ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-7605-5/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICASSP39728.2021.9414875

PARTICLE GIBBS SAMPLING FOR REGIME-SWITCHING STATE-SPACE MODELS

Yousef El-Laham[ac], *Liu Yang*[a], *Heather J. Lynch*[bc], *Petar M. Djurić*[a], *Mónica F. Bugallo*[ac]

[a]Department of Electrical and Computer Engineering, [b]Department of Ecology and Evolution,
[c]Institute for Advanced Computational Sciences
Stony Brook University, Stony Brook, NY 11794-2350
{yousef.ellaham, liu.yang.2, heather.lynch, petar.djuric, monica.bugallo}@stonybrook.edu

ABSTRACT

Regime-switching state-space models (RS-SSMs) are an important class of statistical models that can be used to represent real-world phenomena. Unlike regular state-space models, RS-SSMs allow for dynamic uncertainty in the state transition and observations distributions, making them much more expressive. Unfortunately, there are no existing Bayesian inference techniques for joint estimation of regimes, states, and model parameters in generic RS-SSMs. In this work, we develop a particle Gibbs sampling algorithm for Bayesian learning in RS-SSMs. We demonstrate the proposed inference approach on a synthetic data experiment related to an ecological application, where the goal is in estimating the abundance and demographic rates of penguins in the Antarctic.

***Index Terms*—** particle filtering, regime-switching, state-space models, Gibbs sampling, ecological models

1. INTRODUCTION

State-space models (SSMs) are a popular class of statistical models that relate time-varying latent processes (called *states*) to time-varying collected measurements (called *observations*) [1]. Importantly, these models appear in a plethora of application areas including signal processing [2], neuroscience [3], finance [4], ecology [5], and epidemiology [6]. For example, SSMs are used in ecological applications to model the abundance (i.e., the population) of a species over time, where the states correspond to the true abundance and the observations correspond to raw noise-corrupted counts acquired from surveys [7, 8]. Stochastic filtering methods like Kalman filtering (KF) [9] and particle filtering (PF) [10], [11] provide a means for estimating the latent states of an SSM under the Bayesian framework, where the goal is in obtaining the posterior distribution of the states given the observations. Unfortunately, standard implementations of KF and PF cannot easily accommodate for the estimation of unknowns beyond the latent states, such as static parameters.

Numerous approaches for joint state and static parameter inference in SSMs can be found in the literature (see [12] for a survey). One important work is an offline estimation technique called particle Markov chain Monte Carlo (PMCMC) sampling [13]. PMCMC is an approximate Bayesian inference method that combines sequential Monte Carlo (SMC) with Markov chain Monte Carlo (MCMC) sampling. The simulation technique is used to generate a Markov chain whose stationary distribution is the targeted posterior distribution. The corresponding samples that comprise the Markov chain are taken as samples from the posterior distribution and allow for the computation of Bayesian estimators using a Monte Carlo average. A special case of PMCMC is particle Gibbs sampling (PGS) [13], [14], which generates the samples of the desired Markov chain by alternately sampling from the complete set of conditional posterior distributions of the unknowns in the model. To preserve the theoretical guarantees of PMCMC, PGS samples each state trajectory using a conditional SMC approach that conditions on the previously sampled trajectory, but this sometimes leads to a poorly mixed Markov chain. Modifications to PGS have been proposed to improve on the efficiency of the sampling scheme (e.g., ancestor sampling [15]).

An important extension to SSMs are regime-switching SSMs (RS-SSMs). Under regime-switching, the transition and observation distributions of the SSM can change from one time instant to the next, according to the value of a discrete-valued time-varying parameter referred to as the *regime*. Special cases of RS-SSMs include Markovian switching systems [16], where the time evolution of the regime follows a Markov process. The well-known problem of tracking a maneuvering target is an example of a Markovian switching system [17], where the dynamics of the target can change according to the state of its trajectory. Specialized stochastic filtering methods like the interacting multiple models filter [18] (and its PF counterpart [19]) have been proposed to deal with Markovian switching systems. In general, however, the evolution of the regime need not be Markovian. In [20], a PF algorithm that allowed for joint state and regime estimation in general RS-SSMs was proposed. Unfortunately, no methodology exists for the joint Bayesian estimation of regimes, states, and static parameter estimation for RS-SSMs.

