

A Lagged Particle Filter for Stable Filtering of certain High-Dimensional State-Space Models

BY HAMZA RUZAYQAT¹, AIMAD ER-RAIY¹, ALEXANDROS BESKOS², DAN CRISAN³,
AJAY JASRA¹, & NIKOLAS KANTAS³

¹Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, 23955-6900, KSA. E-Mail: hamza.ruzayqat@kaust.edu.sa,

aimad.erraiy@kaust.edu.sa, ajay.jasra@kaust.edu.sa

²Department of Statistical Science, University College London, London, WC1E 6BT, UK. E-Mail:

a.beskos@ucl.ac.uk

³Department of Mathematics, Imperial College London, London, SW7 2AZ, UK. E-Mail: d.crisan@ic.ac.uk,

n.kantas@ic.ac.uk

Abstract

We consider the problem of high-dimensional filtering of state-space models (SSMs) at discrete times. This problem is particularly challenging as analytical solutions are typically not available and many numerical approximation methods can have a cost that scales exponentially with the dimension of the hidden state. Inspired by lag-approximation methods for the smoothing problem [18, 23], we introduce a lagged approximation of the smoothing distribution that is necessarily biased. For certain classes of SSMs, particularly those that forget the initial condition exponentially fast in time, the bias of our approximation is shown to be uniformly controlled in the dimension and exponentially small in time. We develop a sequential Monte Carlo (SMC) method to recursively estimate expectations with respect to our biased filtering distributions. Moreover, we prove for a class of non-i.i.d. SSMs that as the dimension $d \rightarrow \infty$ the cost to achieve a stable mean square error in estimation, for classes of expectations, is of $\mathcal{O}(Nd^2)$ per-unit time, where N is the number of simulated samples in the SMC algorithm. Our methodology is implemented on several challenging high-dimensional examples including the conservative shallow-water model.

Keywords: Filtering, Sequential Monte Carlo, Lag Approximations, High-Dimensional Particle Filter.

Corresponding author: Hamza Ruzayqat. E-mail: hamza.ruzayqat@kaust.edu.sa

1 Introduction

We are given two sequences of random variables $\{Y_n\}_{n \in \mathbb{N}}$, $\{X_n\}_{n \in \mathbb{N}_0}$, so that $Y_n \in \mathbf{Y} \subseteq \mathbb{R}^{d_y}$, $n \in \mathbb{N}$, and $X_n \in \mathbf{X} \subseteq \mathbb{R}^d$, $n \in \mathbb{N}_0$. We endow \mathbf{X} with an associated σ -field \mathcal{X} . Consider the state-space model (e.g. [5]), where for $n \in \mathbb{N}$, $(A, B) \subseteq \mathbf{X} \times \mathbf{Y}$:

$$\mathbb{P}(X_n \in A | \{y_p, x_p\}_{p \neq n}, y_n) = \int_A f(x_{n-1}, x_n) dx_n, \quad \mathbb{P}(Y_n \in B | \{y_p, x_p\}_{p \neq n}, x_n) = \int_B g(x_n, y_n) dy_n.$$

The initial condition $X_0 = x_0 \in \mathbf{X}$ is assumed given and fixed, $g(x, \cdot)$ is a positive probability density w.r.t. the σ -finite measure dy for each $x \in \mathbf{X}$ and $f(x, \cdot)$ is a positive probability density w.r.t. the σ -finite measure dx for each $x \in \mathbf{X}$. For $n \in \mathbb{N}$ we define the smoothing density:

$$\pi_n(x_{1:n}) := \frac{\prod_{p=1}^n f(x_{p-1}, x_p) g(x_p, y_p)}{\int_{\mathbf{X}^n} \prod_{p=1}^n f(x_{p-1}, x_p) g(x_p, y_p) dx_{1:n}}, \quad (1)$$

where $y_{1:n} := (y_1, \dots, y_n)$ are fixed and known observations. Our objective is to recursively estimate the so-called filter for each $n \in \mathbb{N}$:

$$\pi_{(n)}(\varphi) := \int_{\mathcal{X}^n} \varphi(x_n) \pi_n(x_{1:n}) dx_{1:n},$$

where $\varphi : \mathcal{X} \rightarrow \mathbb{R}$ is assumed π_n -integrable. This is the filtering problem and has several applications in statistics, applied mathematics and engineering; see for instance [5].

The filtering problem is notoriously challenging for a variety of reasons. The main one is that, with the exception of a small class of models, one cannot compute the filter analytically. As a result, there is by now a vast literature on the numerical approximation of the filter; see for instance [5, 8, 9]. The class of algorithm that we focus upon in this article are based on sequential Monte Carlo (SMC) methods. These techniques generate a collection of $N \geq 1$ samples in parallel and combine importance sampling and resampling to numerically approximate expectations w.r.t. the filter. From a mathematical perspective, they are rather well-understood [5, 9], with many convergence results as N grows.

The focus of this article is to study SMC when the dimension of the hidden state, d , is large, for example of the order 10^3 or larger. This problem is often even more problematic than the ordinary filtering problem, due to the high costs of numerical implementation. As noted by several authors [6, 24, 26], the importance sampling method that SMC is based upon can be hugely expensive. The main issue is that for high-dimensional problems of practical interest the proposal and target measure eventually become mutually singular as the dimension increases. As a result, to counteract the weight degeneracy of importance sampling, one may need $N = \mathcal{O}(\kappa^d)$, for some $\kappa > 1$ to obtain reasonable estimators. This exponential cost in the dimension can be prohibitive in practice and *standard* SMC algorithms that consist only of simple importance sampling and resampling recursions are not suitable for high-dimensional problems. There do exist more advanced SMC methods which have been successful [3, 7, 16, 21, 24], but they are often custom designs, taking advantage of useful characteristics of specific problems. This article focusses upon further enhancing these type of methods and focuses on a specific class of State Space Models (SSMs), as opposed to being a universal solution to the high-dimensional filtering problem.

The starting point of this work is the series of the dimension stability results for SMC samplers [10] that were obtained in [1, 2]. These results deal with static¹, i.i.d. target sampling problems and show it is possible to obtain algorithms that scale polynomially in the dimension of the problem and enjoy some type of stability with dimension. The results of this type have been utilized in the approaches of [7, 21], that used a tempering approach to stabilise the weights between two successive observations and inserted a SMC sampler in-between data updates of an ordinary SMC filtering algorithm (see also [14]). The observation in our article is that whilst empirically such approaches may work well when estimating $\pi_{(n)}$, they do not provide a fully satisfying solution, because these methods can exhibit the well-known path degeneracy issue for SMC methods. This is caused by the successive resampling steps and can be briefly described as a lack in diversity in the particles approximating $\pi_n(x_{1:n})$ for times $p < n - CN \log N$ with $C > 0$ being some model specific constant; see [15, 17] for details. Using tempering adds additional resampling steps and as a result tempering-based methods suitable for high dimensions will suffer from path degeneracy. When sticking to estimation of $\pi_{(n)}$, path degeneracy is not usually an issue, but in our high-dimensional setup it prevents extending any type of stability result of the type in [1, 2], with such an extension making up in fact one of the important contributions of the current paper. One potential remedy

¹Note that in contrast the filtering problem is inherently dynamic.

would be to design updates of the entire path, but this is not practical for online algorithms (with fixed computational cost per time).

The contribution of this work is, inspired by the lag-approximation methods for the smoothing problem [11, 18, 20, 23]. We consider a lagged approximation of the smoothing distribution, $\pi_n(x_{1:n})$, that is necessarily biased. This approximation induces an independence between the last $L + 1$ time points of the hidden states of the SSM and the remaining earlier times. The basic premise from there is that this new, but biased target, can be numerically approximated with an algorithm that has a cost that scales polynomially with the dimension. These approximations will induce a bias. However, for SSMs that forget their initial condition exponentially fast in time, this bias will be uniformly controlled in the dimension and exponentially small in time L . In particular, the contributions of this work are:

- We propose an appropriate sequence of biased approximations of the smoothers and an SMC algorithm to numerically approximate the related expectations.
- A proof that for SSMs that can contain dependencies amongst coordinates in the transition density and likelihood, as the dimension $d \rightarrow \infty$, the cost to achieve a stable mean square error in estimation for certain classes of expectations is of $\mathcal{O}(Nd^2)$ per-unit time.
- Numerical implementation of the SMC algorithm on several challenging and high dimensional examples.

The remaining article is structured as follows. In [Section 2](#) we present our methodology. In [Section 3](#) we present our assumptions and theoretical results. In [Section 4](#) the performance of our method is demonstrated on several challenging examples. In the Appendix one can find technical results used in the proofs and a detailed description of the algorithm used in [Section 4](#).

Notation

Let $(\mathbf{X}, \mathcal{X})$ be a measurable space. $\mathcal{P}(\mathbf{X})$ is the collection of probability measures on $(\mathbf{X}, \mathcal{X})$. For $(\mu, \nu) \in \mathcal{P}(\mathbf{X}) \times \mathcal{P}(\mathbf{X})$ we write their total variation distance as $\|\mu - \nu\|_{\text{TV}}$. We use the standard operations $(\mu K)(dz) := \int_{\mathbf{X}} \mu(dx) K(x, dz)$, $\mu(\varphi) := \int_{\mathbf{X}} \varphi(x) \mu(dx)$, for $\mu \in \mathcal{P}(\mathbf{X})$, transition kernel $K : \mathbf{X} \times \mathcal{X} \rightarrow [0, 1]$ and μ -integrable $\varphi : \mathbf{X} \rightarrow \mathbb{R}$. $\mathcal{C}(\mathbf{X})$ is the set of real-valued, continuous, measurable functions on \mathbf{X} . For bounded $\varphi : \mathbf{X} \rightarrow \mathbb{R}$, we set $\|\varphi\|_{\infty} = \sup_{x \in \mathbf{X}} |\varphi(x)|$. For a matrix or vector A , we denote by A^T the transpose of A .

