Semi-Complete Data Augmentation for Efficient State Space Model Fitting

Agnieszka Borowska
University of Glasgow, School of Mathematics and Statistics, UK
and
Ruth King
School of Mathematics, University of Edinburgh, UK

Abstract

We propose a novel efficient model-fitting algorithm for state space models. State space models are an intuitive and flexible class of models, frequently used due to the combination of their natural separation of the different mechanisms acting on the system of interest: the latent underlying system process; and the observation process. This flexibility, however, often comes at the price of more complicated model-fitting algorithms due to the associated analytically intractable likelihood. For the general case a Bayesian data augmentation approach is often employed, where the true unknown states are treated as auxiliary variables and imputed within the MCMC algorithm. However, standard "vanilla" MCMC algorithms may perform very poorly due to high correlation between the imputed states and/or parameters. The proposed method addresses the inefficiencies of traditional approaches by combining data augmentation with numerical integration in a Bayesian hybrid approach. This approach permits the use of standard "vanilla" updating algorithms that perform considerably better than the traditional approach in terms of improved mixing and lower autocorrelation. We apply our semi-complete data augmentation algorithm to different application areas and models, leading to distinct implementation schemes and demonstrating improved efficiency.

Keywords: Bayesian inference; data augmentation; effective sample size; Markov chain Monte Carlo; numerical integration.

1 Introduction

Inference about a latent state governing the dynamics of the system under study given only the observed noisy data is of interest in many contexts, e.g. in applied statistics, ecology, engineering or economics. A very intuitive way of describing such problems is provided by latent process models, also known as state space models (SSM), see Durbin and Koopman (2012) and West and Harrison (1997). Such models are frequently used due to the combination of their natural separation of the different mechanisms acting on the system of interest: the (unobserved) underlying system process; and the observation process. Considering each distinct process separately simplifies the model specification process and provides a very flexible modeling framework. This flexibility, however, typically comes at the price of substantially more complicated fitting of such models to data as for the general non-linear non-Gaussian SSM the associated likelihood is analytically intractable. Only in certain circumstances the associated likelihood can be calculated explicitly: for linear Gaussian systems the likelihood can be obtained by applying the Kalman filter; for hidden Markov models with a discrete state space the likelihood may admit a closed-form but may become unfeasible for a large number of states. In this paper we focus on models for which the likelihood is intractable or for which it may be unfeasible to compute explicitly.

Dominant approaches to intractable likelihood problems include: (i) numerical or Monte Carlo integration to estimate the observed (or marginal) data likelihood; and (ii) data augmentation (DA), based on the complete (or joint) data likelihood of the observed and the imputed unobserved states, see Tanner and Wong (1987). Group (i) includes the sequential Monte Carlo (SMC) methods, see Doucet et al. (2001), which can be used for parameter estimation within a standard Markov chain Monte Carlo (MCMC) algorithm (i.e. particle MCMC, Andrieu et al., 2010). In general, numerical integration is efficient for low dimensional systems. DA approach (ii) has become standard for inference for SSMs within a Bayesian framework, see Frühwirth-Schnatter (1994, 2004). DA treats the true unknown states as auxiliary variables and imputes them within an MCMC algorithm. However, "vanilla" MCMC methods may perform very poorly due to high correlation between the

imputed states and/or parameters, see Hobert et al. (2011) and the references therein.

We propose a novel model-fitting algorithm to circumvent these inefficiencies by combining DA with numerical integration in a Bayesian hybrid approach, where the associated standard "vanilla" algorithms perform substantially more efficiently. The underlying idea is to combine the "good" aspects of both methods by minimizing the problems that arise for each, i.e. highly correlated latent states for DA and the curse of dimensionality for numerical integration. To this end, we utilize the structure of the unknown states and split them into two types: auxiliary variables, imputed within the MCMC algorithm using DA; and "integrable" states, numerically integrated out within the likelihood expression.

The structure of the paper is as follows. Section 2 presents the general SSM specification and the standard approaches to their fitting. Section 3 introduces the proposed semi-complete data augmentation approach, while Section 4 develops a general HMM-based likelihood approximation to the associated likelihood. We demonstrate the efficiency gains from our method in Section 5, based on two empirical applications related to the stochastic volatility model and abundance estimation. Section 6 concludes with a discussion.

2 State space models

Consider a state space model given by $(t = 1, \ldots, T)$

$$\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta} \sim p(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta}), \tag{1}$$

$$\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{\theta} \sim p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{\theta}), \tag{2}$$

$$\boldsymbol{x}_0|\boldsymbol{\theta} \sim p(\boldsymbol{x}_0|\boldsymbol{\theta}). \tag{3}$$

Let $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)$ denote a time series of observations (potentially multivariate) of length T, $\boldsymbol{x} = (\boldsymbol{x}_0, \ldots, \boldsymbol{x}_T)$ a series of latent states (with $\boldsymbol{x}_t = (x_{1,t}, \ldots, x_{D,t})^T$ potentially multivariate, of dimension $D < \infty$, with $x_{d,t} \in \mathcal{X}_d$) and $\boldsymbol{\theta}$ the model parameters for which we put a prior $p(\boldsymbol{\theta})$. To simplify notation, we use p as a general symbol for a probability mass function (pmf) or a probability density function (pdf), possibly conditional.

The system process describing the evolution of $\boldsymbol{x}_t$, the true (unobserved) state of the system over time is defined by distribution (2). The observation process which generates

$\boldsymbol{y}_t$, the observed data given the true underlying states, is specified by distribution (1). This separation of the different mechanisms acting on the system of interest makes SSM a very intuitive and flexible description of time series data. Figure 1 graphically presents the dependencies between states and observations in the SSM. An extensive discussion of SSMs is provided by Durbin and Koopman (2012) and Cappé et al. (2006), where this class of models is called hidden Markov models (HMM).

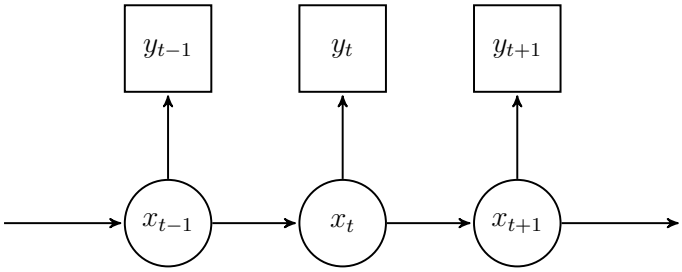


Figure 1: A graphical representation of the general first-order SSM.

Modeling flexibility of SSMs is, however, often offset with the issue of estimating $\boldsymbol{\theta}$, the associated model parameters. The *observed data likelihood* for the system (1)–(3)

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \int p(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta}) d\boldsymbol{x} = \int p(x_0|\boldsymbol{\theta}) \prod_{t=1}^{T} p(y_t|x_t, \boldsymbol{\theta}) p(x_t|x_{t-1}, \boldsymbol{\theta}) dx_0 dx_1 \dots dx_T, \quad (4)$$

is typically not available in closed form due to the necessary integration over the latent variables. This is despite the tractability of $p(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})$, the joint distribution of the data and the auxiliary variables, often referred to as the *complete data likelihood*.

For models with discrete states the observed data likelihood is the likelihood of an HMM, where the states of the chain correspond to distinct values of the latent process, and the transition matrix can be derived from the transition equation (2). This likelihood can be efficiently calculated using the forward algorithm (see Zucchini et al., 2016). However, for systems with multiple processes or processes with a large set of possible states this can lead to this approach being unfeasible.

To overcome the problem of the intractable likelihood, a DA technique is commonly adopted, see Tanner and Wong (1987); Frühwirth-Schnatter (1994, 2004); Hobert (2011). The unknown states $\boldsymbol{x}$ are treated as auxiliary variables and imputed leading to a closed-form complete data likelihood (5) which in a Bayesian framework is used to construct the

joint posterior distribution of $\boldsymbol{\theta}$ and $\boldsymbol{x}$ in (6):

$$p(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta}) = p(x_0|\boldsymbol{\theta}) \prod_{t=1}^{T} p(y_t|x_t, \boldsymbol{\theta}) p(x_t|x_{t-1}, \boldsymbol{\theta}), \tag{5}$$

$$p(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}) p(\boldsymbol{x}|\boldsymbol{\theta}) p(\boldsymbol{\theta}). \tag{6}$$

An MCMC algorithm (or other) can be used to obtain a sample from (6), from which we obtain $p(\boldsymbol{\theta}|\boldsymbol{y})$, the marginal posterior of $\boldsymbol{\theta}$. In practice the random walk Metropolis-Hastings (RW-MH) algorithm is often used and it acts as a "vanilla" MCMC algorithm (see Marin and Robert, 2007, Ch. 4).

However, this approach often results in posterior draws being highly correlated, indicating poor mixing and hence low efficiency of MCMC algorithms. This is particularly the case for SSMs which impose a strong dependence structure on the latent variables and parameters. Single-update algorithms can perform especially poorly and block updates can lead to improved mixing. However, the latter often require defining an appropriate partition of the states and parameters into blocks and specifying an efficient proposal distributions for each block. Thus, bespoke codes often need to be written dependent on model and data.

3 Semi-complete data augmentation

We propose to combine DA with numerical integration within a Bayesian hybrid framework, which we call *semi-complete data augmentation*. A key idea is to separate the latent state $\boldsymbol{x}$ into two components $\boldsymbol{x} = (\boldsymbol{x}_{aug}^T, \boldsymbol{x}_{int}^T)^T$. We will refer to $\boldsymbol{x}_{int}$ and $\boldsymbol{x}_{aug}$ as the "integrated" states and the "augmented" states, respectively. We specify the *semi-complete data likelihood* (SCDL) $p(\boldsymbol{y}, \boldsymbol{x}_{aug}|\boldsymbol{\theta})$ as follows

$$p(\boldsymbol{y}, \boldsymbol{x}_{aug}|\boldsymbol{\theta}) = \int p(\boldsymbol{y}|\boldsymbol{x}_{aug}, \boldsymbol{x}_{int}, \boldsymbol{\theta}) p(\boldsymbol{x}_{aug}, \boldsymbol{x}_{int}|\boldsymbol{\theta}) d\boldsymbol{x}_{int}. \tag{7}$$

The joint posterior distribution of the parameters and augmented states is given by

$$p(\boldsymbol{\theta}, \boldsymbol{x}_{aug}|\boldsymbol{y}) \propto p(\boldsymbol{y}, \boldsymbol{x}_{aug}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{x}_{aug}, \boldsymbol{\theta}) p(\boldsymbol{x}_{aug}|\boldsymbol{\theta}) p(\boldsymbol{\theta}). \tag{8}$$

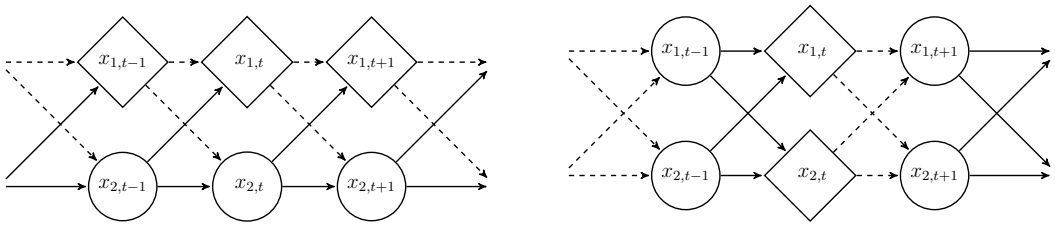
We note that the approach of King et al. (2016), who propose a Bayesian hybrid approach for the particular case of capture-recapture data, is a special case of our general approach proposed here.

Specification of $\boldsymbol{x}_{aug}$, $\boldsymbol{x}_{int}$ Consider a series of latent states $\boldsymbol{x} = \{\boldsymbol{x}_t\}_{t=0}^{T}$ of length $T+1$, where the state at time t is D dimensional: $\boldsymbol{x}_t = (x_{1,t}, \ldots, x_{D,t})^T$, for $t = 0, 1, \ldots, T$. We want to integrate out D_{int} dimensions of the state at time points T_{int}, where $D_{int} \subset \{1, \ldots, D\}$ and $T_{int} \subset \{0, 1, \ldots, T\}$ are "suitably" chosen subsets of dimension and time indices, respectively. Such a "suitable" specification of D_{int} and T_{int} depends on the dependence structure of the model so that the associated integral can be efficiently calculated. For instance, it can be low dimensional or it can be reduced to a product of low-dimensional integrals. We denote the compliments of both subsets D_{aug} and T_{aug}, respectively. We also let T_{int}^{+} and T_{aug}^{+} denote the corresponding sets without the initial observations, i.e. excluding time $t = 0$, and we set $T^* = |T_{int}^{+}|$. The "integrated" and "augmented" states are then defined as the partition of $\boldsymbol{x}$ into $\boldsymbol{x}_{int} = \{x_{d,t}\}_{d \in D_{int}, t \in T_{int}}$ and $\boldsymbol{x}_{aug} = \{x_{d,t}\}_{d \in D_{aug}, t \in T_{aug}}$, where their corresponding elements at time t are denoted by $\boldsymbol{x}_{int,t} = \{x_{d,t}\}_{d \in D_{int}}$ and $\boldsymbol{x}_{aug,t} = \{x_{aug,t}\}_{d \in D_{aug}}$, respectively. As example, consider the two following schemes.

(a) *"Horizontal" integration:* e.g. for a $D = 2$ dimensional state we integrate out the second state at all time periods, so that $D_{int} = \{2\}$ (and hence $D_{aug} = \{1\}$), and $T_{int} = \{0, 1, \ldots, T\}$ (and hence $T_{aug} = T_{int}$), see Figure 2a. We use this scheme in the lapwings data application in Section 5.2.

(b) *"Vertical" integration:* e.g. all D states are integrated out at odd time periods, so that $D_{int} = \{1, \ldots, D\}$ and $T_{int} = \{2t+1\}_{t=0}^{\lfloor T/2 \rfloor}$ (and hence $T_{aug} = \{2t\}_{t=0}^{\lfloor T/2 \rfloor}$), see Figure 2b. We use this scheme in the stochastic volatility (SV) model application in Section 5.1, for $D = 1$ dimensional state.

As we can see, in general T_{int} and T_{aug} do not need to be equal and their elements may not be consecutive numbers. However, we would like to iterate over both sets using the same index. Therefore, we introduce two functions $\tau(t)$ and $a(t)$ such that the image of τ is T_{int}^{+} and the image of a covers T_{aug}^{+}, both defined on $1, 2, \ldots, T^*$. We require τ to be bijective and allow a to take values in the power set of T_{aug}^{+}. The latter characteristic means that $a(t)$ can take two or more values in T_{aug}^{+} but also no value (i.e. $a(t) = \emptyset$). In the two examples above we have $\tau(t) = t$ and $a(t) = t$ for the horizontal integration (a) and $\tau(t) = 2t + 1$ and $a(t) = 2t$ for the vertical integration (b). Additionally, we specify a



(a) **Horizontal integration**. (b) **Vertical integration**

Figure 2: Two examples of an integration/augmentation scheme. Diamonds represent the imputed states, circles – the integrated states. Dashed lines used for the relations *from* the imputed (known) states.

function for observations $o(t)$ with a similar role to τ and a, i.e. allowing us to iterate over the set of observation indices $\{1, \dots, T\}$ using the same index as to iterate over T_{int} and T_{aug}. Thus, we want the image of $o(t)$ to be $\{1, \dots, T\}$, which may consists of elements from both T_{int} and T_{aug}. This means that we need to be able to assign multiple indices from $\{1, \dots, T\}$ to t. Hence, we allow $o(t)$ to take values in the power set of $T_{int} \cup T_{aug}$. For illustration, consider vertical integration (b) together with conditionally independent observations $y_t|\boldsymbol{x}_t \sim p(y_t|\boldsymbol{x}_t)$. For $t = 1, 2, \dots, T^*$ consider states in two different time periods: at $\tau(t) = 2t+1$ for $\boldsymbol{x}_{int}$ and at $a(t) = 2t$ for $\boldsymbol{x}_{aug}$, so for each t we need to account for two different observations, $y_{\tau(t)}$ and $y_{a(t)}$. This means that $o(t) = \{2t, 2t+1\}$ for $t \geq 1$. In this case we also need to account for y_1 so we additionally specify $o(t) = \{2t+1\}$ for $t = 0$. For horizontal integration given in (a) $T_{int} = T_{aug}$, hence $o(t) = t$.

In order to identify conditionally independent latent states to "integrate out", one can use the graphical structure of the problem: Figure 1 can be seen as an directed acyclic graph (DAG), for which the literature on dynamic Bayesian networks (see Murphy, 2002) provides insights regarding the impact of conditioning on a certain node (*d-separation*). In the context of particle filters Doucet et al. (2000) note that the "tractable structure" of some state space models might by analytically marginalized out given imputed other nodes.

Approximate marginal likelihood The SCDL $p(\boldsymbol{y}, \boldsymbol{x}_{aug}|\boldsymbol{\theta})$ in the joint posterior distribution of $\boldsymbol{\theta}$ and $\boldsymbol{x}_{aug}$ in (8) may still be analytically intractable. In this case we can estimate it using simulation-based techniques. Consider a sample of length N of unknown variables of interest (i.e. $\boldsymbol{\theta}$ and $\boldsymbol{x}_{aug}$). Here, N is the number of points used for integration: for a deterministic integration it is the number of evaluation points, for a stochastic, i.e. Monte Carlo (MC), integration it is the number of draws. We use such a sample to compute $\hat{p}_N(\boldsymbol{y}, \boldsymbol{x}_{aug}|\boldsymbol{\theta})$, the N-sample estimator of the SCDL, and consequently to approximate the posterior distribution as $\hat{p}_N(\boldsymbol{\theta}, \boldsymbol{x}_{aug}|\boldsymbol{y}) \propto \hat{p}_N(\boldsymbol{y}, \boldsymbol{x}_{aug}|\boldsymbol{\theta})p(\boldsymbol{\theta})$. We set $\hat{p}_N(\boldsymbol{y}, \boldsymbol{x}_{aug}|\boldsymbol{\theta})$ such that $\hat{p}_N(\boldsymbol{y}, \boldsymbol{x}_{aug}|\boldsymbol{\theta}) \overset{N\to\infty}{\to} p(\boldsymbol{y}, \boldsymbol{x}_{aug}|\boldsymbol{\theta})$, so that $\hat{p}_N(\boldsymbol{\theta}, \boldsymbol{x}_{aug}|\boldsymbol{y}) \overset{N\to\infty}{\to} p(\boldsymbol{\theta}, \boldsymbol{x}_{aug}|\boldsymbol{y})$. Further properties of the resulting estimator depend on the approximation scheme. If it is unbiased and non-negative, standard MCMC algorithms converge to the exact posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{x}_{aug}|\boldsymbol{y})$, which follows from the pseudo-marginal argument, see Beaumont (2003), Andrieu and Roberts (2009) and Andrieu et al. (2010). Pseudo-marginal algorithms are called "exact approximate" and we note that they are the extreme case of our approach with $\boldsymbol{x}_{int} = \boldsymbol{x}$. Whether our approximate MCMC algorithm is "exact approximate" or "just approximate" depends on whether or not $\hat{p}_N(\boldsymbol{y}, \boldsymbol{x}_{aug}|\boldsymbol{\theta})$ is an unbiased and non-negative estimator of the marginal likelihood.