In this work, we propose a novel PGS algorithm that allows for joint estimation in general RS-SSMs. To accommodate for changing regimes, we develop a novel PF algorithm to jointly sample state and regime trajectories, unlike classical PGS approaches, which only allow for sampling state trajectories and assume only a single regime. The proposed inference method is validated on synthetic data for an ecological application. The goal is in estimating the abundance and demographic rates of Antarctic penguins, where the population dynamics are modeled via an RS-SSM.

2. PROBLEM FORMULATION

We consider a state-space formulation that accommodates for dynamically switching regimes. Let $\mathbf{x}_t \in \mathbb{R}^{d_x}$ denote a latent state vector, $\mathbf{y}_t \in \mathbb{R}^{d_y}$ denote an observation vector, and $r_t \in \{1, \ldots, K\}$ denote the regime at time t, where K is finite and assumed to be known.

The authors would like to thank the support of NSF (CCF-1617986, GCR-2021002) and the Stony Brook Office for the Vice President of Research Seed Award. The Stony Brook Research Computing and Cyberinfrastructure and the Institute for Advanced Computational Science at Stony Brook University allowed for access to the high-performance SeaWulf computing system, supported by NSF (OAC-1531492).

Algorithm 1 PGS for Bayesian Learning in RS-SSMs

1: **Initialization:** $\boldsymbol{\gamma}^{(0)}, \boldsymbol{\theta}^{(0)}, \mathbf{x}_{0:T}^{(0)}, r_{1:T}^{(0)}$.
2: **for** $i = 1, \ldots, I$ **do**
3: Run Algorithm 2 with reference trajectories $\mathbf{x}_{0:T}^{(i-1)}$ and $r_{1:T}^{(i-1)}$ to jointly sample $\mathbf{x}_{0:T}$ and $r_{1:T}$:
$$\{\mathbf{x}_{0:T}^{(i)}, r_{1:T}^{(i)}\} \sim p(\mathbf{x}_{0:T}, r_{1:T}|\mathbf{y}_{1:T}, \boldsymbol{\gamma}^{(i-1)}, \boldsymbol{\theta}^{(i-1)}),$$
4: Sample the state transition/measurement parameters:
$$\boldsymbol{\theta}^{(i)} \sim p(\boldsymbol{\theta}|\mathbf{y}_{1:T}, \mathbf{x}_{0:T}^{(i)}, r_{1:T}^{(i)})$$
5: Sample the regime transition parameters:
$$\boldsymbol{\gamma}^{(i)} \sim p(\boldsymbol{\gamma}|\mathbf{y}_{1:T}, r_{1:T}^{(i)})$$
6: **end for**
7: **Return:** $\{\mathbf{x}_{0:T}^{(i)}, r_{1:T}^{(i)}, \boldsymbol{\theta}^{(i)}, \boldsymbol{\gamma}^{(i)}\}_{i=I_0+1}^{I}$

Consider the following set of equations:

$$r_t \sim p(r_t|r_{1:t-1}, \boldsymbol{\gamma}), \tag{1}$$
$$\mathbf{x}_t \sim p_{r_t}(\mathbf{x}_t|\mathbf{x}_{t-1}, \boldsymbol{\alpha}_{r_t}), \tag{2}$$
$$\mathbf{y}_t \sim p_{r_t}(\mathbf{y}_t|\mathbf{x}_t, \boldsymbol{\beta}_{r_t}), \tag{3}$$

for $t = 1, \ldots, T$, where $p(r_1|r_{1:0}, \boldsymbol{\gamma}) = p(r_1|\boldsymbol{\gamma})$ and $\mathbf{x}_0 \sim p(\mathbf{x}_0)$. The expression in (1) is called the *regime transition equation* and characterizes the regime dynamics, i.e., how the switching of regimes is governed. The regime transition equation is assumed to be parameterized by static parameters $\boldsymbol{\gamma}$. The expression in (2) is called the *state transition equation* and characterizes the time-evolution of the latent state vector $\mathbf{x}_t$ (assumed to be Markovian). Finally, the expression in (3) is called the *measurement equation* and relates the hidden states $\mathbf{x}_t$ to the observed measurements $\mathbf{y}_t$. It is important to emphasize that both the transition equation and the measurement equation depend on the regime r_t and the possibly unknown parameters of that regime (i.e., $\boldsymbol{\alpha}_{r_t}$ and $\boldsymbol{\beta}_{r_t}$). The goal is to jointly infer the regime trajectory $r_{1:T} = \{r_1, \ldots, r_T\}$, the latent state trajectory $\mathbf{x}_{0:T} = \{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_T\}$, and all unknown static parameters, i.e., $\boldsymbol{\gamma}$ and $\boldsymbol{\theta} = \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_K, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K\}$.[1]