2 A Lagged Particle Filter

2.1 The Sequence of Approximate Target Distributions

Let $L \in \mathbb{N}$ be given and fixed throughout. We assume that one can construct a stochastic method that can approximate reasonably well expectations w.r.t. arbitrary probabilities on \mathbf{X}^{L+1} , but the method does not work well for higher-dimensional problems. By working well we mean that errors do not grow exponentially or worse as the dimension of the hidden state, d , increases. We will give more precise details and assumptions related to the SSM and SMC components later in [Section 3](#).

Given this idea, we introduce the sequence of targets approximating the correct smoothing distribution $\pi_n(x_{1:n})$ in (1) defined as:

$$\begin{aligned} \widehat{\pi}_n(x_{1:n}) &\propto \alpha_{n-L}(x_{1:n-L}) \times (\mu_{n-L}(x_{n-L+1})g(x_{n-L+1}, y_{n-L+1})) \\ &\times \left(\prod_{p=n-L+2}^n f(x_{p-1}, x_p)g(x_p, y_p) \right), \quad n \in \{L+1, L+2, \dots\}. \end{aligned} \quad (2)$$

Here, $\{\mu_p\}_{p \geq 1}$ is a sequence of user-specified probability densities on \mathbf{X} , giving rise to a corresponding sequence of functions $\{\alpha_p\}_{p \geq 1}$, with the latter started by setting $\alpha_1(x_1) = g(x_1, y_1)f(x_0, x_1)$, and then obeying the recursion:

$$\alpha_{n-L+1}(x_{1:n-L+1}) = \alpha_{n-L}(x_{1:n-L})\mu_{n-L}(x_{n-L+1})g(x_{n-L+1}, y_{n-L+1}), \quad n \in \{L+1, L+2, \dots\}.$$

The construction of the sequence of laws $\{\widehat{\pi}_n\}$ is guided by the following principles. Given index $n \in \{L+1, L+2, \dots\}$, we are interested in the marginal $\widehat{\pi}_{(n)}(x_n)$ being a good approximation of the correct filtering law $\pi_{(n)}(x_n)$. More generally, we focus on the properties of the L -dimensional marginal density denoted $\widehat{\pi}_{(n-L+1:n)}(x_{n-L+1:n})$. Notice that by construction $\widehat{\pi}_n(x_{1:n})$ introduces independence amongst the random variables $X_{1:n-L}$ and $X_{n-L+1:n}$. We aim to produce good approximations of the L -dimensional marginal, i.e. of:

$$\begin{aligned} \widehat{\pi}_{(n-L+1:n)}(x_{n-L+1:n}) &\propto (\mu_{n-L}(x_{n-L+1})g(x_{n-L+1}, y_{n-L+1})) \\ &\times \left(\prod_{p=n-L+2}^n f(x_{p-1}, x_p)g(x_p, y_p) \right). \end{aligned}$$

Note that if $\mu_{n-L}(x_{n-L+1})$ were to be the so-called law $\pi_{(n-L+1|n-L)}(x_{n-L+1})$, that is:

$$\mu_{n-L}(x_{n-L+1}) = \int_{\mathbf{X}} \pi_{(n-L)}(x_{n-L})f(x_{n-L}, x_{n-L+1})dx_{n-L} =: \pi_{(n-L+1|n-L)}(x_{n-L+1}), \quad (3)$$

then $\widehat{\pi}_{(n-L+1:n)}(x_{n-L+1:n}) = \pi_{(n-L+1:n)}(x_{n-L+1:n})$, thus one would be able to perform exact filtering. However, in our context (3) is not of practical use, as even when $n = L+1$, a standard approach for approximating $\int_{\mathbf{X}} \pi_{(n-L)}(x_{n-L})f(x_{n-L}, x_{n-L+1})dx_{n-L}$ would typically have a quadratic cost in the number of particles (see [19] for a more efficient implementation) and, most importantly, can induce errors growing exponentially in d similar to ones seen in [26]. By choosing a different $\mu_{n-L}(x_{n-L+1})$ we aim for an SMC method that approximates uniformly well in the dimension d expectations of functionals $\varphi(x_{n-L+1:n})$ at the expense of introducing a bias. The choice of $\{\alpha_p\}_{p \geq 1}$ is motivated directly from the objective of a standard sequential importance sampling scheme when moving from $\widehat{\pi}_{n-1}$ to $\widehat{\pi}_n$, where the incremental importance weights will only depend on $x_{n-L:n}$, thereby forgetting earlier positions of $x_{1:n}$.

Note that under the choice $\mu_{n-L}(x_{n-L+1}) = f(x_{n-L}, x_{n-L+1})$, the independence described above will typically not hold any more as one reverts back to the original smoothing problem, for which potential Monte-Carlo approximations will degenerate (in general) as the dimension increases.

2.2 Algorithm

We now describe the algorithm that we implement and analyse subsequently. We begin with a generic SMC sampler in Algorithm 1. This algorithm, developed in [10], allows for the approximation of a sequence of densities of interest, all defined on a *common* space.

Algorithm 1 : Sequential Monte Carlo Sampler

1. Input:

- Target density $\kappa(u)$ on state-space \mathbf{U} , dominating σ -finite measure du .
- Initial density $\nu(u)$ on state-space \mathbf{U} , dominating σ -finite measure du .
- Annealing parameters $0 = \phi_1 < \phi_2 < \dots < \phi_p = 1$.
- Sequence of Markov kernels K_2, \dots, K_p , such that K_n has invariant distribution proportional to $\kappa(u)^{\phi_n} \nu(u)^{1-\phi_n} du$, $n \in \{2, \dots, p\}$.
- Number of samples $N \in \mathbb{N}$ and resampling threshold $N^* \in [1, N]$.
- Samples $U^{(1)}, \dots, U^{(N)}$ that approximate ν .

2. Initialize: Set time $k = 1$ and $W_k^{(i)} = 1$, $u_1^{(i)} = U^{(i)}$ for $i \in \{1, \dots, N\}$.

3. Iterate: For $i \in \{1, \dots, N\}$ set:

$$W_{k+1}^{(i)} = W_k^{(i)} \left(\frac{\kappa(u_k^{(i)})}{\nu(u_k^{(i)})} \right)^{\phi_{k+1} - \phi_k}.$$

Compute the Effective Sample Size (ESS):

$$ESS_{k+1} = \frac{\left(\sum_{i=1}^N W_{k+1}^{(i)} \right)^2}{\sum_{i=1}^N (W_{k+1}^{(i)})^2}.$$

If $ESS_{k+1} < N^*$ resample the particles and set $W_{k+1}^{(i)} = 1$ for $i \in \{1, \dots, N\}$, else do nothing. Then, sample $U_{k+1}^{(i)} | u_k^{(i)}$ using $K_{k+1}(u_k^{(i)}, \cdot)$ and set $u_{k+1}^{(i)} = U_{k+1}^{(i)}$, $i \in \{1, \dots, N\}$. Increment $k = k + 1$, if $k = p + 1$ go to the output Step 4. otherwise go to the start of iterate Step 3..

4. Output: Samples U_p^1, \dots, U_p^N and corresponding (unnormalised) weights W_p^1, \dots, W_p^N .

Our proposed lagged particle filter is detailed in [Algorithm 2](#). To shorten the description, when $n \in \{1, \dots, L\}$ we set $\widehat{\pi}_n(x_{1:n}) = \pi_n(x_{1:n})$. There are several important remarks at this point. In terms of the weight expressions $W_{k+1}^{(i)}$ in [Algorithm 1](#), when they appear in [Algorithm 2](#) (at iteration $n > L$), one would have:

$$\begin{aligned} W_{k+1}^{(i)} &= W_k^{(i)} \left(\frac{\widehat{\pi}_n(x_{1:n}^{(i)})}{\widehat{\pi}_{n-1}(x_{1:n-1}^{(i)}) f(x_{n-1}^{(i)}, x_n^{(i)})} \right)^{\phi_{k+1} - \phi_k} \\ &= W_k^{(i)} \left(\frac{\mu_{n-L}(x_{n-L+1}^{(i)}) g(x_n^{(i)}, y_n)}{f(x_{n-L}^{(i)}, x_{n-L+1}^{(i)})} \right)^{\phi_{k+1} - \phi_k}, \quad \phi_{k+1} - \phi_k = \frac{1}{d}. \end{aligned} \quad (4)$$

Notice, importantly, that this expression does not depend upon $x_{1:n-L-1}^{(i)}$. The choice of spacing the temperatures at order $\mathcal{O}(d^{-1})$ apart coincides with the choice in [\[1\]](#) and will prove critical in

the sequel. A key ingredient of [Algorithm 2](#) is the specification and effectiveness of the Markov kernels $K_{2,n}, \dots, K_{d+1,n}$. By construction the variables $X_{n-L+1:n}$ and $X_{1:n-L}$ (and indeed $X_{n-L:n}$ and $X_{1:n-L-1}$) are independent under the approximate target $\hat{\pi}_n(x_{1:n})$. Since the weight function (4) depends on $X_{n-L:n}$ only, this means one needs to only update $X_{n-L:n}$ in the Markov kernel (i.e. the last $(L+1)$ hidden states across time). This is typically implemented via a suitable Markov Chain Monte Carlo (MCMC) step such as Metropolis-Hastings or Gibbs sampling. Moreover, as the Markov kernel can be constructed so that it *does not depend* upon $X_{1:n-L-1}$, the computational cost per iteration of [Algorithm 2](#) is fixed w.r.t. the time index n .