The "just approximate" algorithms, such as a quadrature, can be made arbitrarily close to the true integral by considering sufficiently many points (i.e. as $N \to \infty$). Alternatively, unbiased estimators using an MC approach might be characterized by large MC errors, particularly for a small number of samples, see e.g. Korattikara et al. (2014), Jacob and Thiery (2015). The choice between different likelihood approximation methods fits into the traditional discussion on the bias-variance trade-off.

4 Approximations for MCMC sampling

We focus on the case when $\hat{p}_N(\boldsymbol{y}, \boldsymbol{x}_{aug}|\boldsymbol{\theta})$ can be obtained as a product of one dimensional integrals. This assumption is less restrictive than it may appear at first: the choice of the auxiliary variables can often be made such that this condition is satisfied. There exist several methods to numerically estimate a single one dimensional integral including: (1)

quadrature with fixed nodes; (2) quadrature with adaptive nodes; (3) stochastic (MC) integration. Approaches (1) and (2) can be seen as "binning" of similar values of the integrated state vector within specified ranges ("bins"), which can then be interpreted as states of a (finite-dimensional) first-order HMM. In the context of bins of equal widths such an approach has been successfully applied e.g. by Langrock et al. (2012a) and Langrock et al. (2012b); Langrock and King (2013). For approach (3) the resulting estimator of the complete data likelihood is unbiased and an "exact approximate" algorithm is obtained. We note that in low dimensions all these methods are feasible, however we focus on methods based on the two former approaches as they provide an intuitive interpretation in terms of state transition probabilities and conditional (augmented) observation distributions. There are two cases when such an approximation might be particularly useful. First, when the state vector is discrete but of a large size grouping of its elements into "bins" helps to reduce the size of the problem. Second, for continuous states any form of numerical integration basically reduces to splitting of the state space into "bins", which can then be further combined into larger groups to increase the efficiency of the algorithm.

4.1 Approximation bins as hidden Markov model states

We consider two ways to specify the bins, or quadrature points: a deterministic one, with bins of a fixed size (but varying probability of occurring), and a stochastic one, with bins of a fixed probability (but varying size). To simplify the exposition, we assume that $\boldsymbol{x}_{int,\tau(t)}$ is univariate and we write $x_{int,\tau(t)}$. For multivariate $\boldsymbol{x}_{int,\tau(t)}$ we may consider separate bins for each integrated state dimension $d \in D_{int}$ at time $\tau(t)$. We interpret the bins as states of a latent (first-order) Markov process, which allows us to give the resulting integration/augmentation scheme an HMM embedding.

Fixed bins A straightforward approach to binning is via bins of a fixed size as it relates to a deterministic approximation of the likelihood with a quadrature and allows for a natural HMM interpretation. Discretizing the state space to perform numerical integration dates back to Kitagawa (1987) and is discussed in Zucchini et al. (2016). The state space $\mathcal{X}_{int}$ of

the state to be integrated out is split into B bins of length k (for integer-valued variables we assume $k \in \mathbb{N}$) and e.g. the midpoints of the bins are considered for integration. Then the values that fall in a given bin are approximated by the value of the midpoint of that bin. Such an approach is used by Langrock et al. (2012b) to efficiently approximate the likelihood for stochastic volatility models (with continuous bins) in a classical framework.

For infinitely dimensional states, either discrete or continuous, an "allowed integration range" needs to be specified. For instance, for a normal variable this means setting a lower and an upper bound for the integration b_0 and b_B, while for a Poisson variable only of an upper bound b_B since $b_0 = 0$ in this case. We divide the resulting domain into intervals:

$$\underbrace{[b_0, \dots, b_1)}_{\mathcal{B}_1,\text{ bin } 1}, \underbrace{[b_1, \dots, b_2)}_{\mathcal{B}_2,\text{ bin } 2}, \dots, \underbrace{[b_{j-1}, \dots, b_j)}_{\mathcal{B}_j,\text{ bin } j}, \dots, \underbrace{[b_{B-1}, \cdots, b_B)}_{\mathcal{B}_B,\text{ bin } B}, \quad b_i - b_{i-1} = k, \quad i = 1, \dots, B.$$

For continuous variables $\mathcal{B}_i$ is simply a continuous interval of length k, while for discrete variables it consists of k subsequent integers, e.g. for a Poisson variable we have $\mathcal{B}_i = \{ik, \dots, (i+1)k\}$. We specify the midpoints of the bins as $b_i^* = \frac{b_{i-1}+b_i}{2}$ (for integer-valued variables rounding is required for even k).

We define $\{z_t\}$, $t \in 1, \dots, T^*$, as a B-state, discrete-time (not necessarily homogeneous) Markov chain with transition probabilities $\gamma_{jk,t} = \mathbb{P}(z_t = k | z_{t-1} = j)$ defined as

$$\gamma_{jk,t} := \mathbb{P}(x_{int,\tau(t)} \in \mathcal{B}_k | x_{int,\tau(t-1)} \in \mathcal{B}_j, \boldsymbol{x}_{aug,a(t-1)}).$$

The transition of $z_{t-1} = j$ to $z_t = k$ is equivalent to $x_{int,\tau(t)}$ belonging to bin k given $x_{int,\tau(t-1)}$ was in bin j (and $\boldsymbol{x}_{aug,a(t-1)}$). For computationally intensive probabilities we can further approximate these as $\tilde{\gamma}^*_{jk,t} := p(b_k^* | b_j^*, \boldsymbol{x}_{aug,a(t-1)})$, which for discrete variables means $\mathbb{P}(x_{int,\tau(t)} = b_k^* | x_{int,\tau(t-1)} = b_j^*, \boldsymbol{x}_{aug,a(t-1)})$. To get the valid probability values we normalize the transition probabilities as $\gamma^*_{jk,t} := \tilde{\gamma}^*_{jk,t} / \sum_{c=1}^{B} \tilde{\gamma}^*_{jc,t}$. Notice that this corresponds to treating the values in a bin uniformly. Alternatively, we can compute the transition probabilities between bins directly as follows

$$\mathbb{P}(x_{int,\tau(t)} \in \mathcal{B}_k | x_{int,\tau(t-1)} \in \mathcal{B}_j, \boldsymbol{x}_{aug,a(t-1)}) \propto \int_{\mathcal{B}_k \times \mathcal{B}_j} p(x_{int,\tau(t)} | x_{int,\tau(t-1)}, \boldsymbol{x}_{aug,a(t-1)}) dx_{int,\tau(t-1)} dx_{int,\tau(t)}.$$

However, such an analytical integration will typically be possible only in simple cases.

Adaptive bins An alternative approach is to use adaptive intervals which do not require a bounded integration range. This can be done by transforming the variable of interest to the $[0, 1]$ range by applying a cdf. The bins are specified on the $[0, 1]$ interval and their limits or midpoints are transformed back to obtain the values of the original variable. In particular, quantiles of the distribution associated with the variable of interest can be used.

Suppose $x_{int,\tau(t)} \sim p(\vartheta_{\tau(t)})$, $\tau(t) \in T_{int}$, where $\vartheta_{\tau(t)}$ is a vector of possibly time varying parameters, with the corresponding cdf $F(\vartheta_{\tau(t)})$. Consider a vector of $B+1$ quantiles $\boldsymbol{q} = [q_0, q_2, \ldots, q_B]$. The corresponding B mid-quantiles $\boldsymbol{q}^* = [q_1^*, q_2^*, \ldots, q_B^*]$ are given by $q_i^* = \frac{q_{i-1}+q_i}{2}$ (e.g. $\boldsymbol{q} = [0.0, 0.1, 0.2, \ldots, 1.0]$ and $\boldsymbol{q}^* = [0.05, 0.15, \ldots, 0.95]$). For $F(\vartheta_t)$ continuous and strictly monotonically increasing the bin midpoints at time t are determined by the mid-quantiles as $b_i^* = F^{-1}\left(q_i^* | \vartheta_{\tau(t)}\right)$. For discrete variables one can either use the generalized inverse distribution function, or use a continuous approximation to the associated discrete distribution.

4.2 Hidden Markov model likelihood

Having specified the states of the underlying Markov chain in the previous section, we aim to use them to approximate the joint SCDL (7) by embedding it into an HMM form (below, to ease the notation, we skip $\boldsymbol{\theta}$ in conditioning). We relate each state of the hidden Markov process with the relevant augmented states and observations. This imposes a time structure on the SCDL integral with respect to the "integration time" and thus allows us to cast it into a likelihood of an HMM.

Motivating example Consider the state specification from Figure 2a to which we add conditionally independent observations to result in an SSM (see Online Appendix A.1 for a graphical illustration and more details). We specify $\boldsymbol{x}_{aug} = \{x_{1,t}\}_{t=0}^T =: \boldsymbol{x}_1$ and $\boldsymbol{x}_{int} = \{x_{2,t}\}_{t=0}^T =: \boldsymbol{x}_2$, which corresponds to the "horizontal" integration. Hence, we put $T_{int} = T_{aug} = \{0, 1, \ldots, T\}$, $\tau(t) = t$, $a(t) = t$ and $o(t) = t$. Using the temporal dependence

in this system, the SCDL $p(\boldsymbol{y}, \boldsymbol{x}_{aug})$ can be expressed as

$$p(\boldsymbol{y}, \boldsymbol{x}_{aug}) = p(x_{1,0}) \prod_{t=1}^{T} p(y_t|x_{1,t}) p(x_{1,t}|x_{1,t-1}) = p(x_{1,0}) \prod_{t=1}^{T^*} p(y_{o(t)}|x_{1,a(t)}) p(x_{1,a(t)}|x_{1,a(t-1)}),$$

which is not tractable without integrating out $\boldsymbol{x}_2$. Hence, we marginalise over $\boldsymbol{x}_2$ and approximate the resulting integral using a quadrature with B bins $\mathcal{B}_k$, $k = 1, \ldots, B$, as

$$\begin{aligned} p(\boldsymbol{y}, \boldsymbol{x}_{aug}) = & \int \ldots \int p(x_{1,0}) p(x_{2,0}) \prod_{t=1}^{T^*} p(y_{o(t)}|x_{1,a(t)}, x_{2,\tau(t)}) \\ & \times p(x_{1,a(t)}|x_{1,a(t-1)}, x_{2,\tau(t-1)}) p(x_{2,\tau(t)}|x_{1,a(t-1)}, x_{2,\tau(t-1)}) dx_{2,\tau(T^*)} \ldots dx_{2,\tau(1)} \qquad (9) \\ \approx & \sum_{k_0=1}^{B} \cdots \sum_{k_{T^*}=1}^{B} p(x_{1,0}) p(x_{2,0} \in \mathcal{B}_{k_0}) \prod_{t=1}^{T^*} p(y_{o(t)}|x_{1,a(t)}, x_{2,\tau(t)} \in \mathcal{B}_{k_t}) \\ & \times p(x_{1,a(t)}|x_{1,a(t-1)}, x_{2,\tau(t-1)} \in \mathcal{B}_{k_{t-1}}) p(x_{2,\tau(t)} \in \mathcal{B}_{k_t}|x_{1,a(t-1)}, x_{2,\tau(t-1)} \in \mathcal{B}_{k_{t-1}}). \end{aligned}$$

The above approximation has a natural interpretation in terms of HMM by associating the events $x_{2,\tau(t)} \in \mathcal{B}_k$ with states of a hidden Markov process on B states. The transition matrix of this process is

$$\Gamma_t = \left[\mathbb{P}(x_{2,\tau(t)} \in \mathcal{B}_k | x_{1,a(t-1)}, x_{2,\tau(t-1)} \in \mathcal{B}_l) \right]_{k,l=1,\ldots,B}, \qquad (10)$$

for $t \in 1, 2, \ldots, T^*$. We specify two further matrices for the "augmented data": P_t for the augmented states $\boldsymbol{x}_{aug}$ and Q_t for the real observations $\boldsymbol{y}$, as follows

$$P_t = \text{diag} \left(p(x_{1,a(t)}|x_{1,a(t-1)}, x_{2,\tau(t-1)} \in \mathcal{B}_l) \right)_{l=1,\ldots,B}, \qquad (11)$$

$$Q_t = \text{diag} \left(p(y_{o(t)}|x_{1,a(t)}, x_{2,\tau(t)} \in \mathcal{B}_k) \right)_{k=1,\ldots,B}. \qquad (12)$$

This is different compared to standard HMMs in which only the matrix for $\boldsymbol{y}$ is used. Notice that the conditioning in (10) and (11) is with respect to the previous realizations of the states, whilst for the observations in (12) it is with respect to the current states. Finally, the quadrature based approximation to the SCDL (9) can be expressed as

$$\hat{p}_B(\boldsymbol{y}, \boldsymbol{x}_{aug}) = p(x_{1,0}) \boldsymbol{u}_0 \left(\prod_{t=1}^{T^*} P_t \Gamma_t Q_t \right) \mathbf{1}, \qquad (13)$$

where $\boldsymbol{u}_0 = (\mathbb{P}(x_{2,0} \in \mathcal{B}_1), \ldots, \mathbb{P}(x_{2,0} \in \mathcal{B}_B))$ is the initial distribution of the Markov chain.

General formulation The generic matrices of the HMM-based approximation are

$$\Gamma_t = \left[\mathbb{P}(\boldsymbol{x}_{int,\tau(t)} \in \mathcal{B}_k | \boldsymbol{x}_{int,\tau(t-1)} \in \mathcal{B}_l, \boldsymbol{x}_{aug,a(t-1)})\right]_{k,l=1,\ldots,B},$$
$$P_t = \text{diag}\left(p(\boldsymbol{x}_{aug,a(t)} | \boldsymbol{x}_{int,\tau(t-1)} \in \mathcal{B}_l), \boldsymbol{x}_{aug,a(t-1)})\right)_{l=1,\ldots,B},$$
$$Q_t = \text{diag}\left(p(\boldsymbol{y}_{o(t)} | \boldsymbol{x}_{int,\tau(t)} \in \mathcal{B}_k, \boldsymbol{x}_{aug,a(t)})\right)_{k=1,\ldots,B},$$

for $t \in 1, 2, \ldots, T^*$ and lead to the following form of the HMM approximation

$$\hat{p}_B(\boldsymbol{y}, \boldsymbol{x}_{aug}) = p(x_{1,0})\boldsymbol{u}_0 Q_0 \left(\prod_{t=1}^{T^*} P_t \Gamma_t Q_t\right) \mathbf{1}, \tag{14}$$

which differs from (13) by including $Q_0 := \text{diag}\left(p(y_{o(0)} | x_{int,0} \in \mathcal{B}_k^*)\right)_{k=1,\ldots,B}$, which allows for a dependence of some observations on the initial state of the Markov process. The SV model example in Online Appendix D.3 demonstrates the role of Q_0.

5 Applications

We assess the performance of the proposed SCDA method in two case studies with distinctively different features resulting in different integration schemes. The first application relates to the well-known stochastic volatility model (SV), which is a popular tool to model time-varying volatility especially for financial time series, see Taylor (1994) or Kim et al. (1998). The second application involves the dataset on the Northern lapwing (*Vanellus vanellus*), which has been extensively analyzed in statistical ecology, see Besbeas et al. (2002), Brooks et al. (2004) or King et al. (2008).

We are interested in comparing the performance of the standard DA approach with that of SCDA. For comparability, for each method we use a "vanilla" RW-MH (single-update) algorithm for estimation. We tune each sampler so that the acceptance rates for each element of $\boldsymbol{\theta}$ and the average acceptance rates for each of the imputed states are "reasonable", i.e. 20–40% (Gelman et al., 1996, Roberts and Rosenthal, 2001).

