Supplementary 2: Sequential Monte Carlo methods for data assimilation problems

Kwaku Peprah Adjei[1,2], Robert B. O'Hara[1,2], Nick Isaac[3], Daina Bowler[3], Rob Cooke[3]

[1]Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway

[2]Centre of Biodiversity Dynamics, Norwegian University of Science and Technology, Trondheim, Norway

[3]Center of Ecology and Hydrology, Wallingford, UK

1 Introduction

This document provides some background information needed to understand the Algorithms implemented in the main paper. We first provide information about state space models sequential Monte Carlo methods, and then proceed to describe the bootstrap and auxiliary particle filters as well as the particle MCMC with their algorithms. It must be stated that some of the equations and terminolgies are repeated to enhance the smooth understanding of this document, with fewer referrals to the main paper.

2 State space models

We assume we can obtain a (multivariate) observed time series data denoted by $y_{1:T} = \{y_1, y_2, \dots, y_T\}$, where y_t is a $(k \times 1)$ observed vector. These observations depend on latent (unobserved but in interest) states $x_{1:T} = \{x_1, x_2, \dots, x_T\}$. These latent states are assumed to have a first order Markov structure (the latent state at time t depends on latent state at time $t-1$ only) and the observations at each time t, y_t, given the latent state at that time x_t are independent of previous observations and states.

In summary, we have the following information for the SSM framework:

$$\begin{aligned} \text{Intial state distribution}: \quad & p(x_0|\theta); \quad t = 0 \\ \text{State model}: \quad & p(x_t|x_{t-1}, \theta); \quad t = 1, 2, \dots, T \\ \text{Observation model}: \quad & p(y_t|x_t, \theta); \quad t = 1, 2, \dots, T \end{aligned} \tag{1}$$

where θ are top-level parameters (assumed to be constant in the SSM defined in equation (1)).

Furthermore, we assume that we have already fitted a SSM, either with MCMC or SMC approaches, to the observed data from time $t = 1$ to $t = t_r$, where $t_r < T$. The posterior samples from the fitted SSMs are organised in a number of interations $\times$ t_r matrix of latent state posterior samples and number of interations$\times$number of top level nodes matrix of top-level parameter posterior samples as shown in equation (2).

$$t=1 \qquad \text{Old observation: } y_1,\dots,y_{t_r} \qquad t=t_r \qquad \text{New observation: } y_{t_r+1},\dots,y_T \qquad t=T$$

$$\underbrace{\begin{bmatrix} \hat{x}_1^{(1)} & \hat{x}_2^{(1)} & \dots & \hat{x}_{t_r}^{(2)} \\ \hat{x}_1^{(2)} & \hat{x}_2^{(2)} & \dots & \hat{x}_{t_r}^{(2)} \\ \vdots & \vdots & \dots & \vdots \\ \hat{x}_1^{(n.iter)} & \hat{x}_2^{(n.iter)} & \dots & \hat{x}_{t_r}^{(n.iter)} \end{bmatrix} \begin{bmatrix} \hat{\theta}_r^{(1)} \\ \hat{\theta}_r^{(2)} \\ \vdots \\ \hat{\theta}_r^{(n.iter)} \end{bmatrix}}_{\text{MCMC used for State Space models}} \qquad \underbrace{\begin{bmatrix} x_{t_r+1}^{(1)} & \dots & x_T^{(2)} \\ x_{t_r+1}^{(2)} & \dots & x_T^{(2)} \\ \vdots & \dots & \vdots \\ x_{t_r+1}^{(n.iter)} & \dots & x_T^{(n.iter)} \end{bmatrix} \begin{bmatrix} \theta_{upd}^{(1)} \\ \theta_{upd}^{(2)} \\ \vdots \\ \theta_{upd}^{(n.iter)} \end{bmatrix}}_{\text{SMC used to update latent states and parameters}} \tag{2}$$

where $\hat{\theta}_r$ and $\hat{\mathbf{x}}_{1:t_r} = \{\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_{t_r}\}$ are posterior samples of parameters and latent states from fitted SSM to the observations $y_{1:t_r} = y_1, y_2, \dots, y_{t_r}$; θ_{upd} and $\mathbf{x}_{t_r+1:T} = \{\mathbf{x}_{t_r+1}, \dots, \mathbf{x}_T\}$ are posterior samples of parameters and latent states we are interested in estimating after the new stream of observations $y_{t_r+1:T} = y_{t_r+1}, y_{t_r+2}, \dots, y_T$ are obtained.