Algorithm 2 : A Lagged Particle Filter Algorithm for High-Dimensional Filtering

1. Initialization: Sample $X_1^{(i)}$ i.i.d. from $f(x_0, x_1)$. Run the SMC sampler in [Algorithm 1](#) with:

- Target density $\pi_1(x_1)$.
- Initial density $f(x_0, x_1)$.
- Annealing parameters $\phi_k = (k-1)/d$, $p = d+1$, $k \in \{1, 2, \dots, d+1\}$.
- Sequence of Markov kernels $K_{2,1}, \dots, K_{d+1,1}$, such that $K_{k,1}$ has invariant distribution proportional to:

$$\pi_1(x_1)^{\phi_k} \{f(x_0, x_1)\}^{1-\phi_k} dx_1, \quad k \in \{2, \dots, d+1\}. \quad (5)$$

- $N \in \mathbb{N}$, and resampling threshold $N^* = 1$ (i.e. no resampling).
- Initial samples $X_1^{(i)}$, $i \in \{1, \dots, N\}$.

Let $X_1^{(i)}(1)$, $W^{(i)}(1)$, $i \in \{1, \dots, N\}$, be the samples and weights in the output of [Algorithm 1](#). Set $n = 2$ and go to Step 2. below.

2. Iteration: Resample $X_{1:n-1}^{(i)}(n-1)$ with weights $W^{(i)}(n-1)$, and denote the obtained particles $\check{X}_{1:n-1}^{(i)}(n-1)$, $i \in \{1, \dots, N\}$. Sample new co-ordinate $X_n^{(i)} | \check{X}_{n-1}^{(i)} = \check{x}_{n-1}^{(i)}$ using the transition $f(\check{x}_{n-1}^{(i)}, \cdot)$, $i \in \{1, \dots, N\}$. Run the SMC sampler in [Algorithm 1](#) with:

- Target density $\hat{\pi}_n(x_{1:n})$.
- Initial density $\hat{\pi}_{n-1}(x_{1:n-1})f(x_{n-1}, x_n)$.
- Annealing parameters $\phi_k = (k-1)/d$, $p = d+1$, $k \in \{1, 2, \dots, d+1\}$.
- Sequence of Markov kernels $K_{2,n}, \dots, K_{d+1,n}$, such that $K_{k,n}$ has invariant distribution proportional to:

$$\hat{\pi}_n(x_{1:n})^{\phi_k} \{\hat{\pi}_{n-1}(x_{1:n-1})f(x_{n-1}, x_n)\}^{1-\phi_k} dx_{1:n}, \quad k \in \{2, \dots, d+1\}. \quad (6)$$

- $N \in \mathbb{N}$, $N^* = 1$.
- Initial samples $X_{1:n}^{(i)} := (\check{X}_{1:n-1}^{(i)}(n-1), X_n^{(i)})$, $i \in \{1, \dots, N\}$.

Let $X_{1:n}^{(i)}(n)$, $W^{(i)}(n)$, $i \in \{1, \dots, N\}$, be the samples, weights in the output of [Algorithm 1](#). Set $n = n+1$ and return to the start of the iterate at Step 2. above.

In the next section we will prove that, under certain conditions, [Algorithm 2](#) can estimate fixed dimensional marginals with a cost that is polynomial in dimension. This includes being able to handle the bias that has been introduced by our approximation. Intrinsic to our proof (that is to a large extent contained in the works of [\[1, 2\]](#)) is the fact that for large d (and under certain conditions) the outcome variable $X_{n-L:n}$ resulting from the application of $\mathcal{O}(d)$ tempering steps, will be approximately sampled from the marginal law $\hat{\pi}_{(n-L:n)}$. If one does not use the approximation $\{\hat{\pi}_n\}$ and instead works directly with $\{\pi_n\}$ then the only way to achieve such an ergodicity would be to *update the entire path of samples from time 1 to n* . This will lead to an algorithm that is not recursive or online.

3 Theoretical Results

3.1 Non I.I.D. Targets

To avoid technicalities we henceforth assume that $\mathsf{X} = E^d$ for a compact set $E \subset \mathbb{R}$. We follow closely the approach in [\[1, 2\]](#), developed therein for i.i.d. target distributions. One can generalize the aforementioned analysis, to the case of models with the following structure.

Assumption 3.1. *Throughout, $m \in \mathbb{N}_0$ is fixed and d is such that $d \geq m$.*

i) There exist functions $\bar{g} : E \times \mathsf{Y} \rightarrow \mathbb{R}$ and $\tilde{g} : E^m \times \mathsf{Y} \rightarrow \mathbb{R}$, such that for any $(x, y) \in E^d \times \mathsf{Y}$:

$$g(x, y) = \exp \left\{ \tilde{g}(x^{1:m}, y) + \sum_{j=1}^d \bar{g}(x^j, y) \right\}.$$

ii) There exist functions $\bar{f} : E \times E \rightarrow \mathbb{R}$ and $\tilde{f} : E^m \times E^m \rightarrow \mathbb{R}$, such that for any $(x, z) \in E^{2d}$:

$$f(x, z) = \exp \left\{ \tilde{f}(x^{1:m}, z^{1:m}) + \sum_{j=1}^d \bar{f}(x^j, z^j) \right\}.$$

iii) For each $k \in \mathbb{N}$, there exist functions $\bar{\mu}_k : E \rightarrow \mathbb{R}$, $\tilde{\mu}_k : E^m \rightarrow \mathbb{R}$, such that for any $x \in E^d$:

$$\mu_k(x) = \exp \left\{ \tilde{\mu}_k(x^{1:m}) + \sum_{j=1}^d \bar{\mu}_k(x^j) \right\}.$$

Note we are using here superscripts to index coordinates of a d -dimensional vector x and similar to before $x^{p:m} := (x^p, \dots, x^m)$, for $m > p$. Given the model structure in [Assumption 3.1](#), the sequence of targets defined via [\(5\)](#)-[\(6\)](#), permits for a factorization in terms of co-ordinates $x^{1:m}, x^{m+1}, \dots, x^d$. We will re-express the sequence of targets in a way that exploits this.

To shorten expressions, we will write for $(x^{1:m}, y, x, z^{1:m}, z, k) \in E^m \times Y \times E \times E^m \times E \times \mathbb{N}$:

$$\begin{aligned}
g^{(m)}(x^{1:m}, y) &:= \exp \left\{ \tilde{g}(x^{1:m}, y) + \sum_{j=1}^m \bar{g}(x^j, y) \right\}; \\
g^e(x, y) &:= \exp \left\{ \bar{g}(x, y) \right\}; \\
f^{(m)}(x^{1:m}, z^{1:m}) &:= \exp \left\{ \tilde{f}(x^{1:m}, z^{1:m}) + \sum_{j=1}^m \bar{f}(x^j, z^j) \right\}; \\
f^e(x, z) &:= \exp \left\{ \bar{f}(x, z) \right\}; \\
\mu_k^{(m)}(x^{1:m}) &:= \exp \left\{ \tilde{\mu}_k(x^{1:m}) + \sum_{j=1}^m \bar{\mu}_k(x^j) \right\}; \\
\mu_k^e(x) &:= \exp \left\{ \bar{\mu}_k(x) \right\}.
\end{aligned}$$

Remark 3.1. *To avoid repetitive expressions when separating the initial phase $n \in \{1, 2, \dots, L\}$ from subsequent steps $n \in \{L+1, L+2, \dots\}$, we adopt the convention that distributions when $n < 0$ are to be treated as Dirac measures on the fixed position x_0 at time $n = 0$ (or the subset of co-ordinates of x_0 implied by the context).*

We consider the sequence of tempered approximate target distributions defined via (5) for $n = 1$ and via (6) for general $n > 1$, given the model factorisation introduced by Assumption 3.1. After simple calculations, we define for $(k, n) \in \{2, \dots, d+1\} \times \mathbb{N}$ and for $j \in \{m+1, \dots, d\}$ the distributions $(\tilde{\pi}_{k,n}^{(m)}, \tilde{\pi}_{k,n}) \in \mathcal{P}(E^{m(L+1)}) \times \mathcal{P}(E^{L+1})$ as:

$$\begin{aligned}
\tilde{\pi}_{k,n}^{(m)}(dx_{n-L:n}^{1:m}) &\propto \left(\frac{\mu_{n-L}^{(m)}(x_{n-L+1}^{1:m}) g^{(m)}(x_n^{1:m}, y_n)}{f^{(m)}(x_{n-L}^{1:m}, x_{n-L+1}^{1:m})} \right)^{\phi_k} \mu_{n-L-1}^{(m)}(x_{n-L}^{1:m}) g^{(m)}(x_{n-L}^{1:m}, y_{n-L}) \\
&\quad \times \left(\prod_{p=n-L+1}^{n-1} f^{(m)}(x_{p-1}^{1:m}, x_p^{1:m}) g^{(m)}(x_p^{1:m}, y_p) \right) f^{(m)}(x_{n-1}^{1:m}, x_n^{1:m}), \quad (7)
\end{aligned}$$

$$\begin{aligned}
\tilde{\pi}_{k,n}(dx_{n-L:n}^j) &\propto \left(\frac{\mu_{n-L}^e(x_{n-L+1}^j) g^e(x_n^j, y_n)}{f^e(x_{n-L}^j, x_{n-L+1}^j)} \right)^{\phi_k} \mu_{n-L-1}^e(x_{n-L}^j) g^e(x_{n-L}^j, y_{n-L}) \\
&\quad \times \left(\prod_{p=n-L+1}^{n-1} f^e(x_{p-1}^j, x_p^j) g^e(x_p^j, y_p) \right) f^e(x_{n-1}^j, x_n^j). \quad (8)
\end{aligned}$$