5.1 Financial model: stochastic volatility

The basic form of the SV model is given by

$$y_t|h_t, \boldsymbol{\theta} \sim \mathcal{N}\left(0, \exp(h_t)\right), \qquad t = 1, \ldots, T, \tag{15}$$

$$h_t|h_{t-1}, \boldsymbol{\theta} \sim \mathcal{N}\left(\mu + \phi(h_{t-1} - \mu), \sigma^2\right), \tag{16}$$

where $h_0|\boldsymbol{\theta} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{1-\phi^2}\right)$, $\boldsymbol{\theta} = (\mu, \phi, \sigma^2)^T$. We adopt standard priors (Kim et al., 1998):

$$\mu \sim \mathcal{N}(0, \sigma^2_{\mu 0}), \quad \frac{\phi + 1}{2} \sim \mathcal{B}(\alpha_{\phi 0}, \beta_{\phi 0}), \quad \sigma^2 \sim \mathcal{IG}(\alpha_{\sigma^2 0}, \beta_{\sigma^2 0}),$$

with $\sigma^2_{\mu 0} = 10$, $\alpha_{\phi 0} = 20$, $\beta_{\phi 0} = 1.5$, $\alpha_{\sigma^2 0} = 5/2$, $\beta_{\sigma^2 0} = 0.05/2$. The estimation of the SV model has been considered a challenging problem due to the intractable likelihood

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \int p(\boldsymbol{y}, \boldsymbol{h}) d\boldsymbol{h} = \int p(h_0) \prod_{t=1}^{T} p(y_t|h_t) p(h_t|h_{t-1}) dh_0 dh_1 \ldots dh_T. \tag{17}$$

Some of the previous approaches to tackle this issue include standard DA approach, in which the latent volatilities are imputed in an MCMC scheme, see Kim et al. (1998). The associated complete data likelihood admits a closed form $p(\boldsymbol{y}, \boldsymbol{h}|\boldsymbol{\theta}) = p(h_0) \prod_{t=1}^{T} p(y_t|h_t) p(h_t|h_{t-1})$. An alternative approach is provided by Fridman and Harris (1998) or Langrock et al. (2012b) who propose numerical integration of the latent states. In particular, Langrock et al. (2012b) approximate (17) using an HMM by discretizing the state space of $\boldsymbol{h}$. They perform a numerical integration of the latent states based on a grid of B equally sized bins $B_i = [b_{i-1}, b_i)$, $i = 1, \ldots, B$, with the corresponding midpoints b_i^*. The range of the admissible values for the demeaned volatility, b_0 and b_B, is set to $\pm 5\sigma_h$, where σ_h is the stationary standard deviation of the logvolatility process. This leads to an approximation of (17) via $p(\boldsymbol{y}|\boldsymbol{\theta}) \approx \boldsymbol{u}_0 \prod_{t=1}^{T} \Gamma_t Q_t \mathbf{1}$, where $\Gamma_t = \left[\gamma_{ij,t}\right]_{i,j=1,\ldots,B}$, with

$$\gamma_{ij,t} = \mathbb{P}(h_t - \mu \in B_j | h_{t-1} - \mu = b_i^*) = \Phi\left(\frac{b_j - \phi b_i^*}{\sigma}\right) - \Phi\left(\frac{b_{j-1} - \phi b_i^*}{\sigma}\right),$$

$$Q_t = \operatorname{diag}\left(\varphi\left(\frac{y_t}{\exp((\mu + b_i^*)/2)}\right)\right)_{i=1,\ldots,B},$$

where Φ and φ denote the cdf and the pdf of the standard normal density, respectively. Notice that the transition probabilities are time-constant so that the underlying Markov

chain is homogeneous. Sandmann and Koopman (1998) point out that for the SV model such a form of numerical integration might not be always suitable since a fixed grid cannot efficiently accommodate for periods of low and high volatility. We address this issue by suggesting a more efficient, adaptive HMM-based approximation as an alternative to the fixed bins used by Langrock et al. (2012b).

Finally, we note that for μ and σ^2 Gibbs updates can be performed based on full conditional densities, see Kim et al. (1998). Furthermore, numerous bespoke enhancements for sampling of the hidden states has been devised, see e.g. Kim et al. (1998). However, we provide a general framework requiring only "vanilla" type updates (based on a MH RW algorithm) and consider the full DA as a comparison benchmark.

Dependence structure and SCDL The basic SV specification concerns a single one-dimensional state on the real line and the sampling inefficiency originates from a high persistence of the logvolatility process. In order to break this dependence, we propose to impute $\boldsymbol{h_{2T}}$ and integrate out $\boldsymbol{h_{2T+1}}$, the states in even and odd time periods, respectively. This corresponds to the *vertical* integration scheme with $\boldsymbol{x}_{int} = \boldsymbol{h_{2T+1}}$ and $\boldsymbol{x}_{aug} = \boldsymbol{h_{2T}}$. Without loss of generality we assume that T is odd so that h_T is integrated out and we denote $T^* = \frac{T-1}{2}$; if T is even then we add one extra integration based on uniformly distributed h_{T+1}. The exact SCDL is given by

$$p(y, \boldsymbol{h_{2T}}) = p(h_0) \int p(h_1|h_0)p(y_1|h_0) \left(\prod_{t=1}^{T^*} p(y_{2t+1}|h_{2t+1})p(h_{2t+1}|h_{2t})p(y_{2t}|h_{2t})p(h_{2t}|h_{2t-1}) \right) dh_1 \dots dh_T,$$

and can be split (by conditioning on the even states) into a product $T^* + 1$ of integrals

$$p(y, \boldsymbol{h_{2T}}) = \underbrace{p(h_0)}_{=:C_0} \underbrace{\int p(h_1|h_0)p(y_1|h_0)dh_1}_{=:D_0} \prod_{t=1}^{T^*} \underbrace{p(y_{2t}|h_{2t})}_{=:C_t} \underbrace{\int p(y_{2t+1}|h_{2t+1})p(h_{2t+1}|h_{2t})p(h_{2t}|h_{2t-1})dh_{2t+1}}_{=:D_t}.$$

Since the integrals above are conditionally independent, we have

$$p(y, \boldsymbol{h_{2T}}) = C_0 D_0 \prod_{t=1}^{T^*} C_t D_t = \prod_{t=0}^{T^*} C_t D_t.$$

Hidden Markov model approximations The integrals D_t cannot be evaluated analytically and need to be numerically approximated. We consider two different approaches.

Case (i) Fixed bins This approach follows Langrock et al. (2012b) and consists in relating $z_t = k$, the Markov chain being in state k, to the event $h_{2t+1} - \mu \in \mathcal{B}_k$, the demeaned volatility in an odd time period $2t+1$ falling into the kth bin B_k. We take equally spaced bins, each of length λ. Falling into bin B_k can be specified as e.g. lying in the interval $[b_{k-1}, b_k)$ or being equal to this interval's midpoint $b_k^* = \frac{b_{k-1}+b_k}{2}$. In particular, we consider approximation of the following form

$$D_t \approx \sum_{k=1}^{B} p(y_{2t+1}|h_{2t+1} - \mu = b_k^*) p(h_{2t+2}|h_{2t+1} - \mu = b_k^*) p(h_{2t+1} - \mu \in B_i|h_{2t}). \tag{18}$$

The last term in (18) can be approximated as

$$p(h_{2t+1} - \mu \in B_k|h_{2t}) \approx \Phi\left(\frac{b_k - \phi(h_{2t} - \mu)}{\sigma}\right) - \Phi\left(\frac{b_{k-1} - \phi(h_{2t} - \mu)}{\sigma}\right),$$

which is adopted in Langrock et al. (2012b), or using a simpler midpoint approximation

$$p(h_{2t+1} - \mu \in B_k|h_{2t}) \approx \lambda\varphi\left(\frac{b_k^* - \phi(h_{2t} - \mu)}{\sigma}\right),$$

which we adopt in our application due to computing time.

Case (ii) Adaptive bins Instead of specifying the grid points, we can fix the probabilities for each bin (which in Case (i) needed to be determined) and consider quantiles corresponding to intervals of equal probability. The vector of mid-quantiles $\boldsymbol{q}^*$ determines the midpoints at time $2t+1$, which are given as $\beta_{k,2t+1}^* = \phi(h_{2t} - \mu) + \sigma \cdot \Phi^{-1}(q_k^*)$, $k = 1, \ldots, B$, where h_{2t} is the imputed volatility for the previous time period. We approximate D_t as

$$D_t \approx \sum_{k=1}^{B} \varphi\left(\frac{y_{2t+1}}{\exp((\beta_{k,2t+1}^* + \mu)/2)}\right) \varphi\left(\frac{h_{2t+2} - \mu - \phi\beta_{k,2t+1}^*}{\sigma}\right) \cdot \frac{1}{B},$$

since $\gamma_{ij,t} = \frac{1}{B}$, and these constant transition probabilities from an imputed state cancel out in the acceptance ratios.

Online Appendix D shows how the SCDA scheme easily extends to more complex models, the SV in the mean model of Koopman and Uspensky (2002) or SV with leverage.

Application We consider daily log-returns of the IBM stock from 4th January 2000 to 29th December 2017 (4527 observations, see Figure 3). Below we discuss SCDA for the basic SV model as well as for the extended SVML model (with leverage and SV in the mean, see Online Appendix D). For both models we use adaptive intervals based on 10, 20 and 30 quantiles, while for the SV model we also consider fixed bins based on 20 and 30 intervals (fixed bins turned out to be unfeasible for the SVML model due to the necessary number of bins for stability). For fixed bin we set $b_0 = -4$ and $b_B = 4$ for the integration range. The obtained posterior means for the imputed volatilities suggest that this choice was sufficient, as the estimated state means range roughly from -1 to 3. For each model and method we simulate 50,000 draws after a burn-in of 10,000.

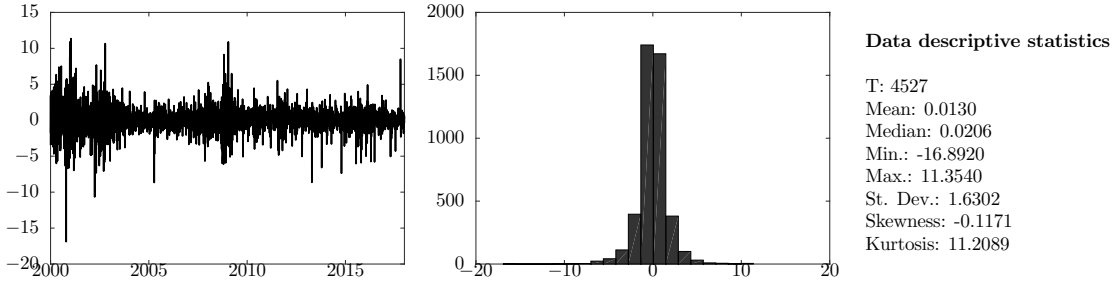


Figure 3: SV model: IBM series, from 4th January 2000 to 29th December 2017.

Tables 1 and 2 present the estimation results for $\boldsymbol{\theta}$ for the SV model and SVML model, respectively (we report on selected volatilities in Online Appendix E.2). For both models all the methods deliver comparable posterior estimates. A good agreement of the HMM-based schemes with the benchmark DA approach demonstrates that the developed methods provide a close approximation to the exact semi-complete data posterior. Interestingly, as few as 10 *adaptive* bins suffice to provide accurate estimates, which contrasts with *minimum* 50 fixed bins considered by Langrock et al. (2012b). This demonstrates the flexibility of the adaptive bins used within the SCDA scheme.

Tables 1 and 2 further reveal that the proposed vertical integration scheme breaks the strong dependence between subsequent states to improve mixing for $\boldsymbol{\theta}$ (Online Appendix E.2 contains the corresponding results for selected imputed volatilities). The effective sample sizes (ESSs, see Online Appendix B) for $\boldsymbol{\theta}$ obtained with the SCDA methods

Table 1: SV model: parameter posterior means, standard deviations and ESSs.

Method	DA	Adapt10	Adapt20	Adapt30	Fixed20	Fixed30
Time [s]	112	1980	2290	2567	1683	2057
μ	0.376	0.382	0.379	0.376	0.381	0.378
std	(0.115)	(0.116)	(0.115)	(0.115)	(0.116)	(0.115)
ESS	3201	5398	5440	**5784**	4504	5484
ϕ	0.962	0.962	0.961	0.961	0.962	0.962
std	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
ESS	114	249	279	121	240	**3014**
σ^2	0.081	0.086	0.085	0.084	0.082	0.081
std	(0.011)	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)
ESS	63	136	143	60.656	147	**168**

are typically higher than for the full DA approach. The only exception is the β parameter of the SVML for which all the methods exhibit excellent mixing with the DA approach slightly outperforming the HMM-based approximations. This high efficiency in the estimations of β is related to the presence of this parameter only in the observation equation hence being less affected by the high autocorrelation of the state process. On the other hand, the second extra parameter of the SVML model, i.e. the leverage parameter ρ, is hard to estimate efficiently. For this parameter the SCDA turns out particularly useful in improving the mixing with the corresponding ESS values being up to 4.5 higher than for the benchmark DA. Online Appendix E.2 provides ACF plots for $\boldsymbol{\theta}$ and the selected volatilities, for the SV and SVML model. As suggested by the ESS values reported in Tables 1–2, in the majority of the cases we observe much quicker decays in the autocorrelations for the SCDA algorithm compared to the "vanilla" DA approach.

However, we note that the computing times are higher for the SCDA approaches, with the computations for the adaptive case based on 10 bins taking roughly 17 times and 7 times longer than for full DA for the basic SV model and the SVML model, respectively. This suggests that the resulting gains in mixing may not necessarily be worth the extra computational cost. However, given the very simple structure of the basic SV model and

Table 2: SVML model: parameter posterior means, standard deviations and ESSs.

Method	DA	Adapt10	Adapt20	Adapt30
Time [s]	203	1511	2039	2497
μ	0.389	0.377	0.375	0.373
std	(0.115)	(0.112)	(0.115)	(0.113)
ESS	2510	**5879**	4836	5446
ϕ	0.96	0.963	0.961	0.96
std	(0.006)	(0.006)	(0.006)	(0.006)
ESS	119	340	**432**	239
σ^2	0.085	0.081	0.084	0.084
std	(0.011)	(0.011)	(0.012)	(0.013)
ESS	49	172	**187**	149.68
β	0.005	0.006	0.006	0.006
std	(0.009)	(0.009)	(0.009)	(0.009)
ESS	**7375**	6969	6692	7186
ρ	-0.286	-0.289	-0.293	-0.292
std	(0.047)	(0.044)	(0.045)	(0.048)
ESS	147	**682**	513	552

not much more complex one of the SVML model, this is perhaps not very surprising. We expect the SCDA approach to be more beneficial for more complex models, with even more involved dependence structure and relatively slower computation time for the benchmark DA approach. This can be already partly seen from shorter *relative* (to DA) computing times for the SCDA methods for the SVML compared to these for the SV model. For instance, the proposed integration scheme for the SV model could be particularly useful for a dynamic factor model with double stochastic volatility (where both the observation and the factor disturbances are subject to stochastic volatility). Due to the complex dependence structure as well matrix computations involved, the standard DA can be expected to perform relatively poorly and be time consuming to run. Then, there are several possibilities how to specify the augmentation-integration scheme, e.g. to fully integrate one of the SV

processes; or interweave between every-second state of both SV processes (e.g. to integrate odd states for one SV process and even states for another SV process).

5.2 Ecological model: lapwing data

We consider $\boldsymbol{y} = (y_1, \ldots, y_T)$, a time series of observations relating to census data (abundance index) of adult British lapwings (*Vanellus vanellus*, see Online Appendix C for details). The lapwings dataset plays an important role in statistical ecology where it has frequently served as an illustration (see King, 2011; Besbeas et al., 2002).

The counts are only estimates of the true unknown population size, which is assumed to change over time according to a first order Markov process. The latent population is related to two times series: for first-years and adults, denoted $\boldsymbol{N_1} = (N_{1,1}, \ldots, N_{1,T})^T$ and $\boldsymbol{N_a} = (N_{a,1}, \ldots, N_{a,T})^T$, respectively. Hence, the latent state is given by $\boldsymbol{x} = (\boldsymbol{N_1}^T, \boldsymbol{N_a}^T)^T$. Following Besbeas et al. (2002) we model the count data via the following SSM

$$y_t | N_{a,t}, \boldsymbol{\theta} \sim \mathcal{N}(N_{a,t}, \sigma_y^2), \qquad t = 1, \ldots, T, \tag{19}$$

$$N_{1,t} | N_{a,t-1}, \boldsymbol{\theta} \sim \text{Poisson}(N_{a,t-1}\rho_{t-1}\phi_{1,t-1}), \tag{20}$$

$$N_{a,t} | N_{1,t-1}, N_{a,t-1}, \boldsymbol{\theta} \sim \text{Bin}\big((N_{1,t-1} + N_{a,t-1}), \phi_{a,t-1}\big), \tag{21}$$

where $N_{1,0} \sim$ Neg–bin$(r_{1,0}, p_{1,0})$ and $N_{a,0} \sim$ Neg–bin$(r_{a,0}, p_{a,0})$. The model is parametrized by the time-varying productivity rate ρ_t, and time-varying survival rates $\phi_{1,t}$ and $\phi_{a,t}$, for first-years and adults, respectively, while $a_{i,0}$ and $p_{i,0}$ are hyperparameters of the prior distribution on the initial state value $N_{i,0}$, $i \in \{1, a\}$.

We let the SMM parameters follow regressions specified by Besbeas et al. (2002)

$$\text{logit}\, \phi_{1,t} = \alpha_1 + \beta_1 f_t, \quad \text{logit}\, \phi_{a,t} = \alpha_a + \beta_a f_t, \quad \log \rho_t = \alpha_\rho + \beta_\rho \tilde{t},$$

where $\tilde{t}$ the normalized time index and f_t denotes the normalized value of frost days $fdays$ in year t, see Online Appendix C for the explanation of this covariate.

To improve the estimation, Besbeas et al. (2002) consider an integrated population model combining the census data with ring-recovery data (see Online Appendix C for the formula of the additional regression parametrized by α_λ and β_λ and further details). We

refer to Besbeas et al. (2002) for a more detailed description of the integrated model. The set of model parameters is collected in a vector $\boldsymbol{\theta} = (\alpha_1, \alpha_a, \alpha_\rho, \alpha_\lambda, \beta_1, \beta_a, \beta_\rho, \beta_\lambda, \sigma_y^2)^T$.

Finally, we set independent vague $\mathcal{N}(0, 100)$ priors for the regression coefficients α_i and β_i, $i \in \{1, a, \rho, \lambda\}$ and $\Gamma^{-1}(a_y, b_y)$ on σ_y^2 with $a_y = 0.001 = b_y$. For the initial states, we set $r_{1,0} = 4$ and $p_{1,0} = 0.98$ so that the prior mean and variance of first-year birds is approximately 200 and $10,000$, respectively; and $r_{a,0} = 111$ and $p_{a,0} = 0.9$, so that the prior mean and variance of adults is approximately $1,000$ and $10,000$, respectively.

System (19)–(21) is non-Gaussian and nonlinear with the associated likelihood unavailable in a closed form. The standard vanilla DA approach leads to poorly mixing MCMC algorithms as demonstrated by King (2011). To this end, we first consider the dependence structure in the model to determine sensible $\boldsymbol{x}_{int}$ and $\boldsymbol{x}_{aug}$.