3 Sequential Monte Carlo (SMC) methods

The SMC methods use sequential importance sampling (SIS) technique to estimate the filtering distributions (Doucet, De Freitas, and Gordon 2001; Michaud et al. 2021). At each time step t, the latent state x_t is proposed from the previous state x_{t-1} from a proposal distribution or importance function, $\pi(x_t|x_{t-1}, y_{1:t}, \theta)$, and posterior samples of x_t are drawn from the proposed samples using importance weights w_t:

$$w_t^{(i)} \propto \frac{p(x_t^{(i)}|x_{t-1}^{(i)}, y_{1:t}, \theta)}{\pi(x_t^{(i)}|x_{t-1}^{(i)}, y_{1:t}, \theta)} \tag{3}$$

and iteratively as:

$$w_t^{(i)} \propto w_{t-1}^{(i)} \frac{p(x_t^{(i)}|x_{t-1}^{(i)},\theta)p(y_{1:t}|x_t^{(i)})}{\pi(x_t^{(i)}|x_{t-1}^{(i)},y_{1:t},\theta)} \tag{4}$$

for $i = 1, 2, \dots, M$ particles.

Since the importance function chosen for SIS is critical to the performance of the SMC method (Arulampalam et al. 2002; Doucet, De Freitas, and Gordon 2001; Michaud et al. 2021), the prior distribution of the latent states are chosen as the importance function. In this case, the importance weights in equation (4) simplifies to:

$$w_t^{(i)} \propto w_{t-1}^{(i)} p(y_t|x_t^{(i)},\theta); \tag{5}$$

for $i = 1, 2, \dots, M$ particles. See Doucet, De Freitas, and Gordon (2001) for the details of the simplification.

This suggests that from time steps $t = t_{r+1}$ to $t = T$ (where we are have new data), we only need information about the weights at $t = t_r$ and the SSM distributions in equation (1) (which are assumed to be known as shown in equation (2)), we can use any of the SMC algorithms, such as bootstrap particle filter to be briefly discussed below, to estimate posterior distribution of latent states and update the top - level parameters. The saved posterior samples of latent state for time steps $t = 1, 2, \dots, t_r$ are assumed to be equally weighted samples (i.e. they have equal weights and we choose $w_t^{(i)} = 1, \forall i$). In summary, our proposed framework uses the following information for generating the posterior samples of the latent states:

$$\begin{cases} w_t^{(i)} = 1 & x_t^{(i)} \text{ from RM} & \text{for } 1 \leq t \leq t_r \\ w_t^{(i)} \text{ defined by equations (3)} - \text{(5)} & x_t^{(i)} \text{ from SMC} & \text{for } t_{r+1} \leq t \leq T \end{cases} \tag{6}$$

where RM denotes a reduced SSM fitted to the observed data $y_{1:t_r}$ using either MCMC or any SMC method to be discussed later.

3.0.1 Bootstrap particle filter

The bootstrap particle filter (hereafter, BPF) re-samples with replacement the M particles $(x_{0:t}^{(i)}; i = 1, 2, \ldots, M)$ from the set of proposed samples $(\tilde{x}_{0:t}^{(i)}; i = 1, 2, \ldots, M)$ according the importance weights $(w_t^{(i)}; i = 1, 2, \ldots, M)$ defined in equations (3) to (5). This re-sampling approach mitigates the particle degeneracy problem, where unimportant particles are propagated through time (Arulampalam et al. 2002; Doucet, De Freitas, and Gordon 2001).

To implement the changes proposed in this paper; the following changes were made to the bootstrap PF algorithm implemented in `nimbleSMC` (Michaud et al. 2021) as shown in Algorithm (1):

3.0.2 Auxiliary particle filter

The auxiliary particle filter (APF hereafter; Pitt and Shephard (1999)) uses the new observation to generate more likely states by including an additional "look-ahead step" (Michaud et al. 2021). At each time step t, this is done by first sampling M particles from weights from time $t-1$ which are calculated using a rough estimate of the likelihood of the current data given the particles from the previous time point (Michaud et al. 2021). The sampled particles are then propagated in time by the proposal distribution $\pi(x_t^{(i)}|\tilde{x}_{t-1}^{(i)}, y_t, \theta)$ and reweighted again using another weights $w_t^{(i)}$, for $i = 1, 2, \ldots, M$ particles. See Pitt and Shephard (1999) and Doucet, De Freitas, and Gordon (2001) for detailed explanation to the APF algorithm.