Given the model structure imposed by Assumption 3.1, the mutation kernels used within the algorithm will be adjusted accordingly. That is, for $(k, n) \in \{2, \dots, d+1\} \times \mathbb{N}$, it is reasonable to select $K_{k,n} : E^{dn} \times \mathcal{B}(E^{dn}) \rightarrow [0, 1]$ defined as:

$$K_{k,n}(x_{1:n}, dz_{1:n}) = \delta_{\{x_{1:n-L-1}\}}(dz_{1:n-L-1}) \times \bar{K}_{k,n}^{(m)}(x_{n-L:n}^{1:m}, dz_{n-L:n}^{1:m}) \prod_{j=m+1}^d \bar{K}_{k,n}(x_{n-L:n}^j, dz_{n-L:n}^j),$$

with $\bar{K}_{k,n}^{(m)} : E^{m(L+1)} \times \mathcal{B}(E^{m(L+1)}) \rightarrow [0, 1]$, $\bar{K}_{k,n} : E^{(L+1)} \times \mathcal{B}(E^{(L+1)}) \rightarrow [0, 1]$ defined so that they preserve $\tilde{\pi}_{k,n}^{(m)}$ and $\tilde{\pi}_{k,n}$ respectively, i.e.:

$$\tilde{\pi}_{k,n}^{(m)} \bar{K}_{k,n}^{(m)} = \tilde{\pi}_{k,n}^{(m)}, \quad \tilde{\pi}_{k,n} \bar{K}_{k,n} = \tilde{\pi}_{k,n}.$$

We denote by $x_{k,p}^{(i),j}(n) \in E$ the i^{th} particle ($i \in \{1, \dots, N\}$), in the j^{th} dimension ($j \in \{1, \dots, d\}$), $k \in \{1, \dots, d\}$ is the SMC sampler time, n is the state-space model time and p an instance across times $\{1, \dots, n\}$. Under the model structure determined in [Assumption 3.1](#), upon recalling the expression for the weights in (4), one has for $n \in \mathbb{N}$:

$$W^{(i)}(n) = \exp \left\{ \frac{1}{d} \sum_{k=1}^d \left(\tilde{g}(x_{k,n}^{(i),1:m}(n), y_n) + \tilde{\mu}_{n-L}(x_{k,n-L+1}^{(i),1:m}(n)) - \tilde{f}(x_{k,n-L}^{(i),1:m}(n), x_{k,n-L+1}^{(i),1:m}(n)) \right) \right. \\ \left. + \frac{1}{d} \sum_{j=1}^d \sum_{k=1}^d \left(\bar{g}(x_{k,n}^{(i),j}(n), y_n) + \bar{\mu}_{n-L}(x_{k,n-L+1}^{(i),j}(n)) - \bar{f}(x_{k,n-L}^{(i),j}(n), x_{k,n-L+1}^{(i),j}(n)) \right) \right\}.$$

In agreement with the convention introduced in [Remark 3.1](#), mappings that involve negative time subscripts should be treated as zeros in the above expression for $W^i(n)$.

3.2 Error Analysis for Lagged Particle Filter

To give our theoretical results we introduce some notation following from [1]. For simplicity, we assume that x_0^j is the same for each j .

We extend the definition, in (7)-(8), of $\tilde{\pi}_{k,n}^{(m)}$ and $\tilde{\pi}_{k,n}$, $n \in \mathbb{N}$, to allow for a *continuous* tempering parameter $s \in [0, 1]$:

$$\tilde{\pi}_{s,n}^{(m)}(dx_{n-L:n}^{1:m}) \propto \left(\frac{\mu_{n-L}^{(m)}(x_{n-L+1}^{1:m})g^{(m)}(x_n^{1:m}, y_n)}{f^{(m)}(x_{n-L}^{1:m}, x_{n-L+1}^{1:m})} \right)^s \mu_{n-L-1}^{(m)}(x_{n-L}^{1:m})g^{(m)}(x_{n-L}^{1:m}, y_{n-L}) \\ \times \left(\prod_{p=n-L+1}^{n-1} f^{(m)}(x_{p-1}^{1:m}, x_p^{1:m})g^{(m)}(x_p^{1:m}, y_p) \right) f^{(m)}(x_{n-1}^{1:m}, x_n^{1:m}), \\ \tilde{\pi}_{s,n}(dx_{n-L:n}^j) \propto \left(\frac{\mu_{n-L}^e(x_{n-L+1}^j)g^e(x_n^j, y_n)}{f^e(x_{n-L}^j, x_{n-L+1}^j)} \right)^s \mu_{n-L-1}^e(x_{n-L}^j)g^e(x_{n-L}^j, y_{n-L}) \\ \times \left(\prod_{p=n-L+1}^{n-1} f^e(x_{p-1}^j, x_p^j)g^e(x_p^j, y_p) \right) f^e(x_{n-1}^j, x_n^j).$$

Similarly, for $s \in [0, 1]$, $n \in \mathbb{N}$, we consider the *continuum* of Markov kernels $\bar{K}_{s,n}^{(m)}$ and $\bar{K}_{s,n}$ that preserve $\tilde{\pi}_{s,n}^{(m)}$ and $\tilde{\pi}_{s,n}$, respectively.

For $n \in \mathbb{N}$ and $x_{n-L:n} \in E^{L+1}$ set:

$$\psi_n(x_{n-L:n}) := \bar{g}(x_n, y_n) + \bar{\mu}_{n-L}(x_{n-L+1}) - \bar{f}(x_{n-L}, x_{n-L+1}).$$

Standard adjustments apply in the case of negative time index, see [Remark 3.1](#). We denote by $\hat{\psi}_{s,n} : E^{L+1} \rightarrow \mathbb{R}$ the solution to the following Poisson equation, $s \in [0, 1]$, $n \in \mathbb{N}$:

$$\hat{\psi}_{s,n} - \bar{K}_{s,n}(\hat{\psi}_{s,n}) = \psi_n - \tilde{\pi}_{s,n}(\psi_n).$$

We set

$$\sigma^2(n) = \int_0^1 \tilde{\pi}_{s,n}(\hat{\psi}_{s,n}^2 - \bar{K}_{s,n}(\hat{\psi}_{s,n})^2) ds.$$

We use the following assumptions as in [2].

(A1) For any $n \in \{1, 2, \dots\}$ there exist $\theta_n \in (0, 1)$, $(\zeta_n^{(m)}, \zeta_n) \in \mathcal{P}(\mathcal{B}(E^{m(L+1)})) \times \mathcal{P}(\mathcal{B}(E^{L+1}))$, such that for each $s \in [0, 1]$:

$$\begin{aligned}\bar{K}_{s,n}^{(m)}(\bar{x}_{n-L:n}^{1:m}, A) &\geq \theta_n \zeta_n^{(m)}(A), \quad \text{for all } (x_{n-L:n}^{1:m}, A) \in E^{m(L+1)} \times \mathcal{B}(E^{m(L+1)}), \\ \bar{K}_{s,n}(\bar{x}_{n-L:n}, A) &\geq \theta_n \zeta_n(A), \quad \text{for all } (\bar{x}_{1:L+1}, A) \in E^{L+1} \times \mathcal{B}(E^{L+1}).\end{aligned}$$

(A2) For any $n \in \{1, 2, \dots\}$ there exists $C_n < \infty$ such that for any $s, t \in [0, 1]$ we have

$$\begin{aligned}\sup_{x_{n-L:n}^{1:m} \in E^{m(L+1)}} \left\| \bar{K}_{s,n}^{(m)}(x_{n-L:n}^{1:m}, \cdot) - \bar{K}_{t,n}^{(m)}(x_{n-L:n}^{1:m}, \cdot) \right\|_{\text{TV}} &\leq C_n |s - t|, \\ \sup_{x_{n-L:n} \in E^{L+1}} \left\| \bar{K}_{s,n}(x_{n-L:n}, \cdot) - \bar{K}_{t,n}(x_{n-L:n}, \cdot) \right\|_{\text{TV}} &\leq C_n |s - t|.\end{aligned}$$

(A3) There exists a $C < \infty$ such that:

- i) $\max\{\sup_{(x^{1:m}, y) \in E^m \times Y} |\tilde{g}(x^{1:m}, y)|, \sup_{(x,y) \in E \times Y} |\bar{g}(x, y)|\} \leq C$;
- ii) $\max\{\sup_{(x^{1:m}, z^{1:m}) \in E^{2m}} |\tilde{f}(x^{1:m}, z^{1:m})|, \sup_{(x,z) \in E^2} |\bar{f}(x, z)|\} \leq C$;
- iii) $\max\{\sup_{(x^{1:m}, k) \in E^m \times \mathbb{N}} |\tilde{\mu}_k(x^{1:m})|, \sup_{(x,k) \in E \times \mathbb{N}} |\bar{\mu}_k(x)|\} \leq C$.