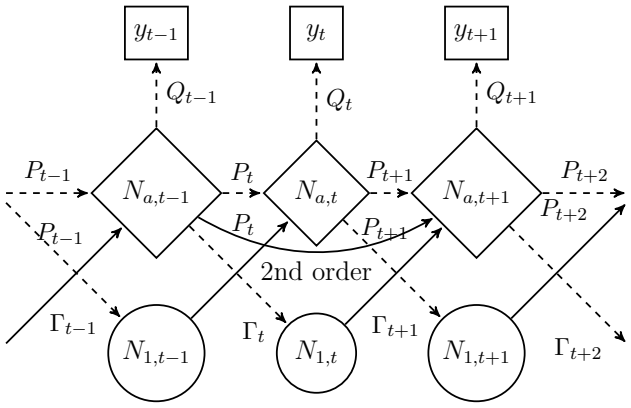


Figure 4: Lapwing data: combining DA and HMM structure. Diamonds – the imputed nodes, squares – the data, circles – the unknown variables. Integrating out $\boldsymbol{N_1}$ leads to a second order HMM on $\boldsymbol{N_a}$. Dashed lines for the relations *from* the imputed states.

The two-dimensional state $(N_{1,t}, N_{a,t})^T$ follows the first-order Markov process with a non-trivial transition kernel. First-year birds in t only feed into adults in $t+1$, however adults in t contribute to both the number of first-years and adults in $t+1$ as well as to the observed y_t. This suggests that reducing the strength of the dependence structure can be obtained by integrating out $\boldsymbol{N_1}$ while imputing $\boldsymbol{N_a}$. This corresponds to the *horizontal* integration scheme with $\boldsymbol{x}_{int} = \boldsymbol{N_1}$ and $\boldsymbol{x}_{aug} = \boldsymbol{N_a}$. The resulting modified dependence structure is presented in Figure 4. Marginalizing over $\boldsymbol{N_1}$ simplifies the analysis as only

$\boldsymbol{N_a}$ need to be considered and they now follow a second-order Markov process. A similar second order structure in this context has also been noted by Besbeas and Morgan (2019).

Hidden Markov Model approximation SCDL for the augmented data $(\boldsymbol{y}^T, \boldsymbol{N_a}^T)^T$ is

$$p(\boldsymbol{y}, \boldsymbol{N_a}|\boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{N_a}, \boldsymbol{\theta})p(\boldsymbol{N_a}|\boldsymbol{\theta}), \tag{22}$$

which is still intractable, thus we approximate it using the HMM embedding in (14). Since $N_{1,t}$ follows a Poisson distribution, we only need to specify a truncation value N^* for the maximum population size for first-years for a fixed bin approach (i.e. we set $b_B = N^*$, with $b_0 = 0$). Since the observations $\boldsymbol{y}$ are conditionally independent from $\boldsymbol{N_1}$ given $\boldsymbol{N_a}$, summing over of $\boldsymbol{N_1}$ is necessary for the second term on the right hand side of (22), to obtain the marginal pmf for $\boldsymbol{N_a}$. The marginal pmf of $\boldsymbol{N_a}$ is

$$p(\boldsymbol{N_a}) = \sum_{\boldsymbol{N_1}} p(N_{a,0})p(N_{1,0}) \prod_{t=1}^{T} p(N_{1,t}|N_{a,t-1})\, p(N_{a,t}|N_{a,t-1}, N_{1,t-1}). \tag{23}$$

We consider the "exact" approximation with the bin size equal to one, in which the approximation error is only due to truncating to N^*. A typical element in the product in (23) can be approximated as (for $t \geq 2$)

$$p(N_{a,t}|\boldsymbol{N}_{a,0:t-1}) = \sum_{k=0}^{N^*} \underbrace{\mathbb{P}(N_{1,t-1} = k|N_{a,t-2})}_{=:u_{k,t-1}} \underbrace{p(N_{a,t}|N_{a,t-1}, N_{1,t-1} = k)}_{=:p_{k,t}}, \tag{24}$$

where $p_{k,t}$ denotes the conditional pmf of $N_{a,t}$ given $N_{1,t-1} = k$ and $N_{a,t-1}$. Next, $u_{k,t}$ (collected in $\boldsymbol{u}_t = [u_{k,t}]_{k=1}^{N^*}$) denotes the "quasi-unconditional" probability of $N_{1,t} = k$ (in the sense of the Markov structure but not in terms of $N_{a,t-1}$) and can be derived as

$$\begin{aligned} u_{k,t} &= \mathbb{P}(N_{1,t} = k|N_{a,t-1}) = \sum_{l=0}^{N^*} \mathbb{P}(N_{1,t-1} = l|\boldsymbol{N}_{a,0:t-1})\mathbb{P}(N_{1,t} = k|N_{1,t-1} = l, \boldsymbol{N}_{a,0:t-1}) \\ &= \sum_{l=0}^{N^*} \underbrace{\mathbb{P}(N_{1,t-1} = l|N_{a,t-2})}_{=u_{l,t-1}} \underbrace{\mathbb{P}(N_{1,t} = k|N_{a,t-1})}_{=:\gamma_{lk,t}}. \end{aligned}$$

In general, we have $\boldsymbol{u}_t = \boldsymbol{u}_{t-1}\Gamma_t$ with $\Gamma_t = [\gamma_{lk,t}]_{l,k=1}^{N^*}$ and $\gamma_{lk,t} = \mathbb{P}(N_{1,t} = k|N_{1,t-1} = l, \boldsymbol{N}_{a,0:t-1})$. However, since in our case $N_{1,t}$'s are mutually independent given $N_{a,t-1}$, the

transition probabilities simplify to $\gamma_{lk,t} = \mathbb{P}(N_{1,t} = k|N_{a,t-1})$ for $k = 0, \ldots, N^* - 1$, while for $k = N^*$ we need $\gamma_{lk,t} = 1 - \sum_{j=0}^{N^*-1} \gamma_{lj,t}$ to ensure a valid probability distribution. Thus, the time varying state transition matrix Γ_t takes a simple form with rows equal to $(\boldsymbol{\gamma}_{11,t}, \ldots, \boldsymbol{\gamma}_{(N^*-1)(N^*-1),t}, \boldsymbol{\gamma}_{N^*N^*,t})$. Next, we express (24) in a convenient matrix notation

$$p(N_{a,t}|\boldsymbol{N}_{a,0:t-1}) = \underbrace{\begin{bmatrix} \gamma_{11,t-1} & \cdots & \gamma_{1N^*,t-1} \\ \vdots & \ddots & \vdots \\ \gamma_{11,t-1} & \cdots & \gamma_{1N^*,t-1} \end{bmatrix}}_{=\Gamma_{t-1}} \underbrace{\begin{bmatrix} p_{1,t} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & p_{N^*,t} \end{bmatrix}}_{=:P_t} \underbrace{\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}}_{\mathbf{1}} = \Gamma_{t-1} P_t \mathbf{1}.$$

Combining (23) and (24) yields the HMM form for the joint pmf of the imputed states

$$p(\boldsymbol{N_a}) = \boldsymbol{u}_0 p(N_{a,0}) \left(\prod_{t=1}^{T} P_t \Gamma_t \right) \mathbf{1},$$

where $\boldsymbol{u}_0 = [p(N_{1,0} = 0), \ldots, p(N_{1,0} = N^*)]^T$ is the initial state distribution.

Since the real observations y_t, conditionally on $N_{a,t}$, are independent of $N_{1,t}$, the observation matrix becomes the identity matrix $\mathbb{I}$ scaled by $p(y_t|N_{a,t})$, i.e. $Q_t = p(y_t|N_{a,t})\mathbb{I}$. Finally, the approximation to the SCDL (22) can be expressed as

$$p(\boldsymbol{y}, \boldsymbol{N_a}|\boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{N_a}) p(\boldsymbol{N_a}) = \boldsymbol{u}_0 p(N_{a,0}) \left(\prod_{t=1}^{T} P_t \Gamma_t Q_t \right) \mathbf{1}.$$

Results We compare the performance of the standard DA approach, in which we impute $\boldsymbol{\theta}$, $\boldsymbol{N_1}$ and $\boldsymbol{N_a}$, with that of the SCDA, in which we impute $\boldsymbol{\theta}$ and $\boldsymbol{N_a}$. For comparability we use a "vanilla" MH RW algorithm for the estimation of the integrated model, with discrete uniform updates for the states and normal updates for the regression coefficients. We use a Gibbs update $\sigma_y^2|\boldsymbol{N_a} \sim \Gamma^{-1}\left(a_y + T/2, b_y + \sum_{t=1}^{T}(y_t - N_{a,t})^2/2\right)$ for the observation variance. For the SCDA we first consider the "exact" integration used in the derivations above, in which the only influence on the posterior is the upper limit which we set $b_B = 679$. This choice of the upper bound is based on the results for first-years from previous studies. We further consider a number of approximate schemes based on fixed and adaptive intervals (with 10, 20 and 30 bins in each case). For adaptive bins we use a normal approximation to the Poisson distribution. Each time we use $100,000$ draws after a burn-in of $10,000$.

Table 3: Lapwing data: absolute (in seconds) and relative (wrt DA) computing times.

Method	DA	Adapt10	Adapt20	Adapt30	Fixed10	Fixed20	Fixed30	Exact
Absolute time	1203	978	1067	1024	1022	1060	1135	2855
Relative time	1.00	0.81	0.89	0.85	0.85	0.88	0.94	2.37

Table 3 summaries computation time for each scheme. As expected, the exact method is the slowest (2.5 times than the full DA approach) as each integration is based on summing 680 elements. All the approximate schemes are faster (10–20%) than the DA approach due to their efficient implementation based on vectorized computations with relatively few elements to be summed every iteration. Table 4 presents the results for the regression parameters in terms of posterior means and standard deviations as well as ESSs and ESSs per second, for DA and selected SDCA approaches. Online Appendix E.2 provides the comparison for all elements of $\boldsymbol{\theta}$ and selected elements of $\boldsymbol{N_a}$ between all the methods.

The results demonstrate the efficiency of the proposed SCDA approach: all the SCDA-based schemes, except the one based on 10 fixed bins, outperform the full DA approach by delivering much higher (up to 4 times) ESSs and ESSs/sec. This is illustrated in Figure 5, which shows the autocorrelation (ACF) plots for the SSM parameters. We refer to Online Appendix E.1 for the ACF plots for selected elements of $\boldsymbol{N_a}$.

6 Discussion

We have presented a new estimation method for state space models using semi-complete data augmentation, designed to increase the efficiency of "vanilla" MCMC algorithms. The main idea behind the approach is to combine data augmentation with numerical integration, where the latter aims at reducing the dependence between the imputed auxiliary variables. The concept relates to general Rao-Blackwellization methods, however we do not require the resulting conditional distribution (given the imputed states) to be analytically integrable, nor the imputed auxiliary variables to be sufficient statistics for the marginalized states.

We propose integration schemes based on the insights from hidden Markov models

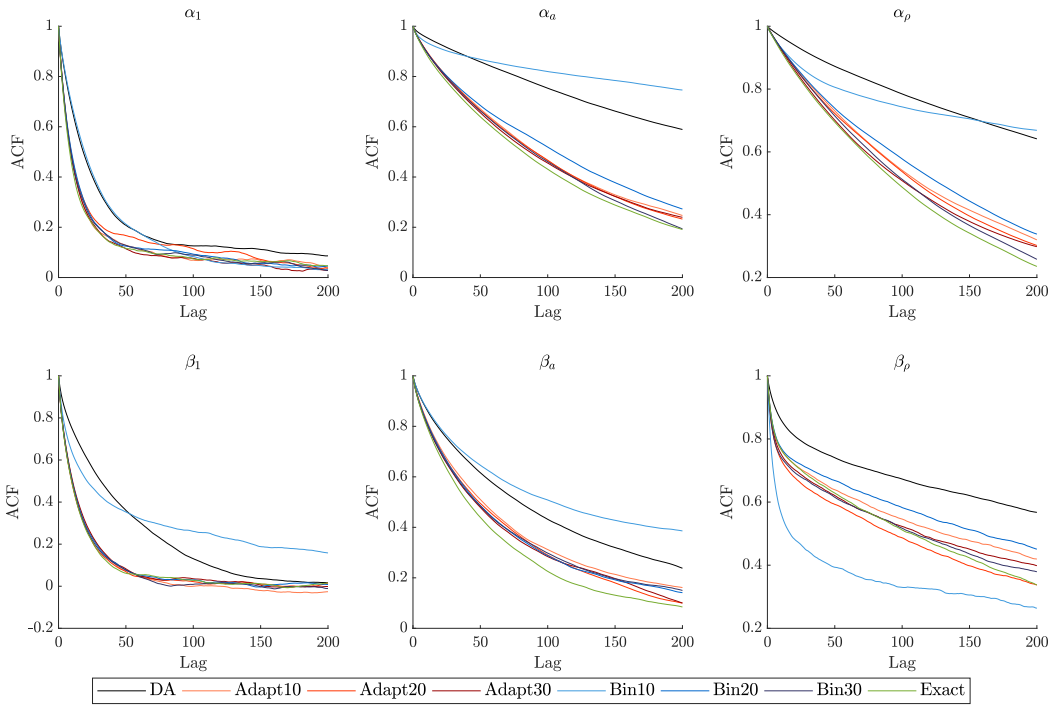


Figure 5: Lapwing data: ACF plots for the SSM regression parameters.

in the sense that we specify new transition probabilities between redefined states, to be numerically integrated out, conditionally on the auxiliary variables. Further efficiency gains can be obtained by "binning". This results in an approximation to SCDL and we note that for continuous states such an approximation is a natural starting point for our approach (as in principle for any MC based analysis). We consider two types of "binning": "fixed bins" based on a pre-specified grid and "adaptive bins" based on e.g. quantiles of the relevant distribution. The latter removes the problem of specifying the "essential domain" required for fixed bins (see Kitagawa, 1987; Langrock et al., 2012b). Adaptive bins are also more suited for problems with highly varying integration ranges, such as the class of SV models, for which fixed bins are unlikely to be efficient (see Sandmann and Koopman, 1998). In our examples a similar accuracy was achieved by using fewer adaptive bins than fixed bins.

The two empirical studies considered demonstrate the gains from applying the SCDA approach compared to the general "vanilla" MCMC algorithm. For the lapwings data model the efficiency gains are substantial, not only in terms of higher effective sample sizes compared to the standard DA technique but also when taking into account the computing

Table 4: Lapwing data: SSM regression parameter posterior means, standard deviations and ESSs. The highest ESS and ESS/sec. for each parameter in bold. Computing times in square brackets.

Method		α_1	α_a	α_ρ	β_1	β_a	β_ρ
DA	Mean	0.547	1.574	-1.189	-0.164	-0.240	-0.348
	(Std)	(0.068)	(0.071)	(0.091)	(0.062)	(0.039)	(0.043)
	ESS	685	124	112	1050	389	106
[1204 s]	ESS/sec.	0.57	0.10	0.09	0.87	0.32	0.09
Adapt10	Mean	0.547	1.564	-1.180	-0.163	-0.239	-0.350
	(Std)	(0.068)	(0.070)	(0.092)	(0.061)	(0.040)	(0.040)
	ESS	1490	390	316	2777	**527**	126
[978 s]	ESS/sec.	1.52	**0.40**	**0.32**	**2.84**	**0.54**	0.13
Fixed10	Mean	0.512	1.441	-1.044	-0.207	-0.205	-0.348
	(Std)	(0.070)	(0.055)	(0.063)	(0.050)	(0.039)	(0.022)
	ESS	942	34	37	181	105	**282**
[1022 s]	ESS/sec.	0.92	0.03	0.03	0.18	0.10	**0.28**
Fixed30	Mean	0.545	1.562	-1.170	-0.162	-0.240	-0.342
	(Std)	(0.069)	(0.073)	(0.095)	(0.061)	(0.039)	(0.040)
	ESS	**1758**	**439**	**329**	**2873**	502	208
[1136 s]	ESS/sec.	**1.55**	0.39	0.29	2.53	0.44	0.18

time (ESS/sec.). For the SV and SVML models, SCDA boosts the mixing, however at the cost of an increased computing time. Nonetheless, for larger models with a more complex dependency structure, such as dynamic factor models with double stochastic volatility, the proposed SCDA method is likely to become much more profitable – also in terms of increased ESS/sec – as discussed in Section 5.1.

The split of the latent states into "auxiliary" and "integrated" variables is model-dependent and should be specified in such a way that the algorithm is efficient. This choice is not unique and multiple approaches may be applied – the efficiency of these will dependent on both the model and data. On the one hand, the imputed states aim to have reduced

correlation, to improve mixing of MCMC algorithms; on the other hand, the numerical integration is over a very low number of dimensions, which in many cases is feasible due to conditional independence of the integration problems. To identify such conditionally independent latent states it can be useful to investigate the underlying graphical structure of the model. In general, high dimensional integration remains a challenging problem, which we leave for further research, noting that quasi Monte Carlo could be useful in this context.

The proposed methodology naturally leads to several topics for further research. First, we aim to investigate bounds of approximation errors in order to quantify the demonstrated higher usefulness of adaptive bins compared to fixed bins. Second, adopting automated methods to identify the correlation structure would make applying the SCDA approach to new models easier and potentially more efficient, especially if the model at hand is complex and/or there are no "natural candidates" for the integrated states. Third, we expect parallelization methods to reduce the increased computing time recorded for the SV models. Since updating a given state is associated with conditioning on only two other states, the previous one and the next one (see Figure 2 in Online Appendix A.2), it is possible to update every second augmented state in parallel. In principle, such an approach could be also adopted for the lapwings data, however there updating every fourth state could only be used due to the second-order dependence in the associated HMM representation.

SUPPLEMENTARY MATERIAL

Online Appendix (SCDAOnlineAppendix.pdf, pdf file) provides more details on the specifications of the HMM-based approximations to the three models considered in the paper (Part A); on computing the effective sample size (Part B); on the lapwings dataset (Part C); the extensions of the basic SV model (Part D). Finally, additional results for both applications are presented (Part E). 11

MATLAB codes (SCDACode.zip, zip file) consist of scripts, functions and data used for the empirical applications. Two README txt files instruct the reader on their use.

Bibliography

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

Andrieu, C. and Roberts, G. (2009), "The Pseudo-Marginal Approach for Efficient Monte Carlo Computations," *The Annals of Statistics*, 37, 697–725.