3. PARTICLE GIBBS WITH REGIME SWITCHING

In this work, a novel PGS approach is proposed for Bayesian estimation of RS-SSMs as defined by (1)-(3). Similar to the standard Gibbs sampler, the proposed simulation scheme obtains samples from the targeted posterior by sampling from each of the conditional posteriors in an alternating fashion. In this implementation, the state and regime trajectories are jointly sampled using a newly proposed conditional regime-switching particle filtering (CRSPF) method. Then, the parameters that comprise the transition and measurement equations are sampled. Finally, the parameters of the regime transition equation are sampled. After I iterations of the algorithm, the first I_0 samples are discarded (burn-in) and the remaining samples are taken as samples from the joint posterior distribution. The proposed sampler is summarized in Algorithm 1.

3.1. Sampling State and Regime Trajectories

Algorithm 1 requires samples from the conditional posterior of the state and regime trajectories $p(\mathbf{x}_{0:T}, r_{1:T}|\mathbf{y}_{1:T})$, where for brevity in the presentation of the PF algorithm presented here, we have dropped the conditioning on $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ in the conditional posterior. For general RS-SSMs, drawing samples directly from this distribution is impossible and an alternative approach must be employed. Here, we consider a regime-switching PF (RSPF) method for simulating the desired trajectories.

A standard RSPF algorithm sequentially builds an approximation to $p(\mathbf{x}_{0:T}, r_{1:T}|\mathbf{y}_{1:T})$ using a set of weighted particle streams. At each time instant t, N regimes are sampled from a proposal distribution $s(r_t|r_{1:t-1})$, which depends on the history of regimes $r_{1:t-1}$:

$$r_t^{(n)} \sim s(r_t|r_{1:t-1}^{(n)}), \quad n = 1, \ldots, N. \tag{4}$$

After the regimes are sampled, states are sampled from another proposal distribution $q_{r_t}(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t)$, which depends on the previous state $\mathbf{x}_{t-1}$, the present regime r_t, and possibly the observations $\mathbf{y}_t$

$$\mathbf{x}_t^{(n)} \sim q_{r_t^{(n)}}(\mathbf{x}_t|\mathbf{x}_{t-1}^{(n)}, \mathbf{y}_t), \quad n = 1, \ldots, N. \tag{5}$$

Importance weights are computed for each of the samples in the set $\{\mathbf{x}_t^{(n)}, r_t^{(n)}\}_{n=1}^N$ according to:

$$\tilde{w}_t^{(n)} = \tilde{w}_{t-1}^{(n)} \frac{p_{r_t^{(n)}}(\mathbf{y}_t|\mathbf{x}_t^{(n)}) p_{r_t^{(n)}}(\mathbf{x}_t^{(n)}|\mathbf{x}_{t-1}^{(n)}) p(r_t^{(n)}|r_{1:t-1}^{(n)})}{q_{r_t^{(n)}}(\mathbf{x}_t^{(n)}|\mathbf{x}_{t-1}^{(n)}, \mathbf{y}_t) s(r_t^{(n)}|r_{1:t-1}^{(n)})} \tag{6}$$

for $n = 1, \ldots, N$. The weights are then normalized as $w_t^{(n)} = \tilde{w}_t^{(n)} / \sum_{j=1}^N \tilde{w}_t^{(j)}$ and the algorithm proceeds to the next time step. In the bootstrap implementation of RSPF, a resampling step is added and the proposals are selected as $s(r_t|r_{1:t-1}) = p(r_t|r_{1:t-1})$ and $q_{r_t}(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t) = p_{r_t}(\mathbf{x}_t|\mathbf{x}_{t-1})$, simplifying the weighting to $\tilde{w}_t^{(n)} = p_{r_t^{(n)}}(\mathbf{y}_t|\mathbf{x}_t^{(n)})$ for all samples. After T steps of the RSPF algorithm, we obtain a discrete random measure comprised of weighted state and regime trajectories, i.e., $\mathcal{X}_T = \{\{\mathbf{x}_{0:T}^{(n)}, r_{1:T}^{(n)}\}, w_T^{(n)}\}_{n=1}^N$. Finally, we can draw samples from the discrete random measure $\mathcal{X}_T$, where a particular state/regime trajectory pair $\{\mathbf{x}_{0:T}^{(j)}, r_{1:T}^{(j)}\}$ is sampled with probability $w_T^{(j)}$. Unfortunately, the simulation of state/regime trajectories cannot be done from a standard RSPF. In the original work on PGS, in order to not impede on the theoretical guarantees of PMCMC (i.e., the invariance of the generated Markov chain in the PGS algorithm), a modification was needed to be made to the sampling step. In particular, it was found that the PF used to sample state trajectories needed to be conditioned on a reference trajectory, where the resulting algorithm was called the conditional PF. One of the contributions of this work is a novel CRSPF scheme that can be integrated in the Gibbs sampler in Algorithm 1 to sample state and regime trajectories. Unlike traditional implementations, the new filter conditions on a reference trajectory for the states *and* a reference trajectory for the regimes.

