMC using Hamiltonian dynamics. In S. Brooks, A. Gelman, G. Jones, and X. Meng (Eds.), *Handbook of Markov Chain Monte Carlo: Methods and Applications*, pp. 113–162. CRC Press.

Newman, K. B., S. T. Buckland, S. T. Lindley, L. Thomas, and C. Fernandez (2006). Hidden process models for animal population dynamics. *Ecological Applications 16*, 74–86.

Newman, K. B., S. T. Buckland, B. J. T. Morgan, R. King, D. L. Borchers, D. J. Cole, P. Besbeas, O. Gimenez, and L. Thomas (2014). *Modelling Population Dynamics: Model Formulation, Fitting and Assessment using State-space Methods*. Springer: New York.

Newman, K. B., C. Fernández, L. Thomas, and S. T. Buckland (2009). Monte Carlo inference for state–space models of wild animal populations. *Biometrics 65*, 572–583.

Osada, Y., T. Kuriyama, M. Asada, H. Yokomizo, and T. Miyashita (2019). Estimating range expansion of wildlife in heterogeneous landscapes: A spatially explicit state-space matrix model coupled with an improved numerical integration technique. *Ecology and Evolution 9*, 318–327.

Patterson, T. A., A. Parton, R. Langrock, P. G. Blackwell, L. Thomas, and R. King (2017). Statistical modelling of individual animal movement: An overview of key methods and a discussion of practical challenges. *AStA Advances in Statistical Analysis 101*, 399–438.

Pedersen, M. W., T. A. Patterson, U. H. Thygesen, and H. Madsen (2011). Estimating animal behavior and residency from movement data. *Oikos 120*, 1281–1290.

Picchini, U. and A. Samson (2018). Coupling stochastic EM and approximate Bayesian computation for parameter inference in state-space models. *Computational Statistics 33*, 179–212.

Pitt, M. K. and N. Shephard (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association 94*, 590–599.

Plummer, M. (2016). *rjags: Bayesian Graphical Models using MCMC*. R package version 4-6.

Polansky, L., K. Newman, and L. Mitchell (2021). Improving inference for nonlinear state-space models of animal population dynamics given biased sequential life stage data. *Biometrics 77*, 352–361.

Ponciano, J. M., M. L. Taper, B. Dennis, and S. R. Lele (2009). Hierarchical models in ecology: confidence intervals, hypothesis testing, and model selection using data cloning. *Ecology 90*, 356–362.

Reich, B. J. and S. K. Ghosh (2019). *Bayesian Statistical Methods*. CRC Press.

Ruiz-Suarez, S., V. Leos-Barajas, I. Alvarez-Castro, and J. Morales (2020). Using approximate Bayesian inference for a "steps and turns" continuous-time random walk observed at regular time intervals. *PeerJ 8*, e8452.

Sarkka, S. (2013). *Bayesian Filtering and Smoothing* (3 ed.). Cambridge University Press.

Schlägel, U. E. and M. A. Lewis (2014). Detecting effects of spatial memory and dynamic information on animal movement decisions. *Methods in Ecology and Evolution 5*, 1236–1246.

Scranton, K., J. Knape, and P. de Valpine (2014). An approximate Bayesian computation approach to parameter estimation in a stochastic stage-structured population model. *Ecology 95*, 1418–1428.

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

Skaug, H. J. (2002). Automatic differentiation to facilitate maximum likelihood estimation in nonlinear random effects models. *Journal of Computational and Graphical Statistics 11*, 458–470.

Sólymos, P. (2010). dclone: Data cloning in R. *The R Journal 2*, 29–37.

Stan Development Team (2021). *Stan Modeling Language Users Guide and Reference Manual*. Version 2.27.

Staples, D. F., M. L. Taper, B. Dennis, and R. J. Boik (2009). Effects of sampling error and temporal correlations in population growth on process variance estimators. *Environmental and ecological statistics 16*, 547–560.

Stoklosa, J., C. Daly, S. D. Foster, M. B. Ashcroft, and D. I. Warton (2015). A climate of uncertainty: accounting for error in climate variables for species distribution models. *Methods in Ecology and Evolution 6*, 412–423.

Thomas, L., S. T. Buckland, K. B. Newman, and J. Harwood (2005). A unified framework for modelling wildlife population dynamics. *Australian & New Zealand Journal of Statistics 47*, 19–34.

Thygesen, U. H., C. M. Albertsen, C. W. Berg, K. Kristensen, and A. Nielsen (2017). Validation of ecological state space models using the Laplace approximation. *Environmental and Ecological Statistics 24*, 317–339.

Urteaga, I., M. F. Bugallo, and P. M. Djurić (2016). Sequential Monte Carlo methods under model uncertainty. In *2016 IEEE Statistical Signal Processing Workshop*, pp. 1–5.

van de Schoot, R., S. Depaoli, R. King, B. Kramer, K. Märtens, M. G. Tadesse, M. Vannucci, A. Gelman, D. Veen, J. Willemsen, and C. Yau (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers 1*, 1–26.

Wan, E. A., R. Van der Merwe, and S. Haykin (2001). The unscented Kalman filter. *Kalman filtering and neural networks 5*, 221–280.

Wang, G. (2007). On the latent state estimation of nonlinear population dynamics using Bayesian and non-Bayesian state-space models. *Ecological Modelling 200*, 521–528.

Watanabe, S. (2013). A widely applicable Bayesian information criterion. *Journal of Machine Learning Research 14*, 867–897.

Watanabe, S. (2021). Information criteria and cross validation for Bayesian inference in regular and singular cases. *Japanese Journal of Statistics and Data Science*, 1–19.

Watanabe, S. and M. Opper (2010). Asymptotic equivalence of Bayes cross validation and

widely applicable information criterion in singular learning theory. *Journal of machine learning research 11*.

Welch, G. and G. Bishop (1995). *An Introduction to the Kalman Filter*. Chapel Hill.

Wikle, C. K., A. Zammit-Mangion, and N. Cressie (2019). *Spatio-temporal Statistics with R*. Chapman and Hall/CRC.

Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature 466*, 1102–1104.

Zucchini, W., I. L. MacDonald, and R. Langrock (2016). *Hidden Markov Models for Time Series An Introduction Using R, Second Edition*. Chapman & Hall/CRC.