Beaumont, M. (2003), "Estimation of Population Growth or Decline in Genetically Monitored Populations," *Genetics*, 164, 1139–1160.

Besbeas, P., Freeman, S. N., Morgan, B. J. T., and Catchpole, E. A. (2002), "Integrating Mark–Recapture–Recovery and Census Data to Estimate Animal Abundance and Demographic Parameters," *Biometrics*, 58, 540–547.

Besbeas, P. and Morgan, B. J. T. (2019), "Exact Inference for Integrated Population Modelling," *Biometrics*, 75, 475–484.

Brooks, S. P., King, R., and Morgan, B. J. T. (2004), "A Bayesian Approach to Combining Animal Abundance and Demographic Data," *Animal Biodiversity and Conservation*, 27, 515–529.

Cappé, O., Moulines, E., and Ryden, T. (2006), *Inference in Hidden Markov Models*, Springer Series in Statistics, Springer New York.

Doucet, A., de Freitas, N., and Gordon, N. (eds.) (2001), *Sequential Monte Carlo Methods in Practice*, Springer.

Doucet, A., Freitas, N. D., Murphy, K., and Russell, S. (2000), "Rao-Blackwellised Particle Filtering for Dynamic Bayesian Networks," in *Proceedings of the Sixteenth conference on Uncertainty in Artificial Intelligence*, pp. 176–183.

Durbin, J. and Koopman, S. J. (2012), *Time Series Analysis by State Space Methods: Second Edition*, Oxford Statistical Science Series, OUP Oxford.

Fridman, M. and Harris, L. (1998), "A Maximum Likelihood Approach for non-Gaussian Stochastic Volatility Models," *Journal of Business & Economic Statistics*, 87, 284—-291.

Frühwirth-Schnatter, S. (1994), "Data Augmentation and Dynamic Linear Models," *Journal of Time Series Analysis*, 15, 183–202.

— (2004), "Efficient Bayesian Parameter Estimation," in *State Space and Unobserved Component Models: Theory and Applications*, eds. A. C. Harvey, S. J. Koopman, and N. Shephard, chap. 7, Cambridge University Press, pp. 123–151.

Gelman, A., Roberts, G. O., and Gilks, W. R. (1996), "Efficient Metropolis Jumping Rule," in *Bayesian Statistics 5: Proceedings of the Fifth Valencia International Meeting*, eds. S. Brooks, J. G. A. Gelman, and X. L. Meng, Oxford University Press, pp. 599–607.

Hobert, J. P. (2011), "The Data Augmentation Algorithm: Theory and Methodology," in *Handbook of Markov Chain Monte Carlo*, eds. S. Brooks, A. Gelman, J. Galin, and X. L. Meng, chap. 10, CRC Press, pp. 253–294.

Hobert, J. P., Royand, V., and Robert, C. P. (2011), "Improving the Convergence Properties of the Data Augmentation Algorithm with an Application to Bayesian Mixture Modeling," *Statistical Science*, 26, 332–351.

Jacob, P. E. and Thiery, A. H. (2015), "On Nonnegative Unbiased Estimators," *The Annals of Statistics*, 43, 769–784.

Kim, S., Shephard, N., and Chib, S. (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *The Review of Economic Studies*, 65, 361–393.

King, R. (2011), "Statistical Ecology," in *Handbook of Markov Chain Monte Carlo*, eds. S. Brooks, J. G. A. Gelman, and X. L. Meng, chap. 17, CRC Press, pp. 419–447.

King, R., Brooks, S. P., Mazzetta, C., Freeman, S. N., and Morgan, B. J. T. (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., McClintock, B. T., Kidney, D., and Borchers, D. (2016), "Capture–recapture Abundance Estimation using a Semi-complete Data Likelihood Approach," *Annals of Applied Statistics*, 10, 264–285.

Kitagawa, G. (1987), "Non-Gaussian State-Space Modeling of Nonstationary Time Series," *Journal of the American Statistical Association*, 82, 1032–1041.

Koopman, S. J. and Uspensky, E. H. (2002), "The Stochastic Volatility in Mean Model: Empirical Evidence from International Stock Markets," *Journal of Applied Econometrics*, 17, 667–689.

Korattikara, A., Chen, Y., and Welling, M. (2014), "Austerity in MCMC Land: Cutting the Metropolis-Hastings Budget," in *Proceedings of the 31st International Conference on International Conference on Machine Learning-Volume 32*, pp. 181–189.

Langrock, R. and King, R. (2013), "Maximum Likelihood Estimation of Mark–Recapture–Recovery Models in the Presence of Continuous Covariates," *The Annals of Applied Statistics*, 7, 1709–1732.

Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., and Morales, J. M. (2012a), "Flexible and Practical Modeling of Animal Telemetry Data: Hidden Markov Models and Extensions," *Ecology*, 93, 2336–2342.

Langrock, R., MacDonald, I. L., and Zucchini, W. (2012b), "Some Nonstandard Stochastic Volatility Models and their Estimation using Structured Hidden Markov Models," *Journal of Empirical Finance*, 147–161.

Marin, J. M. and Robert, C. (2007), *Bayesian Core: a Practical Approach to Computational Bayesian Statistics*, Springer Science & Business Media.

Murphy, K. P. (2002), "Dynamic Bayesian Networks: Representation, Inference and Learning," PhD thesis, University of California, Berkeley.

Roberts, G. O. and Rosenthal, J. S. (2001), "Optimal Scaling for Various Metropolis-Hastings Algorithms," *Statistical Science*, 16, 351–367.

Sandmann, G. and Koopman, S. J. (1998), "Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood," *Journal of Econometrics*, 87, 271–301.

Tanner, M. A. and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–540.

Taylor, S. J. (1994), "Modeling Stochastic Volatility: A Review and Comparative Study," *Mathematical Finance*, 4, 183–204.

West, M. and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models: Second Edition*, Springer-Verlag.

Zucchini, W., MacDonald, I. L., and Langrock, R. (2016), *Hidden Markov Models for Time Series: An Introduction Using R, Second Edition*, Monographs on Statistics and Applied Probability 150, CRC Press.

Online Appendix for: Semi-Complete Data Augmentation for Efficient State Space Model Fitting

Agnieszka Borowska
University of Glasgow, School of Mathematics and Statistics, UK
and
Ruth King
School of Mathematics, University of Edinburgh, UK

A Specification details of the HMM approximations

In this section we present how the general formulation of the HMM-based approximation to SCDL can be applied to the examples discussed in Sections 4 and 5.

A.1 Motivating example from Section 4.2

SSM from Section 4.2 is given by

$$\begin{aligned} y_t|x_{1,t},x_{2,t} &\sim p(x_{1,t},x_{2,t}),\\ x_{1,t+1}|x_{1,t},x_{2,t} &\sim p(x_{1,t},x_{2,t}),\\ x_{2,t+1}|x_{1,t},x_{2,t} &\sim p(x_{1,t},x_{2,t}),\\ x_{i,0} &\sim p(x_{i,0}), i=1,2 \end{aligned}$$

and we aim at imputing $x_{1,t}$ and integrating out $x_{2,t}$, which implies $T_{int}=T_{aug}=\{0,1,\ldots,T\}$. Hence, the index functions $\tau(t)$, $a(t)$ and $o(t)$ are simply identities and we skip them below to simplify the exposition. Figure 1 illustrates the corresponding dependencies.

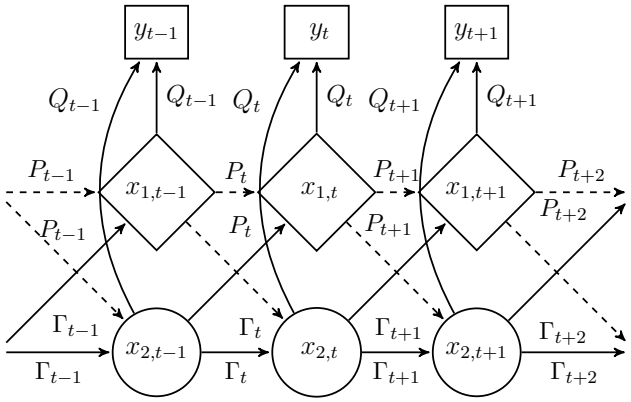


Figure 1: Illustration of combining DA and HMM structure. Conditionally independent observations added to the state specification from Figure 2a Diamonds represent the imputed states, circles – the integrated states. Dashed lines used for the relations *from* the imputed (known) states..

The "quasi marginal" distribution[1] of the imputed state $x_{1,t}$ can be approximated as

$$p(x_{1,t}|\boldsymbol{x}_{1,0:t-1}) \approx \sum_{j=1}^{B} \underbrace{\mathbb{P}(x_{2,t-1} \in \mathcal{B}_j|\boldsymbol{x}_{1,0:t-2})}_{=:u_{j,t-1}} \underbrace{p(x_{1,t}|x_{1,t-1}, x_{2,t-1} \in \mathcal{B}_j)}_{=:p_{j,t}}, \tag{1}$$

where $p_{j,t}$ is the likelihood of the augmented state at t given the imputed state at $t-1$ was in the jth bin (and previous realizations of x_1, but these are treated as known) and $u_{j,t-1}$ is the unconditional probability of the hidden process x_2 falling into the jth bin at $t-1$. This "quasi-unconditional" probability can be expressed as

$$u_{k,t} = \mathbb{P}(x_{2,t} \in \mathcal{B}_k|\boldsymbol{x}_{1,0:t-1}) = \sum_{j=1}^{B} \underbrace{\mathbb{P}(x_{2,t-1} \in \mathcal{B}_j|\boldsymbol{x}_{1,0:t-2})}_{=u_{j,t-1}} \underbrace{\mathbb{P}(x_{2,t} \in \mathcal{B}_k|\boldsymbol{x}_{1,0:t-1}, x_{2,t-1} \in \mathcal{B}_j)}_{=:\gamma_{jk,t}}, \tag{2}$$

which corresponds to the standard result in HMM that the unconditional distributions in subsequent periods are related via the transition matrix $\Gamma_t = [\gamma_{jk,t}]_{k,j=1,\dots,B}$ as follows (see Zucchini et al., 2016, p.16, 32)

$$\boldsymbol{u}_t = \boldsymbol{u}_{t-1}\Gamma_t.$$

[1] i.e. marginal in the sense of the Markov structure, not the augmented states which we treat as known.

Next, the observations are conditionally independent, hence

$$p(y_t|\boldsymbol{x}_{1,0:t}) \approx \sum_{k=1}^{B} u_{k,t} \underbrace{p(y_t|x_{1,t}, x_{2,t} \in \mathcal{B}_k)}_{=:q_{k,t}}, \tag{3}$$

with $q_{k,t}$ denoting the likelihood of the observation at t given the hidden state in the same period t falling into bin k (and the imputed state at t).

Comparing (2) and (3) shows that the distributions of the same period t augmented states $x_{1,t}$ and "real" observations y_t are conditioned on the latent states from different periods, i.e. $t-1$ and t, respectively. This is a consequence of the general dependence structure in SSMs. The transition matrix at t captures this change of the underlying state so that combining of all there parts (1), (2) and (3) results in

$$p(y_t, x_{1,t}|\boldsymbol{x}_{1,0:t-1}) \approx \sum_{j=1}^{B}\sum_{k=1}^{B} u_{j,t-1} p_{j,t} \gamma_{jk,t} q_{k,t}.$$

To compute the HMM-based approximation to the SCDL we consider *forward probabilities* $\alpha_\mathbf{t}$ of the imputed states $x_{1,t}$ and observations y_t (Zucchini et al., 2016, Sec. 2.3.2) defined as

$$\begin{aligned}
\boldsymbol{\alpha}_t &= p(x_{1,0})\boldsymbol{u}_0 \prod_{s=1}^{t} P_s \Gamma_s Q_s, \qquad t = 1, 2, \ldots T^*, \\
\boldsymbol{\alpha}_0 &= p(x_{1,0})\boldsymbol{u}_0 Q_0,
\end{aligned}$$

with $\boldsymbol{u}_0 = (\mathbb{P}(x_{2,0} \in \mathcal{B}_1), \ldots, \mathbb{P}(x_{2,0} \in \mathcal{B}_B))$ being the initial distribution of the latent state and $Q_0 = \mathbb{I}$. It follows from this definition that the forward probabilities can be expressed recursively as

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} P_t \Gamma_t Q_t,$$

so that the required approximation to the SCDL is given by

$$\hat{p}_B(\boldsymbol{y}, \boldsymbol{x_1}) = p(x_{1,0})\boldsymbol{u}_0 \boldsymbol{\alpha}_{T^*} \mathbb{I}.$$

Notice that the transition matrix Γ_t is a full matrix, however in some cases, e.g. the lapwing population model, the transition matrix can take a simpler form e.g. it is "column-wise constant": $\gamma_{lk,t} = \mathbb{P}(x_{2,t} = k|\boldsymbol{x}_{1,0:t-1}),\ \forall l$ (each row is the same). On the other hand, the

augmented observation matrix and the real observation matrix have diagonal forms $P_t = \text{diag}\,(p_{j,t})_{j=1,\dots,B}$ and $Q_t = \text{diag}\,(q_{k,t})_{j=k,\dots,B}$, respectively. Using the notation introduced in Section 4.2 we can write

$$\hat{p}_B(\boldsymbol{y}, \boldsymbol{x_1}) = p(x_{1,0})\boldsymbol{u}_0 Q_0 \prod_{t=1}^{T^*} (P_t \Gamma_t Q_t)\, \mathbf{1}.$$

We can verify the above results be explicitly calculating

$$\begin{aligned}
\boldsymbol{\alpha}_{t-1} P_t \Gamma_t Q_t \mathbf{1} &= \begin{bmatrix} \alpha_{1,t-1} & \dots & \alpha_{B,t-1} \end{bmatrix} \begin{bmatrix} p_{1,t} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & p_{B,t} \end{bmatrix} \begin{bmatrix} \gamma_{11,t} & \dots & \gamma_{1B,t} \\ \vdots & \ddots & \vdots \\ \gamma_{B1,t} & \dots & \gamma_{BB,t} \end{bmatrix} \begin{bmatrix} q_{1,t} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & q_{B,t} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} \alpha_{1,t-1} & \dots & \alpha_{B,t-1} \end{bmatrix} \begin{bmatrix} p_{1,t}\gamma_{11,t} & \dots & p_{1,t}\gamma_{1B,t} \\ \vdots & \ddots & \vdots \\ p_{B,t}\gamma_{B1,t} & \dots & p_{B,t}\gamma_{BB,t} \end{bmatrix} \begin{bmatrix} q_{1,t} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & q_{B,t} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} \alpha_{1,t-1} & \dots & \alpha_{B,t-1} \end{bmatrix} \begin{bmatrix} p_{1,t}\gamma_{11,t}q_{1,t} & \dots & p_{1,t}\gamma_{1B,t}q_{B,t} \\ \vdots & \ddots & \vdots \\ p_{B,t}\gamma_{B1,t}q_{1,t} & \dots & p_{B,t}\gamma_{BB,t}q_{B,t} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} \underbrace{\sum_{j=1}^{B} \alpha_{j,t-1} p_{j,t} \gamma_{j1,t} q_{1,t}}_{=\alpha_{1,t}} & \dots & \underbrace{\sum_{j=1}^{B} \alpha_{j,t-1} p_{j,t} \gamma_{jB,t} q_{B,t}}_{=\alpha_{B,t}} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \boldsymbol{\alpha}_t \mathbf{1}
\end{aligned}$$

and expressing

$$\hat{p}_B(\boldsymbol{y}, \boldsymbol{x_1}) = \sum_{k_0=1}^{B} \sum_{k_1=1}^{B} \cdots \sum_{k_{T^*}=1}^{B} p(x_{1,0}) u_{k_0,0} \prod_{t=1}^{T^*} p_{k_{t-1},t} \gamma_{k_{t-1}k_t,t} q_{k_t,t}.$$

A.2 SV model

Basic SV model The SCDL for the basic SV model can be expressed as

$$\begin{aligned}
p(y, \boldsymbol{h_{2T}}|\boldsymbol{\theta}) = p(h_0) \int p(h_1|h_0) p(y_1|h_0) \Bigg(\prod_{t=1}^{T^*} & p(h_{2t+1}|h_{2t}) p(y_{2t+1}|h_{2t+1}) \\
& \times p(h_{2t}|h_{2t-1}) p(y_{2t}|h_{2t}) \Bigg) dh_1 \dots dh_{T^*}, \qquad (4)
\end{aligned}$$

where $T^* = \frac{T-1}{2}$ (we assume T odd). Since we impute volatilities at even time periods the Markov chain is given by $\{z_t\} = \{h_{2t+1}\}$ for $t = 1, \dots, T^*$ and its transition matrix has the form

$$\Gamma_t = \begin{bmatrix} \mathbb{P}(h_{2t+1} \in \mathcal{B}_1 | h_{2t-1} \in \mathcal{B}_1, h_{2t}) & \dots & \mathbb{P}(h_{2t+1} \in \mathcal{B}_B | h_{2t-1} \in \mathcal{B}_1, h_{2t}) \\ \vdots & \ddots & \vdots \\ \mathbb{P}(h_{2t+1} \in \mathcal{B}_1 | h_{2t-1} \in \mathcal{B}_B, h_{2t}) & \dots & \mathbb{P}(h_{2t+1} \in \mathcal{B}_B | h_{2t-1} \in \mathcal{B}_B, h_{2t}) \end{bmatrix}$$
$$= \begin{bmatrix} \mathbb{P}(h_{2t+1} \in \mathcal{B}_1 | h_{2t}) & \dots & \mathbb{P}(h_{2t+1} \in \mathcal{B}_B | h_{2t}) \\ \vdots & \ddots & \vdots \\ \mathbb{P}(h_{2t+1} \in \mathcal{B}_1 | h_{2t}) & \dots & \mathbb{P}(h_{2t+1} \in \mathcal{B}_B | h_{2t}) \end{bmatrix}.$$

We can see that the rows of Γ_t are the same, which means that the hidden states are conditionally independent given the imputed states. For the augmented observation matrix P_t we have

$$P_t = \text{diag } \left(p(h_{2t} | h_{2t-1} \in \mathcal{B}_j)\right)_{j=1,\dots,B}.$$

The observation matrix has the form

$$Q_t = \text{diag } \left(p(y_{2t}, y_{2t+1} | h_{2t+1} \in \mathcal{B}_j, h_{2t})\right)_{j=1,\dots,B}.$$

Inserting Γ_t, P_t and Q_t in (14) with $\tau(t) = 2t+1$, $a(t) = 2t$ and $o(t) = \{2t, 2t+1\}$ leads to

$$\hat{p}_B(y, \boldsymbol{h_{2T}}) = p(h_0)\mathbf{u}_0 Q_0 \prod_{t=1}^{T^*} P_t \Gamma_t Q_t \mathbf{1}, \tag{5}$$

where $\mathbf{u}_0 = \begin{bmatrix} \mathbb{P}(h_1 \in \mathcal{B}_k | h_0) & \dots & \mathbb{P}(h_1 \in \mathcal{B}_B | h_0) \end{bmatrix}$ and $Q_0 = \text{diag } (y_1 | h_1 \in \mathcal{B}_k)_{k=1,\dots,B}$. Then (5) is an HMM-based approximation to (4) converging to its true value in $B \to \infty$ and $b_0 \to -\infty$, $b_B \to \infty$.