The CRSPF algorithm is similar to the standard RSPF algorithm with the only difference being that exactly one state trajectory *and* one regime trajectory are specified *a priori*. These specified trajectories are called *reference trajectories* and are denoted by $\mathbf{x}'_{0:T}$ (state reference) and $r'_{1:T}$ (regime reference). In the CRSPF algorithm, $N-1$ of the particle streams are sampled as they normally are in the RSPF algorithm, while the Nth particle stream follows the path of the reference trajectories. Analogous to the work in [15], a modification to the CRSPF algorithm can be made to improve the mixing of the Markov chain when it is used in the PGS algorithm. At each time instant $t < T$, there is an additional resampling step called ancestor sampling, where the already generated particle streams are used to

[1] In RS-SSMs, some of the static unknowns can be shared across regimes. An example of such a system is considered in Section 4.

Algorithm 2 Bootstrap CRSPF with Ancestor Sampling

Input: Reference trajectories $\mathbf{x}'_{0:T}$ and $r'_{1:T}$

Initialization: Draw $N-1$ samples of the initial states

$$\mathbf{x}_0^{(n)} \sim p(\mathbf{x}_0), \quad n = 1, \dots, N-1,$$

and set $\mathbf{x}_0^{(N)} = \mathbf{x}'_0$. Set $\tilde{\mathbf{x}}_0^{(n)} = \mathbf{x}_0^{(n)}$ for $n = 1, \dots, N-1$. Compute the initial ancestor weights as

$$\rho_0^{(n)} = \frac{w_0^{(n)} p_{r'_1}(\mathbf{x}'_1 | \mathbf{x}_0^{(n)})}{\sum_{j=1}^{N} w_0^{(j)} p_{r'_1}(\mathbf{x}'_1 | \mathbf{x}_0^{(j)})}, \quad n = 1, \dots, N.$$

Set $\tilde{\mathbf{x}}_0^{(N)} = \mathbf{x}_0^{(j)}$ with probability $\rho_0^{(j)}$.

for $t = 1, \dots, T$ **do**

    Sample regimes: Draw $N-1$ samples of regime indexes

$$r_t^{(n)} \sim \begin{cases} p(r_t), & t = 1 \\ p(r_t | \tilde{r}_{1:t-1}^{(n)}), & t > 1 \end{cases}, \quad n = 1, \dots, N-1.$$

    Set $r_t^{(N)} = r'_t$. If $t > 1$, set $r_{1:t}^{(n)} = (\tilde{r}_{1:t-1}^{(n)}, r_t^{(n)})$ for all n.

    Sample states: Draw $N-1$ samples of the states conditioned on the drawn regimes

$$\mathbf{x}_t^{(n)} \sim p_{r_t^{(n)}}(\mathbf{x}_t | \tilde{\mathbf{x}}_{t-1}^{(n)}), \quad n = 1, \dots, N-1,$$

    and set $\mathbf{x}_t^{(N)} = \mathbf{x}'_t$. Set $\mathbf{x}_{1:t}^{(n)} = (\tilde{\mathbf{x}}_{0:t-1}^{(n)}, \mathbf{x}_t^{(n)})$ for all n.

    Weighting: Compute the weights as

$$\tilde{w}_t^{(n)} = p_{r_t^{(n)}}(\mathbf{y}_t | \mathbf{x}_t^{(n)}), \quad n = 1, \dots, N,$$

    and normalize them as $w_t^{(n)} = \tilde{w}_t^{(n)} / \sum_{j=1}^{N} \tilde{w}_t^{(j)}$ for all n.

    if $t < T$ **then**

        Resampling: Resample particles $\{\tilde{\mathbf{x}}_t^{(n)}, \tilde{r}_t^{(n)}\}_{n=1}^{N-1}$ by sampling $N-1$ times with replacement from $\{\mathbf{x}_t^{(n)}, r_t^{(n)}\}_{n=1}^{N}$ with probabilities equal to the weights $\{w_t^{(n)}\}_{n=1}^{N}$.