To implement the changes proposed in this paper; the following changes were made to the auxiliary PF algorithm implemented in `nimbleSMC` (Michaud et al. 2021) as shown in Algorithm (2):

Algorithm 1 Bootstrap filter with constant top-level nodes θ

for t in $1 : t_r$ and i in $1 : n.iter$ **do**
  for m in $1 : M$ **do**
    Set $w_t^{(m)} := 1$; $x_t^{(m)} := \hat{x}_t^{(i)}$ (the i^{th} posterior sample of latent state in equation (2))
  end for
end for
for t in $t_{r+1} : T$ **do**
  for m in $1 : M$ **do**
    Generate $\tilde{x}_t^{(m)} \sim q(x_t | x_{t-1}^{(m)}, y_t)$

    Calculate unnormalised weight $w_t^{(m)} = \frac{f(\tilde{x}_t^{(m)} | x_{t-1}^{(m)}) g(y_t | \tilde{x}_t^{(m)})}{q(x_t | x_{t-1}^{(m)}, y_t)} \pi_{t-1}^{(m)}$
  end for
  for m in $1 : M$ **do**
    Normalize $w_t^{(m)}$ as $\pi_t^{(m)} := \frac{w_t^{(m)}}{\sum_{i=1}^{M} w_t^{(m)}}$
  end for
  for m in $1 : M$ **do**
    Sample an index j from the set of $i, \dots, M$ with probabilities $\{\pi_t^{(m)}\}_{m=1}^{M}$
    Set $x_t^{(m)} = \tilde{x}_t^{(m)}$
    $\pi_t^{(m)} = \frac{1}{M}$
  end for
  Calculate $\tilde{p}(y_{t|1:t-1}) = \frac{1}{M} \sum_{m=1}^{M} w_t^{(m)}$
end for

Algorithm 2 Auxiliary filter with constant top-level nodes θ

for t in $1 : t_r$ and i in $1 : n.iter$ **do**
  for m in $1 : M$ **do**
    Set $w_t^{(m)} := 1$; $x_t^{(m)} := \hat{x}_t^{(i)}$ (the i^{th} posterior sample of latent state in equation (2))
  end for
end for
for t in $t_{r+1} : T$ **do**
  for m in $1 : M$ **do**
    Generate $\tilde{x}_{t|t-1}^{(m)}$ from either $\tilde{x}_{t|t-1}^{(m)} \sim f(x_t|x_{t-1}^{(m)})$ or $\tilde{x}_{t|t-1}^{(m)} = E(x_t|x_{t-1}^{(m)})$

    Calculate $\hat{p}(y_t|x_{t-1}^{(m)}) = p(y_t|\tilde{x}_{t-1}^{(m)})$

    Calculate unnormalised weight $w_{t|t-1}^{(m)} = \pi_{t-1}^{(m)}\hat{p}(y_t|x_{t-1}^{(m)})$

  end for
  for m in $1 : M$ **do**
    Normalize $w_{t|t-1}^{(m)}$ as $\pi_{t|t-1}^{(m)} := \frac{w_{t|t-1}^{(m)}}{\sum_{i=1}^{M} w_{t|t-1}^{(m)}}$
  end for
  for m in $1 : M$ **do**
    Sample an index j_m from the set of $1, \dots, M$ with probabilities $\{\pi_{t|t-1}^{(m)}\}_{m=1}^{M}$
    Set $\tilde{x}_{t-1}^{(m)} = x_t^{(j_k)}$
    Generate $x_t^{(m)} \sim q(x_t|\tilde{x}_{t-1}^{(m)}, y_t)$
    Calculate unnormalised weight $w_t^{(m)} = \frac{f(\tilde{x}_t^{(m)}|x_{t-1}^{(m)})g(y_t|\tilde{x}_t^{(m)})}{q(x_t|x_{t-1}^{(m)}, y_t)\hat{p}(y_t|x_{t-1}^{(m)})}$
  end for
  for m in $1 : M$ **do**
    Normalise $w_t^{(m)}$ as $\pi_t^{(m)} := \frac{w_t^{(m)}}{\sum_{i=1}^{M} w_t^{(m)}}$
  end for
  Calculate $\tilde{p}(y_{t|t_r:t-1}) = (\frac{1}{M}\sum_{m=1}^{M} w_t^{(m)})(\sum_{m=1}^{M} w_{t|t-1}^{(m)})$
end for