Figure 2 illustrates the HMM approximation and shows dependencies relevant for a single imputation problem.

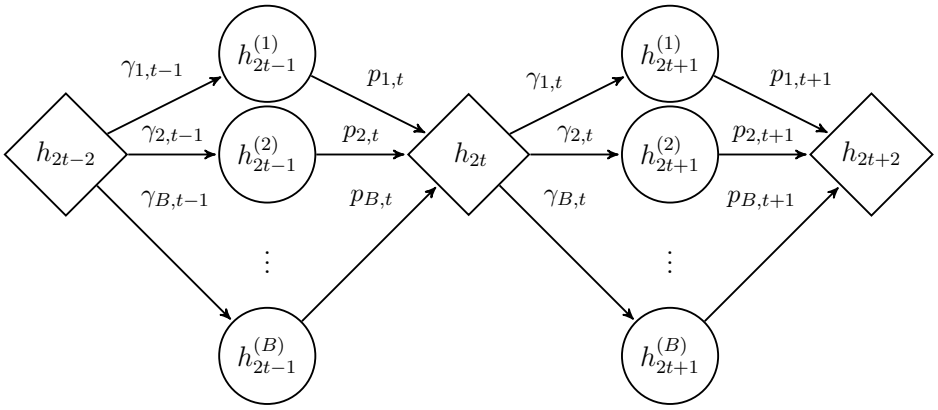


Figure 2: **SV model**: combining DA and the HMM-based integration. Diamonds represent the imputed states, circles – the states being integrated out. $h_t^{(k)}$ denotes $h_t \in \mathcal{B}_k$. The graph presents a single imputation problem of h_{2t} with the associated integrations.

SVML model For the SVML model we only need to adjust the matrices P_t and Q_t as the dependence structure of the observations remains unchanged

$$\Gamma_t = \begin{bmatrix} \mathbb{P}(h_{2t+1} \in \mathcal{B}_1 | h_{2t}, y_{2t}) & \dots & \mathbb{P}(h_{2t+1} \in \mathcal{B}_B | h_{2t}, y_{2t}) \\ \vdots & \ddots & \vdots \\ \mathbb{P}(h_{2t+1} \in \mathcal{B}_1 | h_{2t}, y_{2t}) & \dots & \mathbb{P}(h_{2t+1} \in \mathcal{B}_B | h_{2t}, y_{2t}) \end{bmatrix},$$

$$P_t = \text{diag}\ \left(p(h_{2t} | h_{2t-1} \in \mathcal{B}_j, y_{2t-1})\right)_{j=1,\dots,B}.$$

A.3 Lapwing population model

The approximation for the lapwings model is a special case of scheme used for the general model discussed in the Section A.1, with the transition matrix Γ_t having equal rows. The HMM is here given as $\{z_t\} = \{N_{1,t}\}$ for $t = 0, \dots, T$ and we again set $T_{int} = T_{aug} = \{0, 1, \dots, T\}$, so that the index functions $\tau(t)$, $a(t)$ and $o(t)$ are simply identities. The

transition matrix has the form

$$\Gamma_t = \begin{bmatrix} \mathbb{P}(N_{1,t} = b_1^* | N_{1,t-1} = b_1^*, \boldsymbol{N}_{a,0:t-1}) & \dots & \mathbb{P}(N_{1,t} = b_B^* | N_{1,t-1} = b_1^*, \boldsymbol{N}_{a,0:t-1}) \\ \vdots & \ddots & \vdots \\ \mathbb{P}(N_{1,t} = b_1^* | N_{1,t-1} = b_B^*, \boldsymbol{N}_{a,0:t-1}) & \dots & \mathbb{P}(N_{1,t} = b_B^* | N_{1,t-1} = b_B^*, \boldsymbol{N}_{a,0:t-1}) \end{bmatrix}$$
$$= \begin{bmatrix} \mathbb{P}(N_{1,t} = b_1^* | N_{a,t-1}) & \dots & \mathbb{P}(N_{1,t} = b_B^* | N_{a,t-1}) \\ \vdots & \ddots & \vdots \\ \mathbb{P}(N_{1,t} = b_1^* | N_{a,t-1}) & \dots & \mathbb{P}(N_{1,t} = b_B^* | N_{a,t-1}) \end{bmatrix},$$

with $b_k^* = k$, for $k = 0, \dots, N^*$. We can see that for each column of Γ_t its elements are the same. For the augmented observation matrix P_t we have

$$P_t = \text{diag}\left(p(N_{a,t} | N_{a,t-1}, N_{1,t-1} = b_j^*)\right)_{j=1,\dots,B},$$

so P_t and Γ_t condition on the same hidden state. The observation matrix has a simple form

$$Q_t = p(y_t | N_{a,t}) \mathbb{I}.$$

Inserting Q_t, P_t and Q_t in (14) leads to

$$\hat{p}_B(y, \boldsymbol{N_a}) = p(N_{a,0}) \mathbf{u}_0 Q_0 \prod_{t=1}^{T^*} P_t \Gamma_t Q_t \mathbf{1}, \tag{6}$$

where $\mathbf{u}_0 = \begin{bmatrix} \mathbb{P}(N_{1,0} \in \mathcal{B}_k) & \dots & \mathbb{P}(N_{1,0} \in \mathcal{B}_B) \end{bmatrix}$ and $Q_0 = \mathbb{I}$. Then (6) is an HMM-based approximation to (22) converging to its true value in $B \to \infty$ and $b_B \to \infty$.

B Effective sample size

Since the samples generated by MCMC algorithms are not independent, the standard variance estimator cannot be used to measure the variance of the empirical average delivered by an MCMC algorithm. The asymptotic variance σ^2_{MCMC} of the (stationary) Markov chain $X_1, X_2, \dots$ is given by

$$\sigma^2_{\text{MCMC}} = \text{Var}[X_i] \underbrace{\left(1 + 2 \sum_{k=1}^{\infty} \rho(k)\right)}_{\text{IF}}, \tag{7}$$

where $\rho(k)$ is the kth order serial correlation (Geyer, 2011). The term in the parentheses in (7) is referred to as the *inefficiency factor* (IF, Pitt et al., 2012). High values of autocorrelation, typically reported for MCMC sampling, lead to the standard variance estimator underestimating the true variance σ^2_{MCMC}. A common measure to assess the deterioration in the sampling efficiency due to the draws autocorrelation is the *effective sample size* (ESS) defined as

$$\text{ESS} = \frac{M}{IF},$$

where M is the sample size (Robert and Casella, 2004, Ch. 12.3.5). It indicates what the size of an independent sample would be, had it the same variance as the MCMC sample..

In practice, one typically cannot compute the IF directly and needs to estimate it instead. As noted by Robert and Casella (2004, Ch. 12.3.5) estimation of IF is a "delicate issue", as it contains an infinite sum. A possible solution to this problem is set a cut-off value K for the autocorrelation terms being summed up: $\widehat{IF} = 1 + 2\sum_{k=1}^{K} \hat{\rho}(k)$. The choice of K poses the risk of subjectiveness; setting K to the lowest lag at which $\hat{\rho}(k)$ become insignificant seems to be a reasonable solution suggested by e.g. Kass et al. (1998) or Pitt et al. (2012) and this is the approach we take here.

C Lapwings dataset

The lapwings dataset plays an important role in statistical ecology and has served as an illustration in several handbooks (see King, 2011; King et al., 2010) and papers (e.g. Besbeas et al., 2002) in this field. It was also used as an example of a complex statistical model by e.g. Goudie et al. (2018). One of the main reasons for such a particular interest in this species is a sharp decline in its population in recent years: its European population is considered as *near threatened* by International Union for Conservation of Nature (2018), while in Britain in particular is has been moved to the *red list* of species of conservation concern, see The Royal Society for the Protection of Birds (2018) (i.e. of the highest conservation priority, with species needing urgent action) from the *amber list* (mentioned by previous literature, see Besbeas et al., 2002; Brooks et al., 2004). The population serves

as an indicator species for other farmland birds, giving us an insight into the dynamics of similar bird species.

We follow the approach of Besbeas et al. (2002) and use three datasets for the lapwings application: the count census data for the population index, the weather data on the number of frost days, and the ring-recovery data. Combining of independent sources of data underlies the integrated population modeling (IPM) framework and allows for a more precise parameter estimation. This is due to the survival parameters $\alpha_i, \beta_i, i \in 1, a$, being common to the state space model for the census data and to the ring-recovery model

Census data The census data are derived from the Common Birds Census (CBC) of the British Trust for Ornithology, which recently has been replaced by the Breeding Bird Survey. The dataset is constructed as annual estimates of the number of breeding female lapwings based on annual counts made at a number of sites around the UK. Since only a small fraction of sites are surveyed each year, the index can be seen as a proxy for the total population size. For comparability, we use the same time span as Brooks et al. (2004) and King (2011), i.e. from 1965 to 1998. The choice of the starting year is there motivated by the fact that in earlier years the index protocol was being standardized. Finally we note that year 1965 is associated with time index $t = 3$, for consistency with the ring-recovery data (to be discussed below) which start in 1963. The left panel in Figure 3 presents the census data.

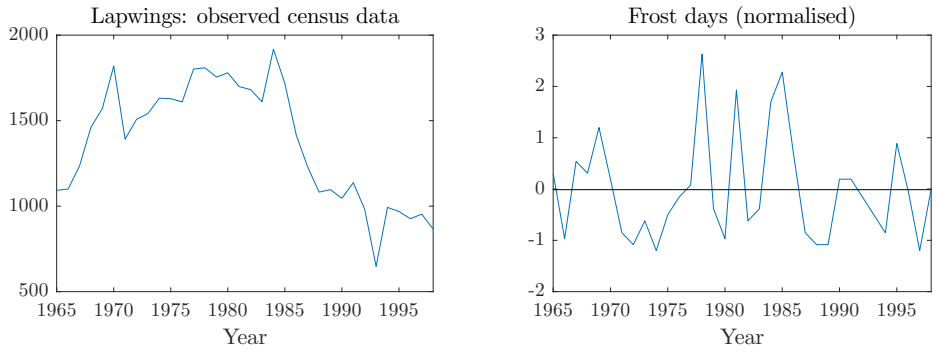


Figure 3: Lapwings census data and normalised frost days.

Weather data For bird species there is a natural relationship between the survival probabilities and the weather conditions, most importantly winter severity. Following Besbeas et al. (2002) we measure this factor for year t by the number of days between April of year t and March of year $(t+1)$ inclusive in which the temperature in Central England fell below freezing and denote it by $fdays_t$. We further normalize $fdays_t$ to obtain f_t which we use as a regressor in the logistic regression for the survival probabilities (see the right panel in Figure 3). As noted by King (2011), normalization of covariates is done to improve the mixing of the sampling scheme and to facilitate the interpretation of the parameters of the logistic regression (intercept and slope).

Ring–recovery data Ring–recovery studies aim at estimating demographic parameters of the population under consideration including first-year survival probabilities and adult survival probabilities. These studies consist in marking individuals (e.g. with a ring or a tag) at the beginning of period t and then releasing them. In subsequent periods $t+1, t+2, \ldots$ the number of dead animals is recorded, where it is assumed that any recovery of a dead animal is immediate. For lapwings, the ringed birds are chicks ("fist-years") and a "period" corresponds to a "bird year" i.e. 12 months from April to March. We analyze the ring-recovery data for the releases from 1963 to 1997, with the recoveries up to 1998.

Ring-recovery data are stored in an array, an example of which is provided in Table 1. The first column corresponds to the number of ringed animals in a given year R_t, $t = 1, \ldots, T-1$, and the subsequent columns report the number of rings $m_{t,s}$ recovered in the interval $(s, s+1]$, $s = 1, \ldots, T-1$, from animals released in year t. Obviously, $m_{t,s} = 0$ for $t > s$. Finally, we denote by $m_{t,T}$ the number of individuals ringed in year t but never seen again (their rings are not recovered), where $m_{t,T} = R_t - \sum_{s=1}^{T-1} m_{t,s}$.

The parameters of interest are $\phi_{1,t}$, $\phi_{a,t}$ and λ_t. The former two are the conditional probabilities of survival until year $t+1$ of a first-year and an adult, respectively, given such an individual is alive in year t. The latter one is the conditional probability of recovering a ring at time t, given the individual dies in the interval $(t-1, t]$. Each row $\boldsymbol{m}_t = \{m_{t,s}\}_{s=1}^{T}$, $t = 1, \ldots, T-1$, of the m-array is multinomially distributed: $\boldsymbol{m}_t \sim \mathcal{MN}(R_t, \boldsymbol{q}_t)$ ($\mathcal{MN}$ denotes the multinomial distribution), where $\boldsymbol{q}_t = \{q_{t,s}\}_{s=1}^{T}$ are the multinomial cell

Year of	Number	Year of Recovery s										
Ringing t	Ringed	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974
1963	1147	14	4	1	2	1	0	1	1	0	0	0
1964	1285		20	3	4	0	1	1	0	0	0	0
1965	1106			10	1	2	2	0	2	2	1	1
1966	1615				9	7	4	2	1	1	0	0
1967	1618					12	1	6	2	0	0	1
1968	2120						9	6	4	0	2	2
1969	2003							10	8	5	3	1
1970	1963								8	3	2	0
1971	2463									4	1	1
1972	3092										7	2
1973	3442											15

Table 1: A fragment of Ring-Recovery Data for lapwings for the years 1963-1973, table from King (2011).

probabilities specified for $t = 1, \ldots, T-1$ as[2]

$$q_{t,s} = \begin{cases} 0, & s = 1, \ldots, t-1, \\ (1-\phi_{1,t})\lambda_t & s = t, \\ \phi_{1,t}\left(\prod_{k=t+1}^{s-1}\phi_{a,k}\right)(1-\phi_{a,s})\lambda_s, & s = t+1, \ldots, T-1, \\ 1 - \sum_{s=1}^{T+1} q_{t,s}, & s = T. \end{cases}$$

The likelihood of the m-array is then given by

$$p(\mathbf{m}|\boldsymbol{\phi_1}, \boldsymbol{\phi_a}, \boldsymbol{\lambda}) \propto \prod_{\mathbf{t=1}}^{\mathbf{T-1}} \prod_{\mathbf{s=t}}^{\mathbf{T}} \mathbf{q}_{\mathbf{t,s}}^{\mathbf{m_{t,s}}}.$$

The array $\boldsymbol{m} = [m_{t,s}]_{t=1,\ldots,T-1}^{s=1,\ldots,T}$ is a sufficient statistic for ring-recovery data.

Following Besbeas et al. (2002) we assume that the time-varying recovery rate λ_t follows a logistic regression given by

$$\text{logit}\,\lambda_t = \log\left(\frac{\lambda_t}{1-\lambda_t}\right) = \alpha_\lambda + \beta_\lambda \tilde{t}_t,$$

[2]For $s = t+1$ we put $\prod_{k=t+1}^{s-1} := 1$.

where $\tilde{t}$ is the normalized time index.

D Extensions of the basic SV model

D.1 SV in the mean

The proposed SCDA scheme easily extends to more complex models, e.g. the popular Stochastic Volatility in the Mean (SVM) model of Koopman and Uspensky (2002) (see also Chan, 2017). Its basic specification is given by

$$y_t|h_t, \boldsymbol{\theta} \sim \mathcal{N}\left(\beta \exp(h_t), \exp(h_t)\right), \tag{8}$$

$$h_{t+1}|h_t \sim \mathcal{N}\left(\mu + \phi(h_t - \mu), \sigma^2\right), \tag{9}$$

$$h_0 \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{1-\phi^2}\right), \tag{10}$$

for $t = 1, \ldots, T$. Hence, the latent volatility process h_t influences both the conditional variance and the conditional mean of the observation series y_t, which is additionally controlled by a scaling parameter β. For the volatility parameters μ, ϕ and σ^2 we adopt the prior specification as for the standard SV model, while for the mean-scaling parameter we specify $\beta \sim \mathcal{N}(0, \sigma^2_{\beta_0})$, with $\sigma^2_{\beta_0} = 10$.