Remark 3.2. To simplify the notation, for $s \in [0, 1]$, $n \in \mathbb{N}$, and $n - L \leq k \leq n$, we denote by $\tilde{\pi}_{s,n,k}^{(m)}$ and $\tilde{\pi}_{s,n,k}$ the marginal of laws $\tilde{\pi}_{s,n}^{(m)}(dx_{n-L:n}^{1:m})$ and $\tilde{\pi}_{s,n}(dx_{n-L:n})$ on $x_{k:n}^{1:m}$ and $x_{k:n}$, respectively. In addition, $\mathbb{V}_{\tilde{\pi}_{s,n,k}^{(m)}}$, $\mathbb{V}_{\tilde{\pi}_{s,n,k}}$ denote the variances of these above laws.

Now we have the following result, that constitutes an extension of [2, Theorem 3.1]. Here, $\|\cdot\|_2$ denotes the standard L_2 -norm on the space of squared-integrable random variables.

Proposition 3.1. Assume Assumption 3.1 and (A1-3) hold. Then for any fixed n and number of particles N with $(n, N) \in \mathbb{N} \times \mathbb{N}$, we have:

i) For any $\varphi^{(m)} \in \mathcal{C}(E^{m(L+1)})$

$$\lim_{d \rightarrow \infty} \left\| \frac{\sum_{i=1}^N \varphi^{(m)}(\tilde{X}_{n-L:n}^{(i),1:m}(n))}{N} - \tilde{\pi}_{1,n,n-L}^{(m)}(\varphi^{(m)}) \right\|_2^2 \leq \frac{\mathbb{V}_{\tilde{\pi}_{1,n,n-L}^{(m)}}[\varphi^{(m)}]}{N} (1 + C_{n,N}(\sigma^2(n))).$$

ii) For any $\varphi \in \mathcal{C}(E^{L+1})$

$$\lim_{d \rightarrow \infty} \left\| \frac{\sum_{i=1}^N \varphi(\tilde{X}_{n-L:n}^{(i),m+1}(n))}{N} - \tilde{\pi}_{1,n,n-L}(\varphi) \right\|_2^2 \leq \frac{\mathbb{V}_{\tilde{\pi}_{1,n,n-L}}[\varphi]}{N} (1 + C_{n,N}(\sigma^2(n)))$$

In both cases above $C_{n,N}(\sigma^2(n)) = \exp\{\sigma^2(n)\} + C_n \exp\{17\sigma^2(n)\}/N^{1/6}$, where C_n is a constant that does not depend on N or $\sigma^2(n)$.

Proof. The proof is given in Appendix B. □

3.3 Error Analysis for Lagged Distribution Approximation

We add an assumption which will be useful below.

(A4) We have the following lower bounds, for some $0 < C < \infty$:

- i) $\min\{\inf_{(x,y) \in E^m \times Y} \tilde{g}(x^{1:m}, y), \inf_{(x,y) \in E \times Y} \bar{g}(x, y)\} \geq C$.
- ii) $\min\{\inf_{(x^{1:m}, z^{1:m}) \in E} \tilde{f}(x^{1:m}, z^{1:m}), \inf_{(x,z) \in E^2} \bar{f}(x, z)\} \geq C$.

We write $\tilde{\pi}_{(n)}^{(m)}$ (resp. $\tilde{\pi}_{(n-L+1|n-L)}^{(m)}$) and $\tilde{\pi}_{(n)}$ (resp. $\tilde{\pi}_{(n-L+1|n-L)}$) for the *true* filtering distribution (resp. *true* predictive distribution as defined in (3)) in the first m dimensions and in the $(m+1)^{\text{th}}$ -dimension respectively. Then the following result is standard in the literature, see e.g. [9, Proposition 4.3.7].

Proposition 3.2. *Assume (A3-4). Then there exist $(\rho, C) \in (0, 1) \times (0, \infty)$ such that:*

i) For any $(d, L, n) \in \mathbb{N}^3$ and any bounded function $\varphi^{(m)} : E^m \rightarrow \mathbb{R}$:

$$\left| [\tilde{\pi}_{1,n,n}^{(m)} - \tilde{\pi}_{(n)}^{(m)}](\varphi^{(m)}) \right| \leq C \|\varphi^{(m)}\|_{\infty} \|\mu_{n-L}^{(m)} - \tilde{\pi}_{(n-L+1|n-L)}^{(m)}\|_{\text{TV}} \times \rho^L.$$

ii) For any $(d, L, n) \in \mathbb{N}^3$ and any bounded function $\varphi : E \rightarrow \mathbb{R}$:

$$\left| [\tilde{\pi}_{1,n,n} - \tilde{\pi}_{(n)}](\varphi) \right| \leq C \|\varphi\|_{\infty} \|\mu_{n-L}^e - \tilde{\pi}_{(n-L+1|n-L)}\|_{\text{TV}} \times \rho^L.$$

Remark 3.3. *Condition (A3)(i) can be relaxed to the following: for any $y \in Y$, there exists a $C < \infty$ such that $\max\{\sup_{x^{1:m} \in E^m} \tilde{g}(x, y), \sup_{x \in E} \bar{g}(x, y)\} \leq C$. This weaker condition suffices for Proposition 3.1, but not for Proposition 3.2.*

3.4 Total Error of Lagged Approach

As a result of Propositions 3.1–3.2 it easily follows that for $n \in \mathbb{N}_0$

$$\begin{aligned} \lim_{d \rightarrow \infty} \left\| \sum_{i=1}^N \frac{\varphi^{(m)}(\tilde{X}_n^{(i), 1:m}(n))}{N} - \tilde{\pi}_{(n)}^{(m)}(\varphi^{(m)}) \right\|_2^2 &\leq \\ &\frac{\mathbb{V}_{\tilde{\pi}_{1,n,n}^{(m)}}[\varphi^{(m)}]}{N} \left(1 + C_{n,N}(\sigma^2(n))\right) + C \|\varphi^{(m)}\|_{\infty}^2 \|\mu_{n-L}^{(m)} - \tilde{\pi}_{(n-L+1|n-L)}^{(m)}\|_{\text{TV}}^2 \times \rho^L, \end{aligned}$$

and

$$\begin{aligned} \lim_{d \rightarrow \infty} \left\| \frac{\sum_{i=1}^N \varphi(\tilde{X}_n^{(i), m+1}(n))}{N} - \tilde{\pi}_{(n)}(\varphi) \right\|_2^2 &\leq \\ &\frac{\mathbb{V}_{\tilde{\pi}_{1,n,n-L}[\varphi]}}{N} \left(1 + C_{n,N}(\sigma^2(n))\right) + C \|\varphi\|_{\infty}^2 \|\mu_{n-L}^e - \pi_{(n-L+1|n-L)}\|_{\text{TV}}^2 \times \rho^L. \end{aligned}$$

Thus, for given $\epsilon > 0$, to obtain an MSE of $\mathcal{O}(\epsilon^2)$ one needs to choose L large enough, *independently of d* and then a number of particles large enough, *independently of d* . In particular, choosing $L = \mathcal{O}(|\log(\epsilon)|)$ and $N = \mathcal{O}(\epsilon^{-2})$ given an MSE is $\mathcal{O}(\epsilon^2)$. The cost to achieve this MSE is of the order $\mathcal{O}(d^2 \epsilon^{-2} (n + |\log(\epsilon)|^2))$. Due to the structure of the model, we expect that this cost is a best case scenario. Note also, that one cannot in practice increase L indefinitely, as typically the mixing rate of the Markov kernels $\{K_{k,n}\}_{(k,n) \in \{2, \dots, d+1\} \times \{1, 2, \dots\}}$ would fall with L ; also, one cannot afford to run the algorithm with L too large.

4 Numerical Results

4.1 Model

Let $X_n \in \mathbb{R}^d$, for $n = 1, \dots, T$, and $Y_m \in \mathbb{R}^{d_y}$, for $m = 1, \dots, M$, for some $T, M \in \mathbb{N}$. Assume that $X_0 = x_0 \in \mathbb{R}^d$ is given. We consider the following SSM:

$$X_n = q_n(X_{n-1}) + R_1^{1/2}W_n, \quad n = 1, \dots, T, \quad (9)$$

$$Y_m = CX_{m\hat{k}} + R_2^{1/2}V_m, \quad m = 1, \dots, M, \quad (10)$$

where $R_1 \in \mathbb{R}^{d \times d}$, $R_2 \in \mathbb{R}^{d_y \times d_y}$ are symmetric positive definite matrices, and $C \in \mathbb{R}^{d_y \times d}$. The