        Ancestor sampling: Compute the ancestor weights using (7). Set $(\tilde{\mathbf{x}}_t^{(N)}, \tilde{r}_t^{(N)}) = (\mathbf{x}_t^{(j)}, r_t^{(j)})$ with probability $\rho_t^{(j)}$.

    end if

end for

Set $\{\mathbf{x}^\star_{0:T}, r^\star_{1:T}\} = \{\mathbf{x}_{0:T}^{(j)}, r_{1:T}^{(j)}\}$ with probability $w_T^{(j)}$.

Return: $\{\mathbf{x}^\star_{0:T}, r^\star_{1:T}\}$

assign suitable artificial histories to the reference trajectories $\mathbf{x}'_{t:T}$ and $r'_{t:T}$. This resampling step is done to improve the sampling efficiency of the CRSPF algorithm, so that the sampled latent state and regime trajectories do not repeatedly collapse to the reference trajectories $\mathbf{x}'_{t:T}$ and $r'_{t:T}$. The ancestor weights $\{\rho_t^{(n)}\}_{n=1}^{N}$ used to resample from the past trajectories can be computed as:

$$\rho_t^{(n)} = \frac{\tilde{w}_t^{(n)} p_{r'_{t+1}}(\mathbf{x}'_{t+1} | \mathbf{x}_t^{(n)}) p(r'_{t+1} | r_{1:t}^{(n)})}{\sum_{j=1}^{N} \tilde{w}_t^{(j)} p_{r'_{t+1}}(\mathbf{x}'_{t+1} | \mathbf{x}_t^{(j)}) p(r'_{t+1} | r_{1:t}^{(j)})}, \tag{7}$$

for $n = 1, \dots, N$. A bootstrap implementation of the resulting conditional regime-switching particle filter (CRSPF) with ancestor sampling is summarized in Algorithm 2.

3.2. Sampling the Transition/Observation Parameters

Under different regimes, the variation of state transition and measurement models is reflected in the parameters $\boldsymbol{\alpha}_{r_t}$ and $\boldsymbol{\beta}_{r_t}$, respectively. In order to sample the parameter vector $\boldsymbol{\theta}$ at each iteration, we need to derive the conditional distributions of every single one of parameters. Obviously, for each different SSM, a particular derivation is necessary. Let us consider $p(\boldsymbol{\alpha}_{r_t} | \boldsymbol{\theta}_{-\boldsymbol{\alpha}_{r_t}}, \mathbf{y}_{1:T}, \mathbf{x}_{0:T}, r_{1:T})$ and $p(\boldsymbol{\beta}_{r_t} | \boldsymbol{\theta}_{-\boldsymbol{\beta}_{r_t}}, \mathbf{y}_{1:T}, \mathbf{x}_{0:T}, r_{1:T})$. Since the transition parameters are only related to latent states and regimes, their conditional distributions can be simplified to $p(\boldsymbol{\alpha}_{r_t} | \boldsymbol{\alpha}_{-r_t}, \mathbf{x}_{0:T}, r_{1:T})$. Similarly, the conditional distributions of observation parameters have the simpler form $p(\boldsymbol{\beta}_{r_t} | \boldsymbol{\beta}_{-r_t}, \mathbf{y}_{1:T}, \mathbf{x}_{0:T}, r_{1:T})$. Notice that both transition and observation parameters might be shared across regimes in most of regime-switching SSMs and in that case the conditional distributions can be further simplified as $p(\boldsymbol{\alpha} | \mathbf{x}_{0:T})$ and $p(\boldsymbol{\beta} | \mathbf{y}_{1:T}, \mathbf{x}_{0:T})$. There is also a chance that we are not able to sample parameters directly from their conditional distributions. In this situation, one must seek a numerical approximation, such as Metropolis-Hastings and importance sampling to obtain approximate samples.