D.2 SV with leverage

The basic SV or SVM models can be extended to allow for *leverage* effects, i.e. a feedback from past logreturns to the current value of the volatility process. This effect is typically modeled as a negative correlation between the last period logreturns and the current value of volatility. The motivation behind the leverage effect is that the volatility in financial markets may adapt differently to positive and negative shocks/news (affecting logreturns), where large negative shocks are likely to increase the volatility. The SV model with leverage (SVL) has been frequently analyzed in the literature, see Jungbacker and Koopman (2007), Meyer and Yu (2000), Yu (2005), Durbin and Koopman (2012, Section 9.5.5.) or Zucchini

et al. (2016, Section 20.2.3). For convenience, we rewrite the basic SV model (15)–(16) as

$$\begin{aligned} y_t &= \exp(h_t/2)\varepsilon_t, & \varepsilon_t &\sim \mathcal{N}(0,1), \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \eta_t, & \eta_t &\sim \mathcal{N}\left(0, \sigma^2\right), \\ h_1 &\sim \mathcal{N}\left(\mu, \frac{\sigma^2}{1-\phi^2}\right), \end{aligned}$$

for $t = 1, \ldots, T$. The only difference between the SVL model and the basic specification of the SV model is that now the error terms ε_t and η_t are assumed to be correlated: $\text{corr}[\varepsilon_t, \eta_t] = \rho \neq 0$, with ρ typically estimated to be negative. This apparently slight modification has, however, substantial effect on the dependence structure in the model (see Figure 4) and hence the conditional distribution of h_t. To derive the latter several reformulations of the model has been proposed (Jungbacker and Koopman, 2007 or Meyer and Yu, 2000), however we will use the treatment provided by Zucchini et al. (2016, Section 20.2.3). These authors use the basic regression lemma for normal variables to show that

$$h_t | h_{t-1}, y_{t-1}, \mu, \phi, \sigma^2, \rho \sim \mathcal{N}\left(\mu + \phi(h_{t-1} - \mu) + \frac{\rho\sigma y_{t-1}}{\exp(h_{t-1}/2)}, \sigma^2(1-\rho^2)\right) \tag{11}$$

(the details of the derivation are provided in the next subsection). Formulation (11) is particularly convenient for "reusing" the derived integration scheme for the basic SV model, as we only need to adjust the transition probabilities in the approximation to D_t.

D.3 Modifications to the HMM-based approximation

The proposed HMM-based approximation to SCDL can be easily adapted to allow for both extension by simply modifying the components of the matrices Γ_t, P_t and Q_t specified in (10)–(12). Notice that for the SVM model the dependence structure of the state is the same as for the basic SV model, hence the core of the integration/imputation scheme remains unchanged. What needs to be adjusted is the observation density, which can be done is a straightforward manner. The modification for the SVL model requires adjusting of the transition probabilities and the pdfs of the augmented states. Below we present the required modifications for the largest model, allowing for both SV in the mean and for the leverage effect (which we refer to as the SVML model).

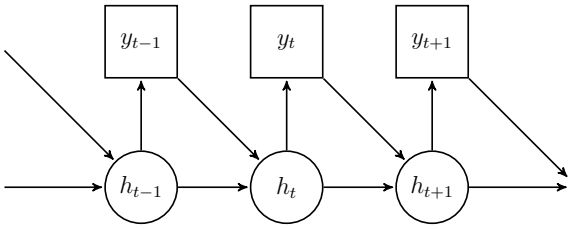


Figure 4: SV model with leverage: modified dependence structure due to feedback from the logreturns y_{t-1} to logvolatilities h_t.

Below, we skip $\boldsymbol{\theta}$ in the conditioning to simplify notation. Following Zucchini et al. (2016), we aim at deriving $p(h_{t+1}|h_t, y_t)$, the conditional distribution of h_{t+1} given h_t and y_t. Since $y_t = \exp(h_t/2)\varepsilon_t$ we can replace conditioning on y_t by conditioning on ε_t. Moreover, conditioning on h_t is equivalent to conditioning on η_t and adding the mean $\mu + \phi(h_t - \mu)$. Hence, we are interested in the the distribution of η_t given ε_t.

The distribution of $\eta_t|\varepsilon_t$ can be obtained using the basic result from multivariate normal regression, which we recall below for convenience:

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix} \right) \Rightarrow x|y \sim \mathcal{N}\left(\mu_x + \frac{\sigma_{xy}}{\sigma_y^2}(y - \mu_y), \sigma_x^2 - \frac{\sigma_{xy}^2}{\sigma_y^2} \right).$$

Hence, we obtain

$$\eta_t|\varepsilon_t \sim \mathcal{N}\left(0 + \frac{\rho\sigma}{1}(y - 0), \sigma^2 - \frac{\rho^2\sigma^2}{1} \right) = \mathcal{N}(\rho\sigma\varepsilon_t, \sigma^2(1 - \rho^2))$$

so that

$$h_{t+1}|h_t, \varepsilon_t \sim \mathcal{N}\big(\mu + \rho(h_t - \mu) + \rho\sigma\varepsilon_t, \sigma^2(1 - \rho^2)\big).$$

Finally, we can express the latter in terms of the actual observation y_t rather than the unobserved disturbance ε_t. For the basic SV model this becomes

$$h_{t+1}|h_t, y_t \sim \mathcal{N}\left(\mu + \rho(h_t - \mu) + \rho\sigma\frac{y_t}{\exp(h_t/2)}, \sigma^2(1 - \rho^2) \right),$$

which is the result reported in Section D.2, while for the SVM we have

$$h_{t+1}|h_t, y_t \sim \mathcal{N}\left(\mu + \rho(h_t - \mu) + \rho\sigma\frac{y_t - \beta\exp(h_t)}{\exp(h_t/2)}, \sigma^2(1 - \rho^2) \right).$$

E Additional results

E.1 Lapwings data

Figure 5 illustrates the posterior means and 95% credible intervals (CI) for the adult population comparing the accuracy of the full DA with that of the SCDA methods (separately for the adaptive intervals and fixed bins). We can see that all the methods deliver virtually the same posterior means and comparable 95% symmetric CI, with only the fixed bin case with 10 bins deviating slightly from all other methods. Interestingly, 10 adaptive bins give very comparable estimates to the other approaches in this case, indicating an increased accuracy of the adaptive approach.

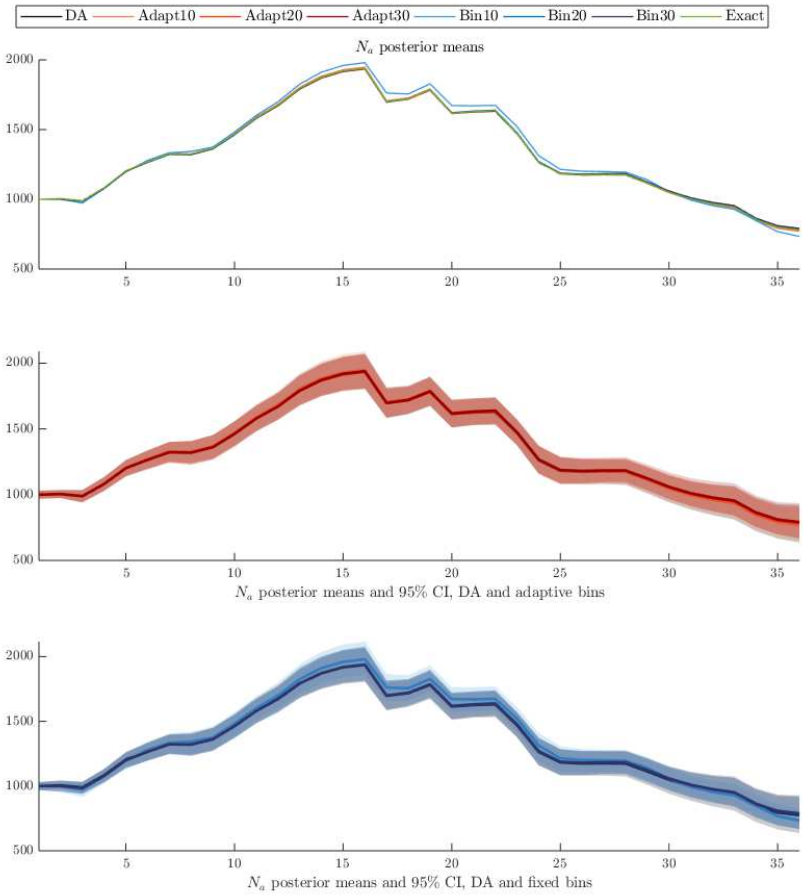


Figure 5: Lapwings data: the posterior means and 95% CI for the adult population.

Tables 2 and 3 report the estimation results for all the methods considered for the parameters and selected imputed states, respectively. Figure 6 shows the ACF plots for selected imputed states.

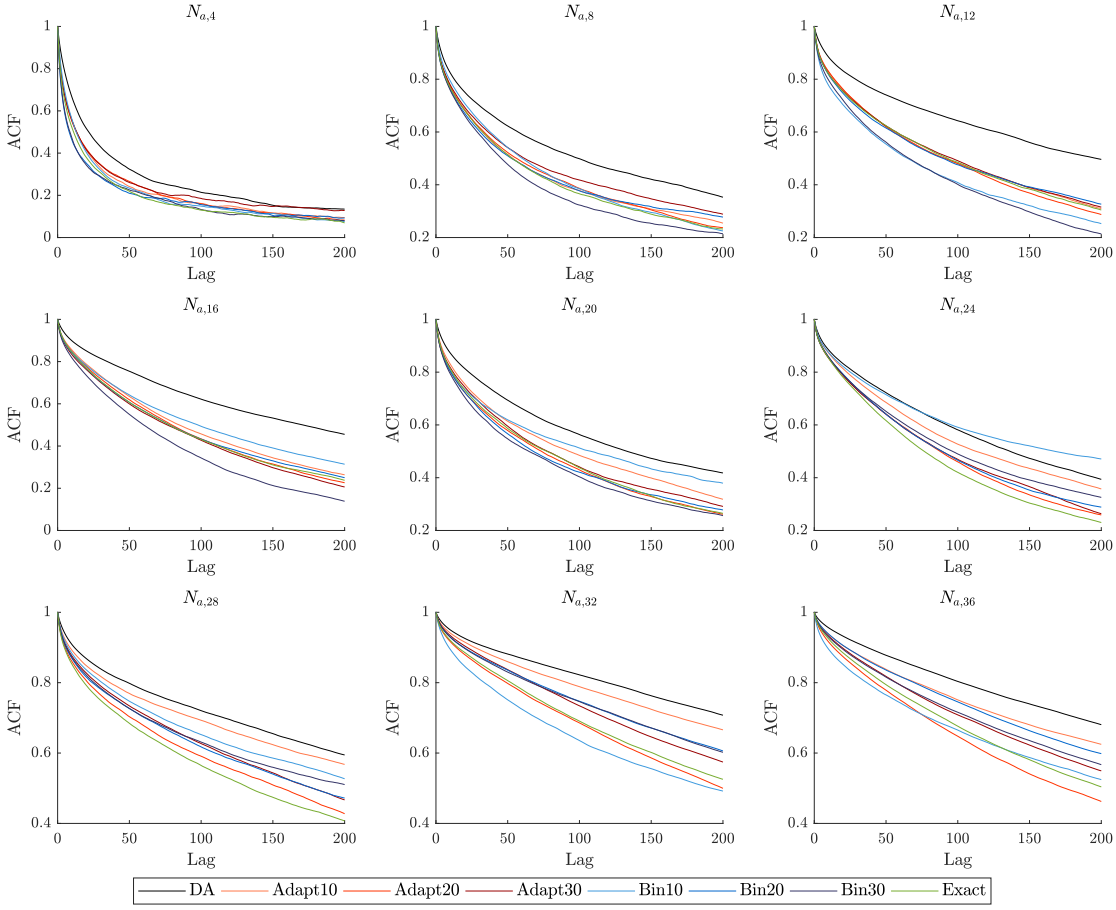


Figure 6: Lapwings data: ACF plots for the adult population.

Table 2: Lapwings data: parameter posterior means, standard deviations and ESSs. The highest ESS and ESS/sec. for each parameter in bold. Computing times (in seconds) in square brackets.

Method		α_1	α_a	α_ρ	α_λ	β_1	β_a	β_ρ	β_λ	σ_y^2
DA	Mean	0.547	1.574	-1.189	-4.578	-0.164	-0.240	-0.348	-0.364	30180
	(Std)	(0.068)	(0.071)	(0.091)	(0.035)	(0.062)	(0.039)	(0.043)	(0.04)	(8890)
	ESS	685	124	112	1089	1050	389	106	8206	1245
[1204 s]	ESS/sec.	0.57	0.10	0.09	0.90	0.87	0.32	0.09	6.82	1.03
Adapt10	Mean	0.547	1.564	-1.180	-4.580	-0.163	-0.239	-0.350	-0.364	30355
	(Std)	(0.068)	(0.070)	(0.092)	(0.035)	(0.061)	(0.040)	(0.040)	(0.040)	(8928)
	ESS	1490	390	316	**3035**	2777	527	126	7492	1852
[978 s]	ESS/sec.	1.52	**0.40**	**0.32**	**3.10**	**2.84**	0.54	0.13	7.66	1.89
Adapt20	Mean	0.544	1.564	-1.173	-4.581	-0.162	-0.238	-0.342	-0.363	30002
	(Std)	(0.069)	(0.072)	(0.094)	(0.035)	(0.060)	(0.039)	(0.039)	(0.040)	(8759)
	ESS	1359	395.695	324.918	2720.786	2685.188	586.964	243.425	8212	2075
[1068 s]	ESS/sec.	1.27	0.37	0.30	2.55	2.515	**0.55**	0.23	7.69	1.94
Adapt30	Mean	0.542	1.561	-1.166	-4.581	-0.162	-0.241	-0.339	-0.363	30311
	(Std)	(0.069)	(0.071)	(0.092)	(0.036)	(0.061)	(0.039)	(0.040)	(0.040)	(8888)
	ESS	1438	322	243	2736	2471	564	196	7146	2129
[1025 s]	ESS/sec.	1.40	0.31	0.24	2.67	2.41	0.55	0.19	6.97	2.08
Fixed10	Mean	0.512	1.441	-1.044	-4.599	-0.207	-0.205	-0.348	-0.353	29992
	(Std)	(0.070)	(0.055)	(0.063)	(0.034)	(0.050)	(0.039)	(0.022)	(0.040)	(8837)
	ESS	942	34	37	562	181	105	**282**	8771	1627
[1022 s]	ESS/sec.	0.92	0.03	0.04	0.55	0.18	0.10	**0.28**	**8.58**	1.59
Fixed20	Mean	0.546	1.570	-1.179	-4.579	-0.170	-0.240	-0.343	-0.364	30156
	(Std)	(0.069)	(0.069)	(0.090)	(0.035)	(0.061)	(0.039)	(0.040)	(0.040)	(8802)
	ESS	1250	270	210	2328	2566	525	139	8582	2270
[1060 s]	ESS/sec.	1.18	0.25	0.20	2.19	2.42	0.49	0.13	8.09	2.14
Fixed30	Mean	0.545	1.562	-1.170	-4.580	-0.162	-0.240	-0.342	-0.363	30012
	(Std)	(0.069)	(0.073)	(0.095)	(0.035)	(0.061)	(0.039)	(0.040)	(0.040)	(8698)
	ESS	**1758**	**438**	**329**	2902	**2873**	502	208	7613	2706
[1136 s]	ESS/sec.	**1.55**	0.39	0.29	2.55	2.52	0.44	0.18	6.70	**2.38**
Exact	Mean	0.545	1.564	-1.175	-4.580	-0.162	-0.240	-0.345	-0.363	30063

Table 3: Lapwings data: posterior means, standard deviations and ESSs. The highest ESS and ESS/sec. for each state in bold. Computing times (in seconds) in square brackets.

Method		Na_4	Na_8	Na_{12}	Na_{16}	Na_{20}	Na_{24}	Na_{28}	Na_{32}	Na_{36}	Na_{min}	Na_{max}
DA	Mean	1083.51	1325.45	1674.38	1935.84	1614.73	1264.61	1174.20	964.85	776.15	1113.76	1083.51
	(Std)	(26.31)	(43.08)	(52.06)	(67.99)	(53.97)	(53.68)	(49.46)	(59.57)	(72.07)	(50.97)	(26.31)
	ESS	460	179	120	134	147	154	59	62	68	55	460
[1204 s]	ESS/sec.	0.38	0.15	0.10	0.11	0.12	0.13	0.05	0.05	0.06	0.05	0.38
Adapt10	Mean	1083.46	1326.45	1681.57	1947.34	1621.63	1268.14	1174.99	962.14	770.63	1113.57	1083.46
	(Std)	(26.25)	(43.50)	(51.51)	(68.96)	(53.63)	(56.49)	(49.99)	(56.98)	(66.91)	(51.06)	(26.25)
	ESS	880	**393**	318	335	295	234	43	46	51	42	879
[978 s]	ESS/sec.	**0.90**	**0.40**	0.32	0.34	0.30	0.24	0.04	0.05	0.05	0.043	**0.90**
Adapt20	Mean	1081.74	1320.73	1670.04	1934.08	1615.29	1268.92	1181.09	973.61	785.42	1121.17	1081.74
	(Std)	(27.45)	(44.99)	(52.41)	(67.96)	(51.42)	(53.69)	(46.37)	(51.32)	(61.67)	(46.41)	(27.45)
	ESS	792	301	263	413	299	311	174	150	**161**	167	792
[1068 s]	ESS/sec.	0.74	0.28	0.25	0.39	0.28	0.29	0.16	0.14	0.15	0.16	0.74
Adapt30	Mean	1081.74	1319.29	1670.94	1938.90	1617.39	1268.43	1184.04	978.16	791.52	1124.32	1081.74
	(Std)	(27.37)	(45.49)	(51.54)	(65.25)	(53.62)	(54.57)	(48.40)	(56.59)	(68.24)	(49.64)	(27.37)
	ESS	597	278	247	434	326	306	181	155	164	**171**	597
[1025 s]	ESS/sec.	0.58	0.27	0.24	0.42	0.32	0.30	**0.18**	0.15	**0.16**	**0.17**	0.58
Fixed10	Mean	1075.91	1343.03	1699.24	1979.99	1671.88	1313.00	1194.68	954.96	733.69	1140.26	1075.91
	(Std)	(26.81)	(43.44)	(50.66)	(62.41)	(48.44)	(54.65)	(42.54)	(39.70)	(41.44)	(43.18)	(26.81)
	ESS	868	310	327	243	173	108	97	**163**	92	88	868
[1022 s]	ESS/sec.	0.85	0.30	0.32	0.24	0.17	0.10	0.10	**0.16**	0.09	0.09	0.85
Fixed20	Mean	1079.80	1319.96	1671.12	1939.17	1619.88	1270.62	1183.22	976.20	788.92	1123.72	1079.80
	(Std)	(25.97)	(44.14)	(51.98)	(67.01)	(52.05)	(54.00)	(46.78)	(53.32)	(64.39)	(47.5)	(25.97)
	ESS	785	279	225	331	344	328	74	58	64	67	785
[1060 s]	ESS/sec.	0.74	0.26	0.21	0.31	0.32	**0.31**	0.07	0.05	0.06	0.06	0.74
Fixed30	Mean	1079.48	1320.18	1671.22	1936.26	1615.844	1268.040	1181.895	975.230	787.55	1121.97	1079.48
	(Std)	(25.72)	(43.24)	(48.00)	(62.93)	(50.327)	(53.921)	(46.464)	(53.46)	(65.12)	(47.10)	(25.72)
	ESS	**911**	370	**374**	**505**	347	247	111	90	98	102	**911**
[1136 s]	ESS/sec.	0.80	0.33	**0.33**	**0.44**	**0.30**	0.22	0.10	0.08	0.09	0.09	0.80
Exact	Mean	1083.13	1324.13	1675.32	1939.63	1615.88	1265.87	1176.69	968.36	780.24	1116.33	1083.13
	(Std)	(27.19)	(44.55)	(52.06)	(67.38)	(53.86)	(54.55)	(46.09)	(54.11)	(66.99)	(47.25)	(27.19)
	ESS	902	349	293	366	**402**	**418**	**197**	122	117	168	902
[2855 s]	ESS/sec.	0.32	0.12	0.10	0.13	0.14	0.15	0.07	0.04	0.04	0.06	0.32