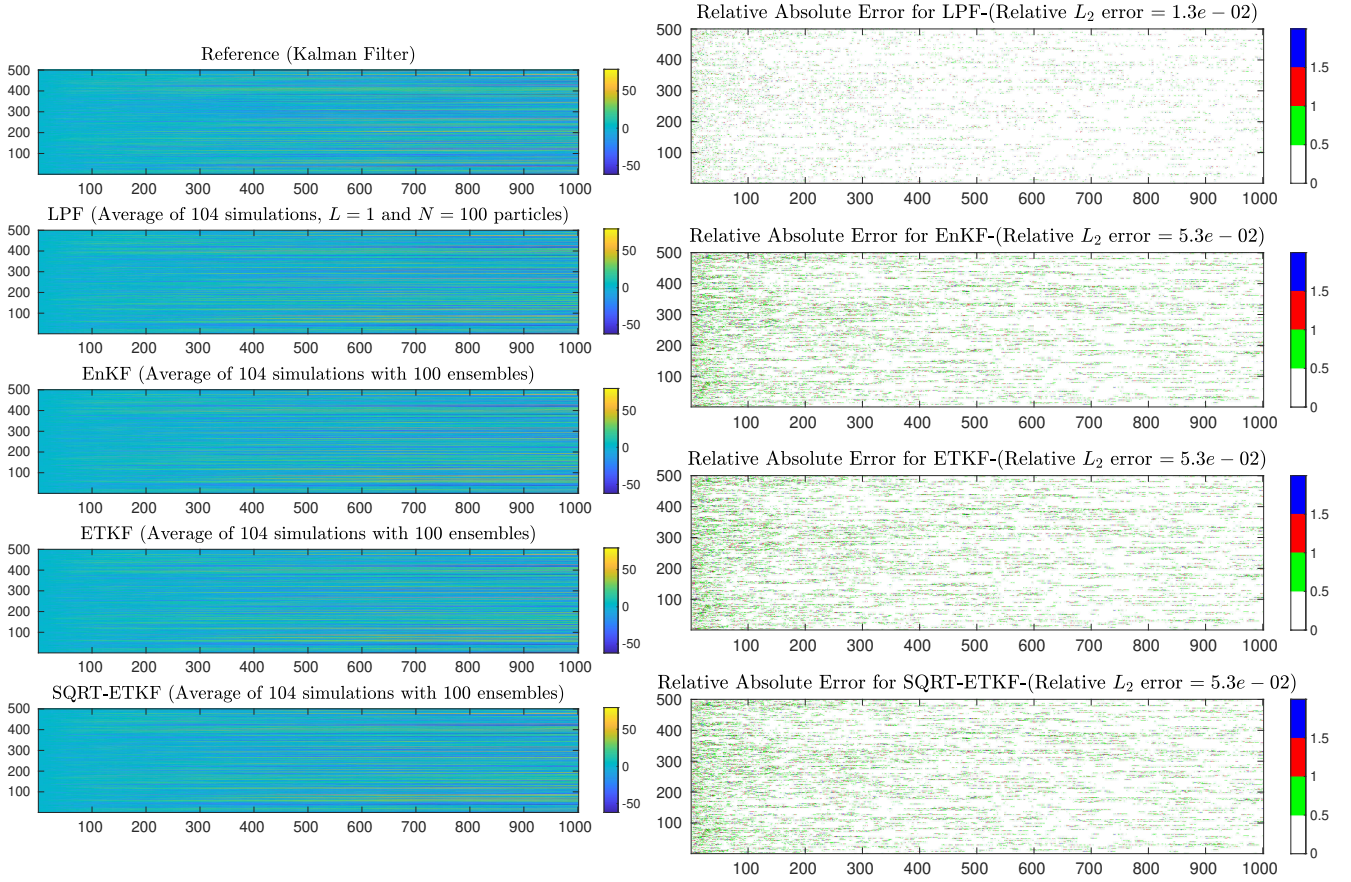


Figure 1: (Linear-Gaussian Model) A comparison amongst the different filters. The horizontal axis represents the time parameter n , and the vertical axis represents the signal's coordinates. The left panel shows approximations of $\pi_n(\varphi)$ with $\varphi(x_{1:n}) = x_n$ as obtained by the KF, LPF, EnKF, ETKF and ETKF-SQRT algorithms. The right panel shows the corresponding relative absolute errors and the corresponding relative L_2 -error is shown in the titles. For readability, the color of the relative absolute errors that are less than 0.5 is set to white.

random variables W_n and V_m are sampled from $\mathcal{N}(0, I_d)$ and $\mathcal{N}(0, I_{d_y})$, respectively, where I_r is the identity matrix of size $r \times r$. The positive integer \hat{k} is the time frequency at which the signal or part of it is observed (e.g. if $\hat{k} = 3$, the signal is observed at times 3, 6, 9, ...) The function $q_n : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is possibly nonlinear.

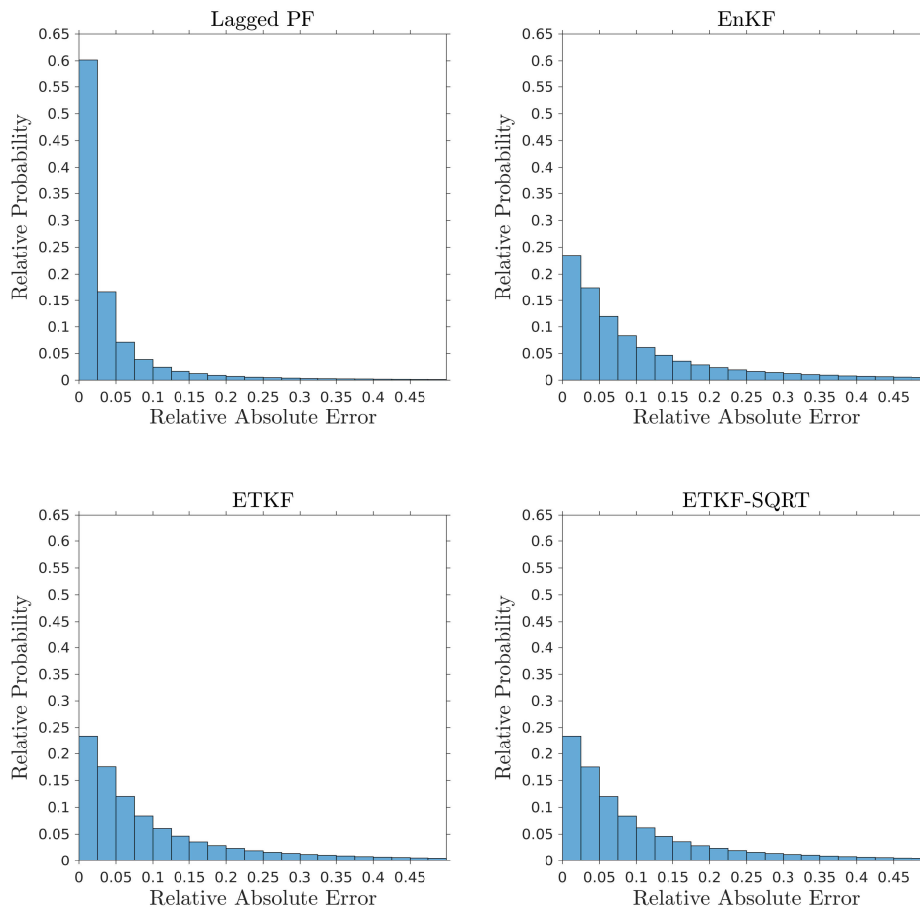


Figure 2: (Linear-Gaussian Model) A histogram of the relative absolute errors for each filter. The relative probability here is defined as the number of elements in the bin divided by the total number of elements, which is $d \times (T + 1)$.

The target smoothing distribution is

$$\pi_n(x_{1:n}) \propto \prod_{i=1}^n f(x_{i-1}, x_i) g(x_i, y_i), \quad (11)$$

where from the above SSM,

$$f(x_{n-1}, x_n) = \frac{1}{\sqrt{(2\pi)^d \det [R_1]}} \exp \left\{ -\frac{1}{2} [x_n - q_n(x_{n-1})]^T R_1^{-1} [x_n - q_n(x_{n-1})] \right\}, \quad (12)$$

$$g(x_{m\hat{k}}, y_m) = \frac{1}{\sqrt{(2\pi)^{d_y} \det [R_2]}} \exp \left\{ -\frac{1}{2} [y_m - Cx_{m\hat{k}}]^T R_2^{-1} [y_m - Cx_{m\hat{k}}] \right\}. \quad (13)$$

We will implement an enhancement of [Algorithm 2](#) that includes adaptive tempering and resampling. Resampling occurs whenever the ESS \mathcal{E} is less than a given threshold N^* . This way the algorithm calculates the annealing parameters ϕ_k on the fly. A detailed description of the algorithm used can be found in [Algorithm 3](#) in the Appendix. There we also specify the Markov kernels $K_{k,n}$ as Gaussian random walk Metropolis updates. The proposal density covariance matrix Σ_m is proportional to $(2.38^2/d)\alpha(\phi_k)I_d$ modified adaptively so that the average acceptance rate over all particles is in the range 0.15 - 0.25. Here $\alpha(\phi_k)$ is some function of the annealing parameter