3.0.3 Particle Markov Chain Monte Carlo

The discussions so far has assumed that θ to be constant. In most ecological applications, these parameters are stochastic. For example, these parameters can be covariate effects we may be interested in making inferences about. Using the Bayesian framework, the joint likelihood of the latent states and these parameters is $p(\theta, x_{1:t}|y_{1:t}) = p_\theta(x_{1:t}|y_{1:t})p(\theta)$; where $p(\theta)$ is the prior distribution for the top-level parameters .

The particle MCMC (Andrieu, Doucet, and Holenstein 2010) makes it possible to jointly sample from the posterior distribution of the states and the top-level parameters θ. This algorithm first proposes a value for the top-level parameter ($\theta^\star$) from a proposal distribution ($\pi(\theta^\star|\theta)$) and fits SMC method described above with the proposed value $\theta^\star$. This proposed value is accepted or rejected based on a Metropolis Hasting acceptance ratio (MHAR) defined as:

$$\begin{aligned}
MHAR &= \frac{p_\theta(y_{1:t})p(\theta^\star)\pi(\theta|\theta^\star)}{p_{\theta^\star}(y_{1:t})p(\theta)\pi(\theta^\star|\theta)};\\
\text{where} \quad p_\theta(y_{1:t}) &= p_\theta(y_1)\prod_{t=2}^{T} p_\theta(y_n|y_{n-1}) \\
&= p_\theta(y_1)\prod_{t=2}^{T}\sum_{i=1}^{M} w^i_{t|\theta}
\end{aligned} \tag{7}$$

where $p_\theta(y_{1:t})$ is the marginal distribution of the observed data given the parameter θ, $w^i_{t|\theta}$ is the i^{th} importance weight at time t given the value of θ, $p(\theta^\star)$ and $p(\theta)$ is the prior distribution of $\theta^\star$ and θ respectively and $\pi(.|.)$ is the proposal distribution for the parameters θ.

Adapting the MHAR defined in equation (7) and weights in equation (6) to our proposed framework, we propose $\theta^\star$ from $\pi(\theta^\star|\hat{\theta}_r)$ and the proposed parameter value ($\theta^\star$) is now accepted or rejected with MHAR:

$$MHAR_{upd} = \frac{p(\theta^\star)\pi(\hat{\theta}_r|\theta^\star)\prod_{t=t_r+1}^{T}\sum_{i=1}^{M} w^i_{t|\theta^\star}}{p(\hat{\theta}_r)\pi(\theta^\star|\hat{\theta}_r)\prod_{t=t_r+1}^{T}\sum_{i=1}^{M} w^i_{t|\hat{\theta}_r}}; \tag{8}$$

where $w^i_{t|\theta}$ is the i^{th} importance weight at time t given the value of θ, $p(\theta^\star)$ and $p(\hat{\theta}_r)$ is the prior distribution of $\theta^\star$ and $\hat{\theta}_r$ respectively, $\pi(.|.)$ is the proposal distribution for the parameters θ and M is the number of parameters. See Supplementary Information 1 for details of the MHAR defined by equations (7) and (8).

We implement this by making changes to the *random walk block sampler* in the `nimbleSMC` package (Michaud et al. 2021) as shown in Algorithm (3).

Algorithm 3 Particle MCMC using the proposed updated model

for i in $1 : n.iterations$ **do**
  Generate $\theta^\star \sim \pi(\theta|\hat{\theta}_r^{(i)})$
  Run a particle filter with proposed SMC in Algoirithm (1) to estimate the marginal likelihood $\tilde{p}(y_{1:T}|\theta^\star)$
  At the last time step T, draw $x^\star_{t_r+1:T} \sim p(x_{t_r+1:T}|y_{t_r+1:T}, \theta^\star)$ from the full particle filter history
  Compute the Metropolis Hasting acceptance ratio in an equation (8) and choose $a^\star = \min(1, \text{MHARupd})$
  Generate $r \sim unif(0, 1)$
  if $a^\star > r$ **then**
    Set $\theta^{(i)} := \theta^\star$; $x^{(i)}_{1:t_r} = x_{\{1:t_r\}|\hat{\theta}_r^{(i)}}$; $x^{(i)}_{t_r+1:T} = x^\star_{t_r+1:T|\theta^\star}$
  else
    Set $\theta^{(i)} := \hat{\theta}_r^{(i)}$; $x^{(i)}_{1:t_r} = x_{\{1:t_r\}|\hat{\theta}_r^{(i)}}$; $x^{(i)}_{t_r+1:T} = x_{\{t_r+1:T\}|\hat{\theta}_r^{(i)}}$
  end if
end for