3.3. Sampling the Regime Switching Parameters

There are many different possible regime transition models. Practitioners should choose the regime dynamics such that the conditional distribution for $\boldsymbol{\gamma}$ can easily be derived. In our previous work [20], we discussed three regime dynamics from simple to complex (e.g. independent, Markovian, and Pólya urn). In this paper, we will consider that the finite regimes are independent from each other. When the number of possible regimes is $K \geq 2$, we can develop a multinoulli switching model to govern the regime transition and propose the conditional posterior of regime transition parameters $\boldsymbol{\gamma}$. Under multinoulli switching, the regime transition equation is a multinoulli distribution with probability parameters $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_K]^\intercal$, i.e.,

$$p(r_t | \boldsymbol{\gamma}) = \prod_{k=1}^{K} \gamma_k^{\mathbb{1}(r_t = k)}, \quad r_t \in \{1, \dots, K\}, \tag{8}$$

with $\gamma_k \geq 0$ for all k, and $\sum_{k=1}^{K} \gamma_k = 1$. Since the regimes at each time instant are assumed to be independent, the joint distribution $p(r_{1:T} | \boldsymbol{\gamma})$ is

$$p(r_{1:T} | \boldsymbol{\gamma}) = \prod_{k=1}^{K} \gamma_k^{c_k}, \tag{9}$$

where $c_k = \sum_{t=1}^{T} \mathbb{1}(r_t = k)$ is the total number of occurrences of the kth regime and we define $\mathbf{c} = [c_1, \dots, c_K]^\intercal$. If we assume a Dirichlet prior over the probability vector $\boldsymbol{\gamma}$, i.e., $p(\boldsymbol{\gamma}) = \mathrm{Dirichlet}(\boldsymbol{\xi}_0)$ with $\boldsymbol{\xi}_0 = [\xi_{0,1}, \dots, \xi_{0,K}]^\intercal$ and $\xi_{0,k} > 0$ for all k, then by conjugacy, the conditional posterior of $\boldsymbol{\gamma}$ is also Dirichlet and is given by

$$p(\boldsymbol{\gamma} | r_{1:T}) = \mathrm{Dirichlet}(\boldsymbol{\xi}) \tag{10}$$

where $\boldsymbol{\xi} = \boldsymbol{\xi}_0 + \mathbf{c}$. Therefore, we can straightforwardly sample the model transition parameters under multinoulli switching.

As we can see, when $K = 2$, the multinoulli switching coincides with the Bernoulli switching, where there are only two possible regimes, i.e., $r_t \in \{1, 2\}$ for all t. Thus, the joint distribution $p(r_{1:T} | \gamma)$ can straightforwardly be determined as

$$p(r_{1:T} | \gamma) = \prod_{t=1}^{T} p(r_t | \gamma) = \gamma^{c_1} (1 - \gamma)^{T - c_1}, \tag{11}$$

where $c_1 = \sum_{t=1}^{T} \mathbb{1}(r_t = 1)$ is the total number of occurrences of regime 1 and $c_2 = T - c_1$ is the total number of occurrences of regime 2. Suppose that a prior distribution $p(\gamma) = \mathrm{Beta}(a_0, b_0)$ where $a_0, b_0 > 0$, by conjugate prior it turns out that the conditional

K	ψ	ϕ_1	ϕ_2	ϕ_3	$\sigma_s^2 \times 10^3$	$\sigma_c^2 \times 10^3$	MSE
1	(0.782, 0.799)	(0.430, 0.499)	-	-	(3.50, 14.3)	(26.5, 67.4)	72176
2	(0.787, 0.792)	(0.326, 0.342)	(0.518, 0.536)	-	(0.669, 1.83)	(0.644, 1.97)	1513
3	(0.787, 0.792)	(0.227, 0.342)	(0.331, 0.532)	(0.519, 0.725)	(0.671, 1.84)	(0.548, 1.92)	1487
	0.79	0.33	0.52	-	0.9	0.9	-

Table 1: 95% Bayesian CIs of parameters and MSE in predictive mean of states.

posterior distribution is also a Beta distribution, i.e.,

$$p(\gamma|r_{1:T}) = \mathrm{Beta}(a, b), \tag{12}$$

where $a = a_0 + c_1$ and $b = b_0 + c_2$.

4. EXAMPLE: SWITCHING POPULATION MODEL

We validate our method with a synthetic data experiment for an ecological application, where an RS-SSM is used to model a time series of abundance of Antarctic penguins. Let $S_{j,t}$ and $C_{j,t}$ denote the number of stage j adult penguins and penguin chicks at year t, respectively, and $S_t = \sum_{j=3}^{J} S_{j,t}$ and $C_t = \sum_{j=1}^{J-2} C_{j,t}$ represent the total number of adults and chicks irrespective of age. Consider the following multinoulli switching transition process for modeling the evolution of penguin populations over T years:

$$r_t \sim \mathrm{Multinoulli}(\boldsymbol{\gamma}), \tag{13}$$
$$S_{1,t} \sim \mathrm{Binomial}(0.5C_{t-1}, \psi), \tag{14}$$
$$S_{j,t} \sim \mathrm{Binomial}(S_{j-1,t-1}, \psi),\ j = 2, \ldots, J, \tag{15}$$
$$S_{J,t} \sim \mathrm{Binomial}(S_{J-1,t-1} + S_{J,t-1}, \psi), \tag{16}$$
$$C_{j,t} \sim \mathrm{Binomial}(2S_{j,t}, \phi_{r_t}),\ j = 1, \ldots, J-2, \tag{17}$$

where ψ denotes the survival rate and ϕ_{r_t} denotes the reproductive rate of regime r_t (hence, we assume that the reproductive rates switch). Importantly, penguins cannot be aged in the field and census counts capture only the total number of adults and chicks, irrespective of age. Errors in counting are assumed proportional to abundance and we model the total number in each class as:

$$\tilde{S}_t \sim \mathcal{N}(S_t, \sigma_s^2 S_t^2), \tag{18}$$
$$\tilde{C}_t \sim \mathcal{N}(C_t, \sigma_c^2 C_t^2). \tag{19}$$

The unknowns of the model are the regimes $r_t \in \{1, \ldots, K\}$, the states $\mathbf{x}_t = [S_{1,t}, \ldots, S_{J,t}, C_{1,t}, \ldots, C_{J-2,t}]^\intercal$, the measurement/transition parameters $\boldsymbol{\theta} = [\psi, \phi_1, \ldots, \phi_K, \sigma_s^2, \sigma_c^2]^\intercal$ and the regime transition probabilities $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_K]$. Assuming $J = 5$ stages, we simulate a $T = 40$ year synthetic dataset under the following parameter settings: $K = 2$ possible regimes, $\psi = 0.79$, $\phi_1 = 0.33, \phi_2 = 0.52, \sigma_s^2 = \sigma_c^2 = 0.0009, \gamma_1 = 0.3$, and $\gamma_2 = 0.7$. Under the assumption that we are unaware of the true number of possible regimes, we run our inference algorithm varying the total number of possible regimes ($K = 1$, $K = 2$, and $K = 3$) for a total of $I = 50000$ iterations, using a burn-in period of $I_0 = 25000$, and $N = 500$ particles in the CRSPF algorithm. We assume vague priors for each parameter and preserve the identifiability of reproductive rates by enforcing the monotonicity constraint $\phi_1 < \phi_2 < \cdots < \phi_K$.

The results of the experiment are summarized in Table 1, where we show Bayesian credible intervals (CIs) of the posteriors of the transition/measurement parameters and the mean squared error (MSE) in the posterior mean of the states. We can see that when only a single regime is used, the reproductive rate falls between the simulated values of 0.33 and 0.52, and the model explains the "switching" in the simulated time-series by an overestimation in the variance parameters σ_s^2 and σ_c^2, but this leads to a high MSE in the state estimation. When the correct number of regimes is assumed ($K = 2$), we can see that the CIs for all parameters contain the true values, indicating that the inference method was successful in recovering the system parameters. Since the jumps in the time-series are now appropriately explained by switching in the reproductive rate parameters, there is now a much lower MSE in the state estimation. Finally, when more than enough regimes are assumed ($K = 3$), we arrive at an interesting result. Mainly, the intermediate switching parameter ϕ_2 is multimodal, with modes close to the true reproductive rates of 0.33 and 0.52 (see Fig. 1). Therefore, for switching models like this, we can use multimodality as a diagnostic for choosing the total number of possible regimes. It is also important to note that despite this multimodality, the MSE in the state estimate is comparable than that of the model assuming the correct number of possible regimes.

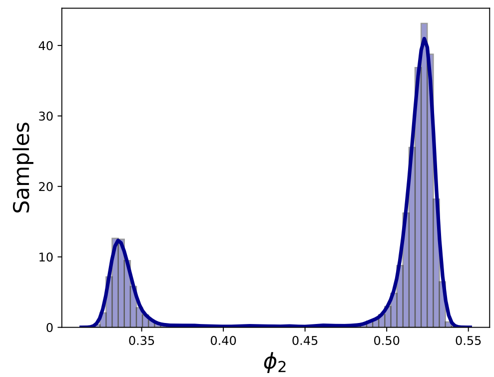

Fig. 1: Multimodality in reproductive rates for $K = 3$ regimes.