E.2 SV model

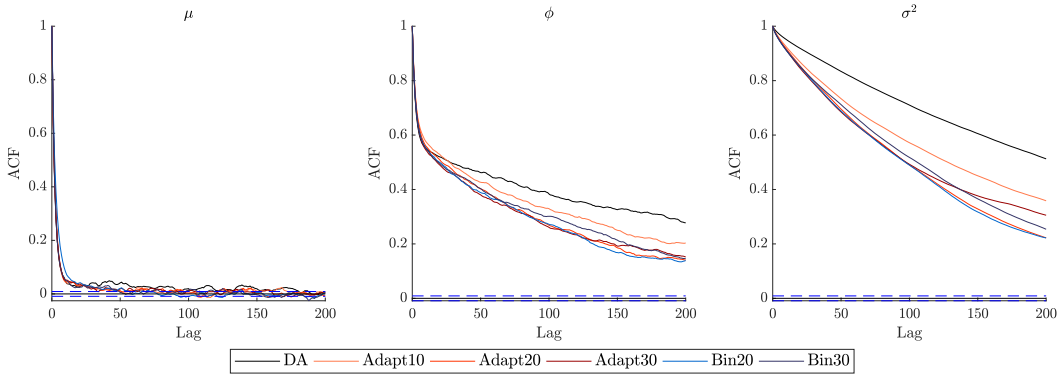
The top panel in Figure 7 presents the ACF plots for the parameters of the SV model and the bottom panel – for the SVML model. Tables 4 and 5 report the estimation results for selected imputed volatilities in the SV and SVML model, receptively. The top panel in Figure 8 presents the ACF plots for these volatilities for the SV model, while the bottom one for the SVML model.

Table 4: SV model: volatility posterior means, standard deviations and ESSs. Computing times (in seconds) in square brackets.

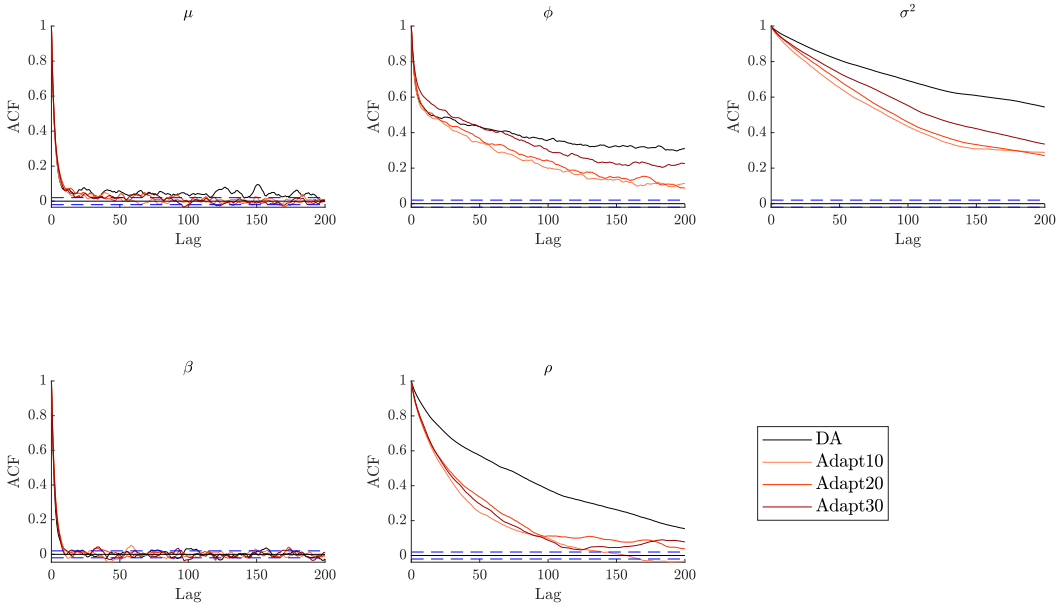
Method		h_{450}	h_{950}	h_{1450}	h_{1950}	h_{2450}	h_{2950}	h_{3450}	h_{3950}	h_{4450}
DA	Mean	0.974	0.259	-0.009	-0.006	0.257	1.083	0.011	0.733	-0.782
	(Std)	(0.444)	(0.417)	(0.404)	(0.492)	(0.466)	(0.449)	(0.485)	(0.457)	(0.454)
[112 s]	ESS	458	488	573	302	395	501	301	377	428
Adapt10	Mean	1.002	0.213	0.0104	-0.051	0.243	1.064	-0.022	0.753	-0.781
	(Std)	(0.434)	(0.431)	(0.432)	(0.492)	(0.469)	(0.445)	(0.479)	(0.453)	(0.482)
[1980 s]	ESS	1251	1490	1438	1246	1046	**1510**	**1426**	1338	1309
Adapt20	Mean	0.978	0.228	0.020	-0.040	0.241	1.059	-0.014	0.742	-0.805
	(Std)	(0.434)	(0.447)	(0.4358)	(0.504)	(0.467)	(0.442)	(0.489)	(0.448)	(0.467)
[2290 s]	ESS	**1667**	1224	1435	**1330**	1286	1460	1157	1290	1403
Adapt30	Mean	0.961	0.225	0.014	-0.052	0.240	1.101	0.014	0.745	-0.782
	(Std)	(0.439)	(0.438)	(0.431)	(0.499)	(0.457)	(0.442)	(0.482)	(0.436)	(0.474)
[2567 s]	ESS	1583	1638	1521	1270	**1410**	1608	1359	**1458**	1331
Fixed20	Mean	0.979	0.207	0.018	-0.048	0.251	1.069	0.010	0.756	-0.808
	(Std)	(0.437)	(0.428)	(0.428)	(0.490)	(0.471)	(0.441)	(0.480)	(0.448)	(0.482)
[1683 s]	ESS	1419	1243	**1633**	1223	1179	1456	1395	1217	1171
Fixed30	Mean	0.961	0.228	0.008	-0.048	0.238	1.075	-0.011	0.731	-0.819
	(Std)	(0.424)	(0.424)	(0.428)	(0.501)	(0.464)	(0.445)	(0.487)	(0.437)	(0.464)
[2057 s]	ESS	1152	**1760**	1367	1158	1396	1422	1292	1223	**1539**

Table 5: SVML model: volatility posterior means, standard deviations and ESSs. Computing times (in seconds) in square brackets.

Method		h_{450}	h_{950}	h_{1450}	h_{1950}	h_{2450}	h_{2950}	h_{3450}	h_{3950}	h_{4450}
DA	Mean	0.961	0.478	-0.163	0.011	0.448	1.026	-0.010	0.659	-0.963
	(Std)	(0.429)	(0.405)	(0.449)	(0.517)	(0.407)	(0.439)	(0.479)	(0.431)	(0.506)
[203 s]	ESS	651	421	540	277	621	634	507	349	346
Adapt10	Mean	0.955	0.454	-0.198	-0.022	0.413	1.024	-0.004	0.682	-0.984
	(Std)	(0.416)	(0.402)	(0.423)	(0.481)	(0.408)	(0.408)	(0.435)	(0.447)	(0.468)
[1511 s]	ESS	1551	1193	1184	1199	**1550**	**1860**	1278	1140	1089
Adapt20	Mean	0.920	0.470	-0.127	0.008	0.419	1.016	-0.055	0.677	-0.929
	(Std)	(0.400)	(0.405)	(0.450)	(0.489)	(0.419)	(0.421)	(0.463)	(0.435)	(0.493)
[2039 s]	ESS	1708	1280	1152	1009	1493	1622	1353	1231	971
Adapt30	Mean	0.915	0.409	-0.136	-0.034	0.402	1.030	-0.053	0.697	-0.948
	(Std)	(0.404)	(0.398)	(0.439)	(0.486)	(0.403)	(0.430)	(0.438)	(0.451)	(0.483)
[2497 s]	ESS	**1860**	**1334**	**1846**	**1227**	1487	1550	**1353**	**1322**	**1205**

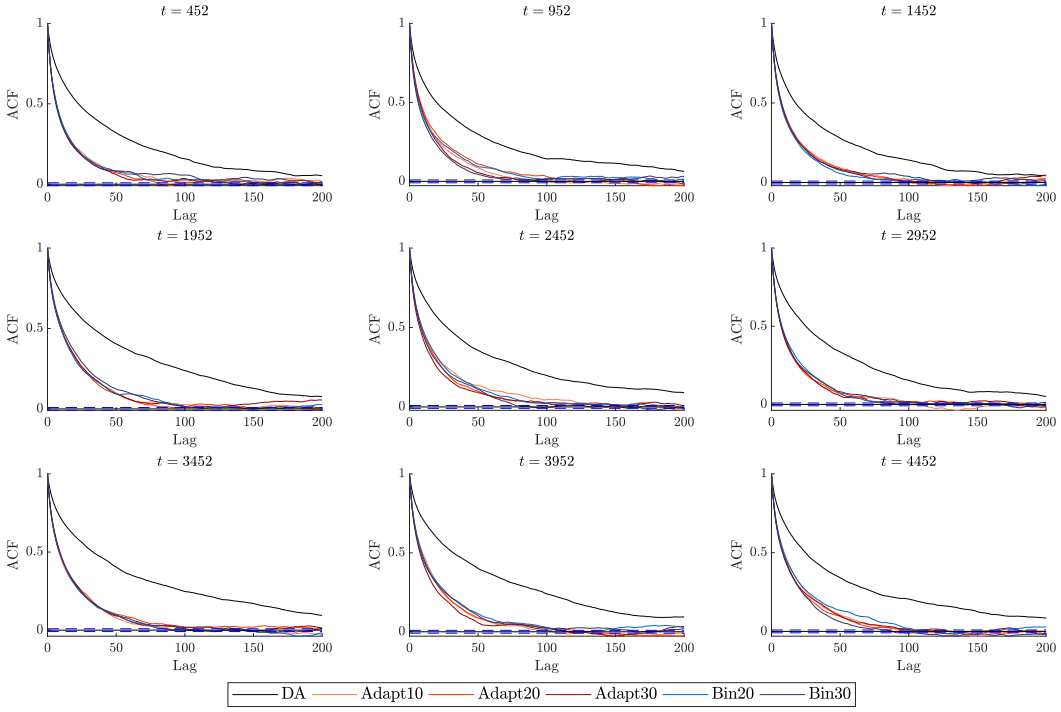


(a) SV model.

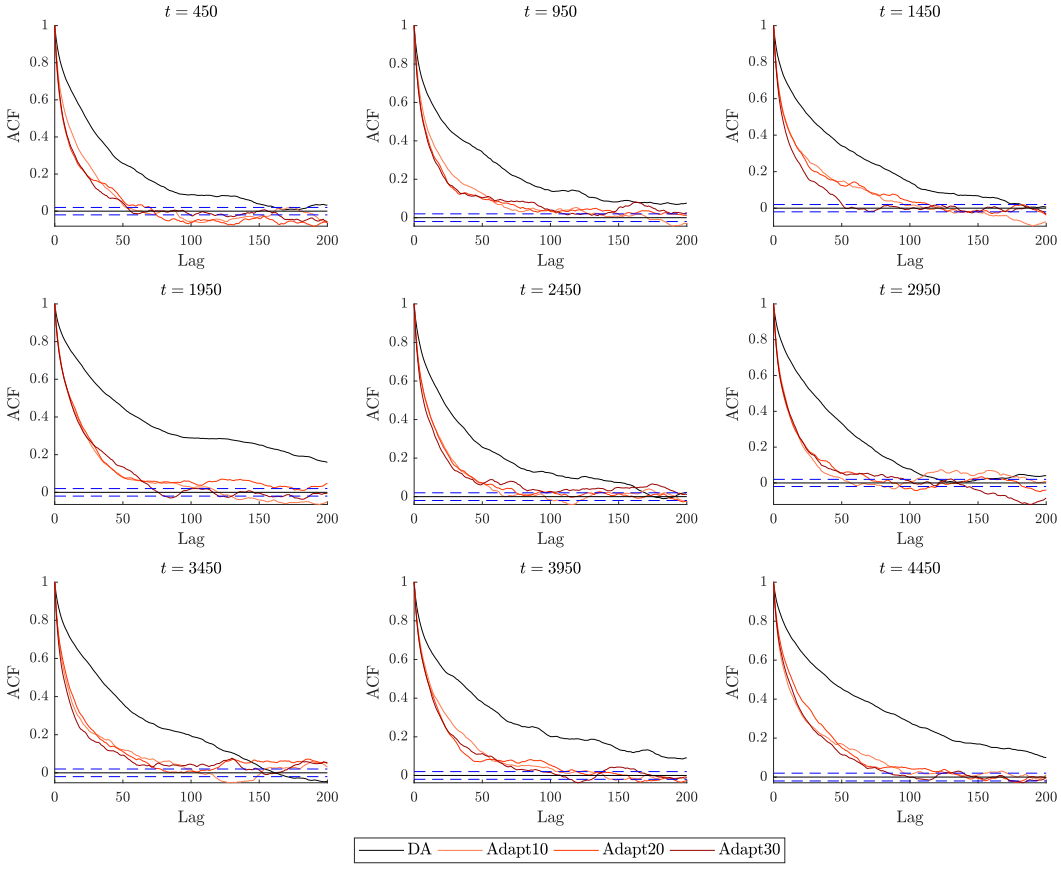


(b) SVML model.

Figure 7: SV and SVML model: ACF plots for parameters.



(a) SV model.



(b) SVML model.

Figure 8: SV and SVML model: ACF plots for selected volatilities.

Bibliography

Besbeas, P., Freeman, S. N., Morgan, B. J. T., and Catchpole, E. A. (2002), "Integrating Mark–Recapture–Recovery and Census Data to Estimate Animal Abundance and Demographic Parameters," *Biometrics*, 58, 540–547.

Brooks, S. P., King, R., and Morgan, B. J. T. (2004), "A Bayesian Approach to Combining Animal Abundance and Demographic Data," *Animal Biodiversity and Conservation*, 27, 515–529.

Chan, J. C. C. (2017), "The Stochastic Volatility in Mean Model with Time-Varying Parameters: An Application to Inflation Modeling," *Journal of Business & Economic Statistics*, 35, 17–28.

Durbin, J. and Koopman, S. J. (2012), *Time Series Analysis by State Space Methods: Second Edition*, Oxford Statistical Science Series, OUP Oxford.

Geyer, C. J. (2011), "Introduction to Markov Chain Monte Carlo," in *Handbook of Markov Chain Monte Carlo*, eds. S. Brooks, J. G. A. Gelman, and X. L. Meng, chap. 4, Chapman and Hall/CRC, pp. 3–48.

Goudie, R. J. B., Presanis, A. M., Lunn, D., Angelis, D. D., and Wernisch, L. (2018), "Joining and Splitting Models with Markov Melding," *Bayesian Analysis*.

International Union for Conservation of Nature (2018), "The IUCN Red List of Threatened Species: lapwing," http://www.iucnredlist.org/details/22693949/0. Accessed: 2018-08-05.

Jungbacker, B. and Koopman, S. J. (2007), "Monte Carlo Estimation for Nonlinear Non-Gaussian State Space Models," *Biometrika*, 94, 827–839.

Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. M. (1998), "Markov Chain Monte Carlo in Practice: A Roundtable Discussion," *The American Statistician*, 52, 93–100.

King, R. (2011), "Statistical Ecology," in *Handbook of Markov Chain Monte Carlo*, eds. S. Brooks, J. G. A. Gelman, and X. L. Meng, chap. 17, Chapman and Hall/CRC, pp. 419–447.

King, R., Morgan, B., Gimenez, O., and Brooks, S. (2010), *Bayesian Analysis for Population Ecology*, Chapman and Hall/CRC.

Koopman, S. J. and Uspensky, E. H. (2002), "The Stochastic Volatility in Mean Model: Empirical Evidence from International Stock Markets," *Journal of Applied Econometrics*, 17, 667–689.

Meyer, R. and Yu, J. (2000), "BUGS for a Bayesian Analysis of Stochastic Volatility Models," *Econometrics Journal*, 3, 198–215.

Pitt, M. K., Silva, R. S., Giordani, P., and Kohn, R. (2012), "On Some Properties of Markov Chain Monte Carlo Simulation Methods Based on the Particle Filter," *Journal of Econometrics*, 171, 134–151.

Robert, C. P. and Casella, G. (2004), *Monte Carlo Statistical Methods: Second Edition*, Springer Texts in Statistics.

The Royal Society for the Protection of Birds (2018), "The Red List of Conservation Concern: lapwing," https://www.rspb.org.uk/birds-and-wildlife/wildlife-guides/bird-a-z/lapwing/. Accessed: 2018-08-05.

Yu, J. (2005), "On Leverage in a Stochastic Volatility Models," *Journal of Econometrics*, 127, 165–178.

Zucchini, W., MacDonald, I. L., and Langrock, R. (2016), *Hidden Markov Models for Time Series: An Introduction Using R, Second Edition*, Monographs on Statistics and Applied Probability 150, CRC Press.