As discussed in the main paper, we proposed a Gibbs - Metropolis Hasting sampler for the models with more than five model parameters. To do so, we split the set of model parameters θ into two mutually exclusive sets θ_1 and θ_2. We assume that θ_2 depends on θ_1 and latent state distribution x. Therefore, employing the theory behind Gibbs sampling, we define a Metropolis Hastings block sampler to sample values of θ_1 and given the value(s) of θ_1, define a Metropolis Hastings block sampler to sample values of θ_2. We implement this by making changes to the *random walk block sampler* in the `nimbleSMC` package (Michaud et al. 2021) as shown in Algorithm (4):

Algorithm 4 Particle MCMC using the proposed updated model

for i in $1 : n.iterations$ **do**
  Generate $\theta_1^\star \sim \pi(\theta_1 | \widehat{\theta}_{1|r}^{(i)})$
  Compute the Metropolis Hastings acceptance ratio in equation (7) and choose $a^\star = \min(1, \text{MHAR})$
  Generate $r \sim unif(0, 1)$
  if $a^\star > r$ **then**
    Set $\theta_1^{(i)} := \theta_1^\star$
  else
    Set $\theta_1^{(i)} := \widehat{\theta}_{1|r}^{(i)}$
  end if

  Generate $\theta_2^\star \sim \pi(\theta_2 | \widehat{\theta}_{2|r}^{(i)})$
  Set $\theta^\star := (\theta_1^{(i)}, \theta_2^\star)$
  Run a particle filter with proposed SMC in Algoirithm (1) to estimate the marginal likelihood $\tilde{p}(y_{1:T} | \theta^\star)$
  At the last time step T, draw $x_{t_r+1:T}^\star \sim p(x_{t_r+1:T} | y_{t_r+1:T}, \theta^\star)$ from the full particle filter history
  Compute the Metropolis Hasting acceptance ratio in equation (8) and choose $a^\star = \min(1, \text{MHARupd})$
  Generate $r \sim unif(0, 1)$
  if $a^\star > r$ **then**
    Set $\theta^{(i)} := \theta^\star$; $x_{1:t_r}^{(i)} = x_{\{1:t_r\}|\hat{\theta}_r^{(i)}}$; $x_{t_r+1:T}^{(i)} = x_{t_r+1:T|\theta^\star}^\star$
  else
    Set $\theta^{(i)} := \widehat{\theta}_r^{(i)}$; $x_{1:t_r}^{(i)} = x_{\{1:t_r\}|\hat{\theta}_r^{(i)}}$; $x_{t_r+1:T}^{(i)} = x_{\{t_r+1:T\}|\hat{\theta}_r^{(i)}}$
  end if
end for

4 References

Andrieu, Christophe, Arnaud Doucet, and Roman Holenstein. 2010. "Particle Markov Chain Monte Carlo Methods." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72 (3): 269–342.

Arulampalam, M Sanjeev, Simon Maskell, Neil Gordon, and Tim Clapp. 2002. "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian

Tracking." *IEEE Transactions on Signal Processing* 50 (2): 174–88.

Doucet, Arnaud, Nando De Freitas, and Neil Gordon. 2001. "An Introduction to Sequential Monte Carlo Methods." *Sequential Monte Carlo Methods in Practice*, 3–14.

Michaud, Nicholas, Perry de Valpine, Daniel Turek, Christopher J Paciorek, and Dao Nguyen. 2021. "Sequential Monte Carlo Methods in the Nimble and nimbleSMC r Packages." *Journal of Statistical Software* 100: 1–39.

Pitt, Michael K, and Neil Shephard. 1999. "Filtering via Simulation: Auxiliary Particle Filters." *Journal of the American Statistical Association* 94 (446): 590–99.