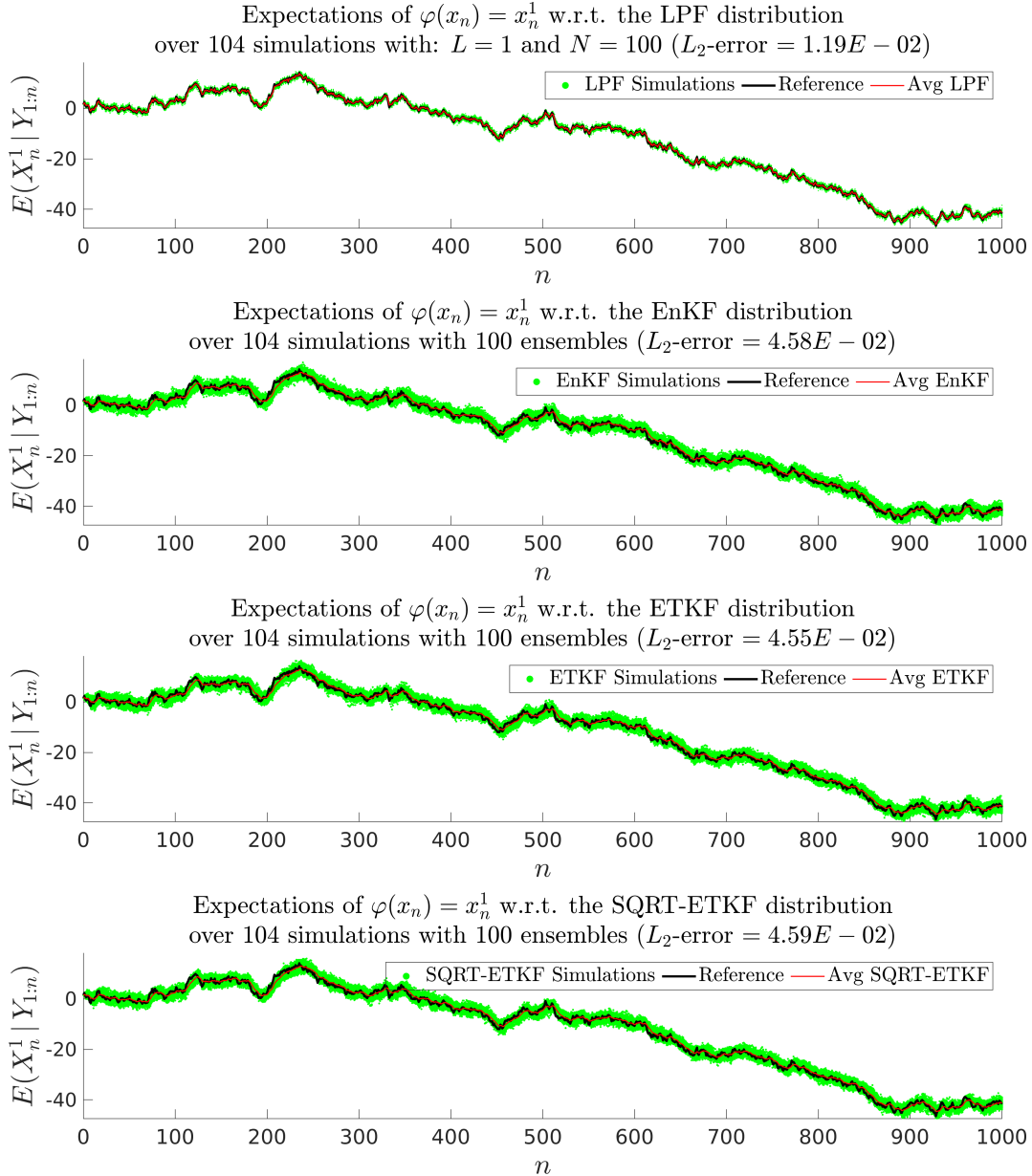


Figure 3: (Linear-Gaussian Model) A comparison of approximations of $\pi_n(\varphi)$ with $\varphi(x_{1:n}) = x_n^1$ as found by LPF, EnKF, ETKF and ETKF-SQRT. The green circle/dots represent the cloud of different simulations of each filter. The black curve is the reference from the KF and the red curve is the average over 104 independent simulations of each method.

ϕ_k , e.g. $\alpha(\phi_k) = (\phi_k + 2)/(\phi_k + 1)$. We use an S iterations of $K_{k,n}$, with S varying between 15 and 25.

In the next subsection, we test the algorithm on the above model with three choices of the function q_n and three corresponding different scenarios for the observations. In the first example, the signal is fully observed and the time frequency is $\hat{k} = 1$. For the second example, the signal is fully observed and $\hat{k} = 3$. In the third example, we observe around 40% of the state coordinates and the time frequency is $\hat{k} = 1$. Then, we compare results from the lagged particle filter (LPF) [Algorithm 3](#), the ensemble Kalman filter (EnKF) [[12](#), [13](#)], the ensemble transform Kalman filter (ETKF) [[4](#)] and the ETKF with square-root (ETKF-SQRT) [[22](#), [25](#)].

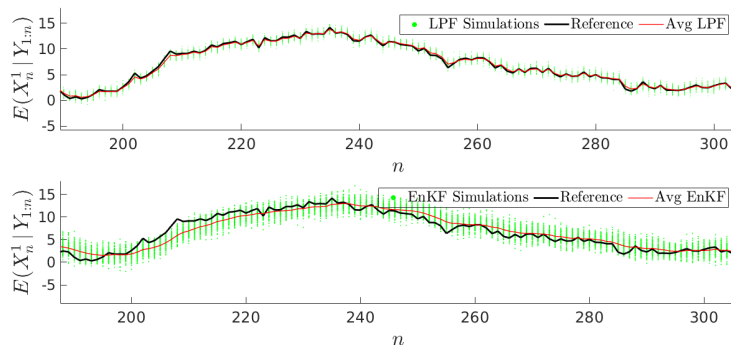


Figure 4: (Linear-Gaussian Model) We zoom in the region $[190, 305] \times [-5, 15]$ in Figure 3 for the LPF and EnKF (the rest of the methods are very similar to the EnKF.)

4.1.1 Linear-Gaussian Model

Here, we set $q_n(x) = x$, $T = 10^3$, $N = 100$, $N^* = 0.8 N$, $d = d_y = 500$, $R_1^{1/2} = \frac{1}{\sqrt{2}}I_d$, $R_2^{1/2} = 0.1I_{d_y}$, $C = I_d$, $X_0 = 1.5 \cdot \mathbf{1}$, where $\mathbf{1}$ is a vector of ones. We consider a lag of 2, i.e., $L = 1$. For $n \geq L + 1$, we take $\mu_{n-L}(x_{n-L+1}) = \mathcal{N}(x_{n-L+1}; m_{n-L+1}, P_{n-L+1})$, where m_{n-L+1} and P_{n-L+1} are the mean and the covariance of the Kalman predictor distribution at time step $n - L + 1$. The time frequency of the observations is $\hat{k} = 1$, and thus $M = 10^3$.

In the left panel of Figure 1, we plot the average of $\pi(\varphi)$ with $\varphi(x_{1:n}) = x_n$ after 104 independent runs. Here the subscript \cdot refers to the filter of interest (LPF, EnKF, ETKF, ETKF-SQRT). Since this model is linear-Gaussian, we use the Kalman filter (KF) as a reference. The right panel of Figure 1 shows the relative absolute error for each filter. Additionally, the relative L_2 -error $\|\pi(\varphi) - \pi_{KF}(\varphi)\|_2 / \|\pi_{KF}(\varphi)\|_2$ is shown at the title of each subfigure. We elaborate further in Figure 2 and present a histogram of the relative absolute errors. The histogram shows that for the same number of particles/ensembles, around 60% of the LPF errors are less than 0.025 and only 23% in the rest of the methods. Finally, Figure 3 shows a cloud plot of estimates of $\mathbb{E}(x_n^1 | y_{1:n})$ for each independent run over time, where $1 \leq n \leq T$ and x_n^1 being the first coordinate of x_n . The LPF method tracks the KF very well on average (as illustrated in Figure 4) and it shows considerable less variability between different independent runs. Similar comparisons will be carried for the other examples to follow.

4.1.2 Lorenz 96 Model

We test the algorithm on the Lorenz 96 model [27], perturbed by a Gaussian noise, along with noisy observations. The function $q_n(X_{n-1})$ is given by the 4th-order Runge Kutta approximation of the Lorenz 96 system

$$\frac{dX_{n-1}^i}{dt} = X_{n-1}^{i-1}(X_{n-1}^{i+1} - X_{n-1}^{i-2}) - X_{n-1}^i + 8, \quad (14)$$

when run one step in time with step size Δ_t . Here X_{n-1}^i is the i th-component of X_{n-1} with periodic conditions, $X_{n-1}^0 = X_{n-1}^d$, $X_{n-1}^{-1} = X_{n-1}^{d-1}$ and $X_{n-1}^{d+1} = X_{n-1}^1$.

For the numerical results, we set $\mu_{n-L}(x_{n-L+1}) = \mathcal{N}(x_{n-L+1}; m_{n-L+1}, P_{n-L+1})$ for $n \geq L + 1$, where m_{n-L+1} and P_{n-L+1} are the mean and covariance of the ETKF-SQRT predictor distribution at the time step $n - L + 1$. We set $N = 100$, $N^* = 0.8N$, $T = 10^3$, $\hat{k} = 3$ (i.e. $M = 333$), $d = d_y = 40$, $R_1^{1/2} = 0.1I_d$, $R_2^{1/2} = 0.1I_{d_y}$, $\Delta_t = 0.05$, $L = 1$, $C = I_d$, $X_0^{20} = 8.01$, $X_0^i = 8$ for $1 \leq i < 20$, $20 < i \leq d$.