5. CONCLUSIONS AND FUTURE WORK

In this work, we presented a novel inference algorithm for Bayesian learning in regime-switching state-space models (RS-SSMs). The proposed method is a particle Gibbs sampler that employs a new conditional regime-switching particle filtering method for jointly sampling state and regime trajectories. The conditional posterior distributions for the regime switching parameters are also derived in the case of independent switching, while the derivation of all other parameters are specific to the application of interest. The sampling method was validated on a synthetic data experiment involving an ecological application, where the goal was to estimate the abundance and demographic rates of penguins whose dynamics were modeled with an RS-SSM. Open questions regrading how the number of regimes is chosen remain to be answered. Future work will consider non-parametric extensions of the proposed algorithm, possibly based on Dirichlet process mixture models, in which the number of regimes is assumed unknown and therefore a part of the inference problem.

6. REFERENCES

[1] C. K. Carter and R. Kohn, "On Gibbs sampling for state space models," *Biometrika*, vol. 81, no. 3, pp. 541–553, 09 1994.

[2] T. Schon, F. Gustafsson, and P. . Nordlund, "Marginalized particle filters for mixed linear/nonlinear state-space models," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2279–2289, 2005.

[3] L. Paninski, Y. Ahmadian, D. G. Ferreira, S. Koyama, K. R. Rad, M. Vidne, J. Vogelstein, and W. Wu, "A new look at state-space models for neural data," *Journal of Computational Neuroscience*, vol. 29, pp. 107–126, 2010.

[4] D. Creal, "A survey of sequential monte carlo methods for economics and finance," *Econometric Reviews*, vol. 31, no. 3, pp. 245–296, 2012.

[5] M. Dowd, "A sequential Monte Carlo approach for marine ecological prediction," *Environmetrics: The official journal of the International Environmetrics Society*, vol. 17, no. 5, pp. 435–455, 2006.

[6] C. Jégat, F. Carrat, C. Lajaunie, and H. Wackernagel, "Early detection and assessment of epidemics by particle filtering," in *geoENV VI–Geostatistics for Environmental Applications*, pp. 23–35. Springer, 2008.

[7] J.S. Clark and O.N. Bjørnstad, "Population time series: Process variability, observation errors, missing values, lags, and hidden states," *Ecology*, vol. 85, no. 11, pp. 3140–3150, 2004.

[8] G. Tavecchia, P. Besbeas, T. Coulson, B.J.T. Morgan, and T.H. Clutton-Brock, "Estimating population size and hidden demographic parameters with state space modeling," *The American Naturalist*, vol. 173, no. 6, pp. 722–733, 2009.

[9] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[10] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *IEE Proceedings F (radar and signal processing)*. IET, 1993, vol. 140, pp. 107–113.

[11] P. M. Djurić, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Miguez, "Particle filtering," *IEEE Signal Processing Magazine*, vol. 20, no. 5, pp. 19–38, 2003.

[12] N. Kantas, A. Doucet, S. S. Singh, and J. M. Maciejowski, "An overview of sequential Monte Carlo methods for parameter estimation in general state-space models," *IFAC Proceedings Volumes*, vol. 42, no. 10, pp. 774 – 785, 2009, 15th IFAC Symposium on System Identification.

[13] C. Andrieu, A. Doucet, and R. Holenstein, "Particle Markov chain Monte Carlo methods," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 72, no. 3, pp. 269–342, 2010.

[14] N. Chopin and S. Singh, "On particle Gibbs sampling," *Bernoulli*, vol. 21, pp. 1855–1883, 2015.

[15] F. Lindsten, M. I. Jordan, and T. B. Schön, "Particle Gibbs with ancestor sampling," *Journal of Machine Learning Research*, vol. 15, no. 63, pp. 2145–2184, 2014.

[16] C. B. Chang and M. Athans, "State estimation for discrete systems with switching parameters," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-14, no. 3, pp. 418–425, 1978.

[17] R. Karlsson and N. Bergman, "Auxiliary particle filters for tracking a maneuvering target," in *Proceedings of the 39th IEEE Conference on Decision and Control (Cat. No.00CH37187)*, 2000, vol. 4, pp. 3891–3895 vol.4.

[18] H. A. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE transactions on Automatic Control*, vol. 33, no. 8, pp. 780–783, 1988.

[19] Y. Boers and J. N. Driessen, "Interacting multiple model particle filter," *IEE Proceedings-Radar, Sonar and Navigation*, vol. 150, no. 5, pp. 344–349, 2003.

[20] Y. El-Laham, L. Yang, P. M. Djuric, and M. F. Bugallo, "Particle filtering under general regime switching," *arXiv preprint arXiv:2009.04551*, 2020.