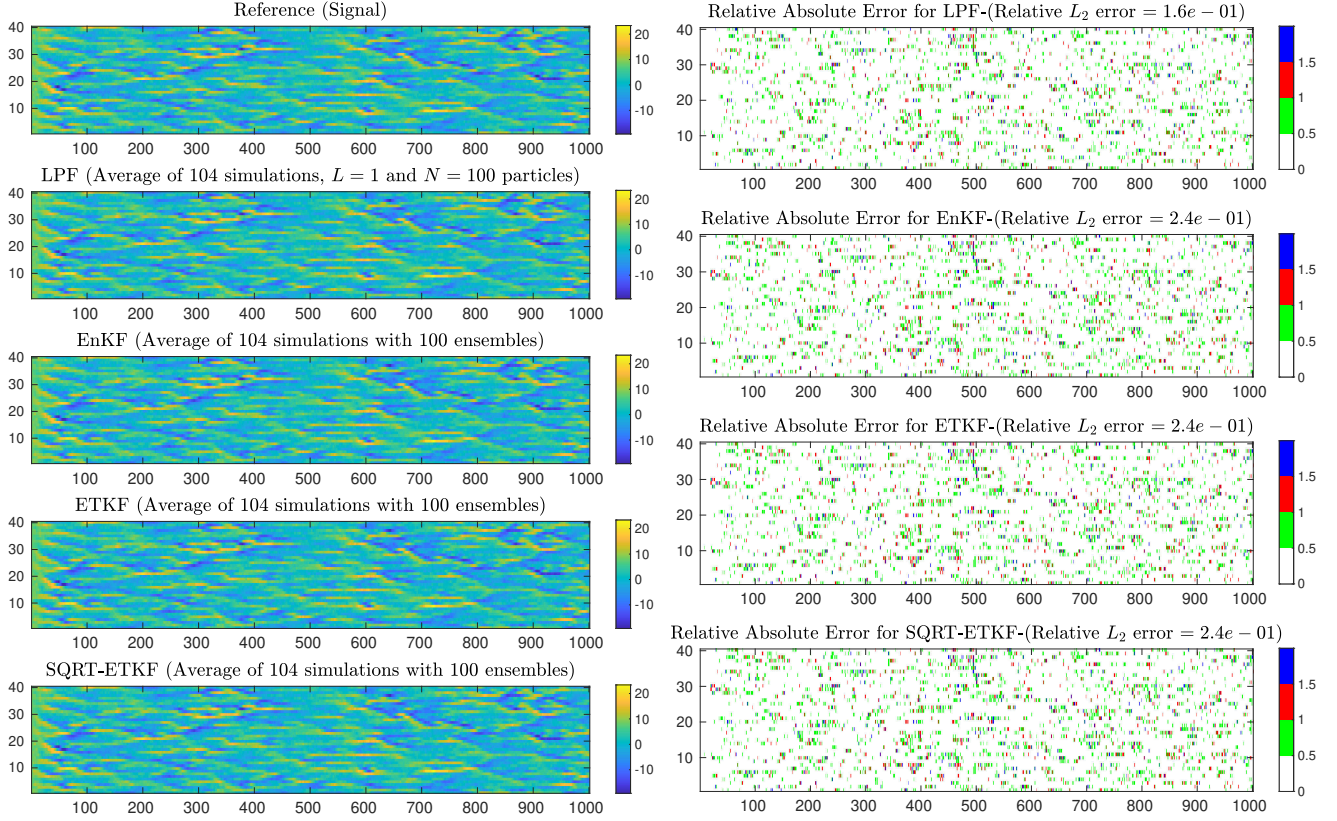


Figure 5: (Lorenz 96 Model) A comparison between the different filters. Details as in Figure 1 except the top-left panel that shows the true signal (9) as the reference.

Our results are presented similarly to before. The left and right panels of Figure 5 show approximations of $\pi_n(\varphi)$ with $\varphi(x_{1:n}) = x_n$ and the associated relative absolute errors for LPF, EnKF, ETKF and ETKF-SQRT. The reference in this example is the true signal from which the data was generated. In Figure 6, we present as before a histogram of the relative errors and observe that for the same number of particles/ensembles about 31% of the LPF errors are less than 0.025 and 29% in the rest of the methods. Finally Figure 7 shows the approximations of $\pi_n(\varphi)$ against time with $1 \leq n \leq T$ and $\varphi(x_{1:n}) = x_n^1$. We note that in this example the LPF performs marginally better than the rest of the methods.

4.1.3 Conservative Shallow-Water Model

We apply the algorithm on the conservative shallow-water model [28]. Function $q_n(X_{n-1})$ is the finite volume (FV) solution of the shallow-water equations (SWE) given below (one step in time with step size Δ_t). Let $(x, y) \in \Omega = [0, a]^2$, for some $a > 0$, and let $h(t, x, y)$, $v(t, x, y)$ and $u(t, x, y)$ represent the fluid column height, the fluid's horizontal velocity in the x -direction and the fluid's horizontal velocity in the y -direction, respectively, at position (x, y) and time $t > 0$.

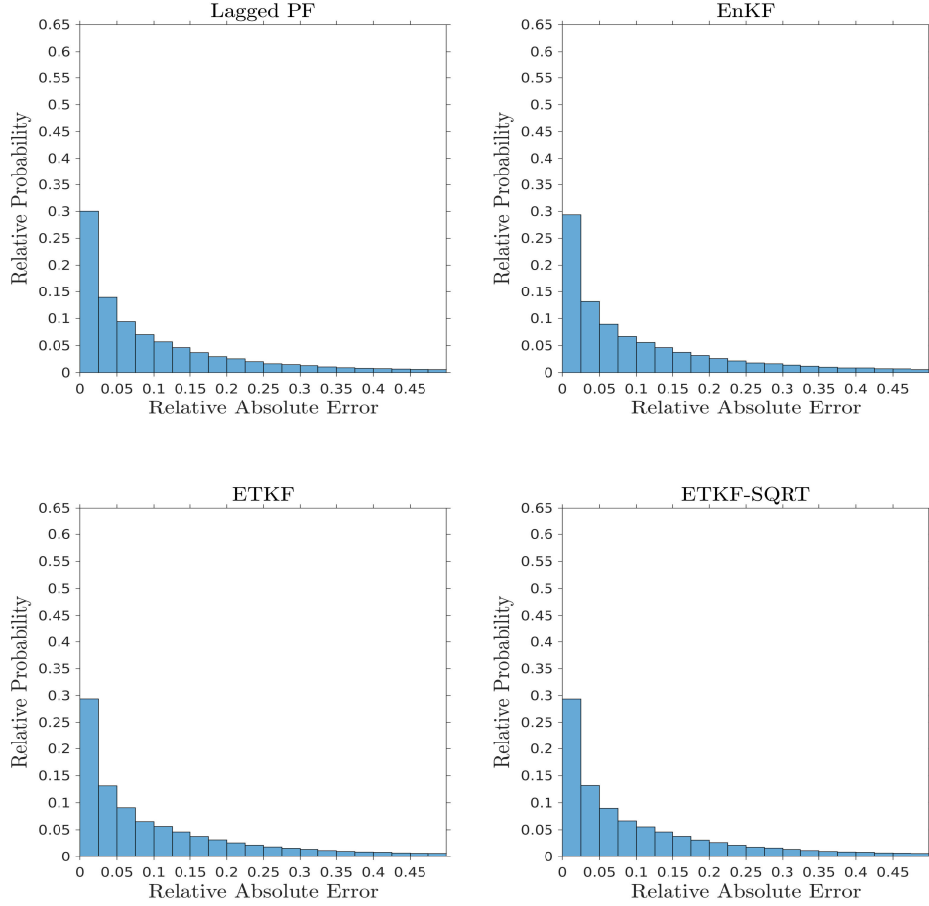


Figure 6: (Lorenz 96 Model) A histogram of the relative absolute errors for each filter. Details as in Figure 2.

The conservative form of SWE is as follows

$$\begin{aligned} \frac{\partial h}{\partial t} + \frac{\partial(hu)}{\partial x} + \frac{\partial(hv)}{\partial y} &= 0, \\ \frac{\partial(hu)}{\partial t} + \frac{\partial}{\partial x}(hu^2 + \frac{1}{2}gh^2) + \frac{\partial(huv)}{\partial y} &= 0, \\ \frac{\partial(hv)}{\partial t} + \frac{\partial(huv)}{\partial y} + \frac{\partial}{\partial x}(hv^2 + \frac{1}{2}gh^2) &= 0, \end{aligned}$$

where g is the gravitational acceleration. To write the equations in a compact form, we introduce three vectors $U = [h, hu, hv]^\top$, $A(U) = [hu, hu^2 + \frac{1}{2}gh^2, huv]^\top$ and $B(U) = [hv, huv, hv^2 + \frac{1}{2}gh^2]^\top$. Under this notation, we have

$$\frac{\partial U}{\partial t} + \frac{\partial A(U)}{\partial x} + \frac{\partial B(U)}{\partial y} = 0. \quad (15)$$

Given the number of grid points in the x and y directions, d_g , we set $\Delta_x = \Delta_y = a/d_g$. We will refer to the grid $\{(x_i, y_j) \in \Omega : x_i = (i-1)\Delta_x, y_j = (j-1)\Delta_y, \text{ for } i, j \in \{1, \dots, d_g\}\}$ as the physical grid. Consider the uniform grid with FV cells $I_{i,j} = [x_{i-1/2}, x_{i+1/2}] \times [y_{j-1/2}, y_{j+1/2}]$ centered at $(x_i, y_j) = (\frac{x_{i-1/2} + x_{i+1/2}}{2}, \frac{y_{j-1/2} + y_{j+1/2}}{2})$, for all $i, j \in \{0, \dots, d_g + 1\}$, with grid size of

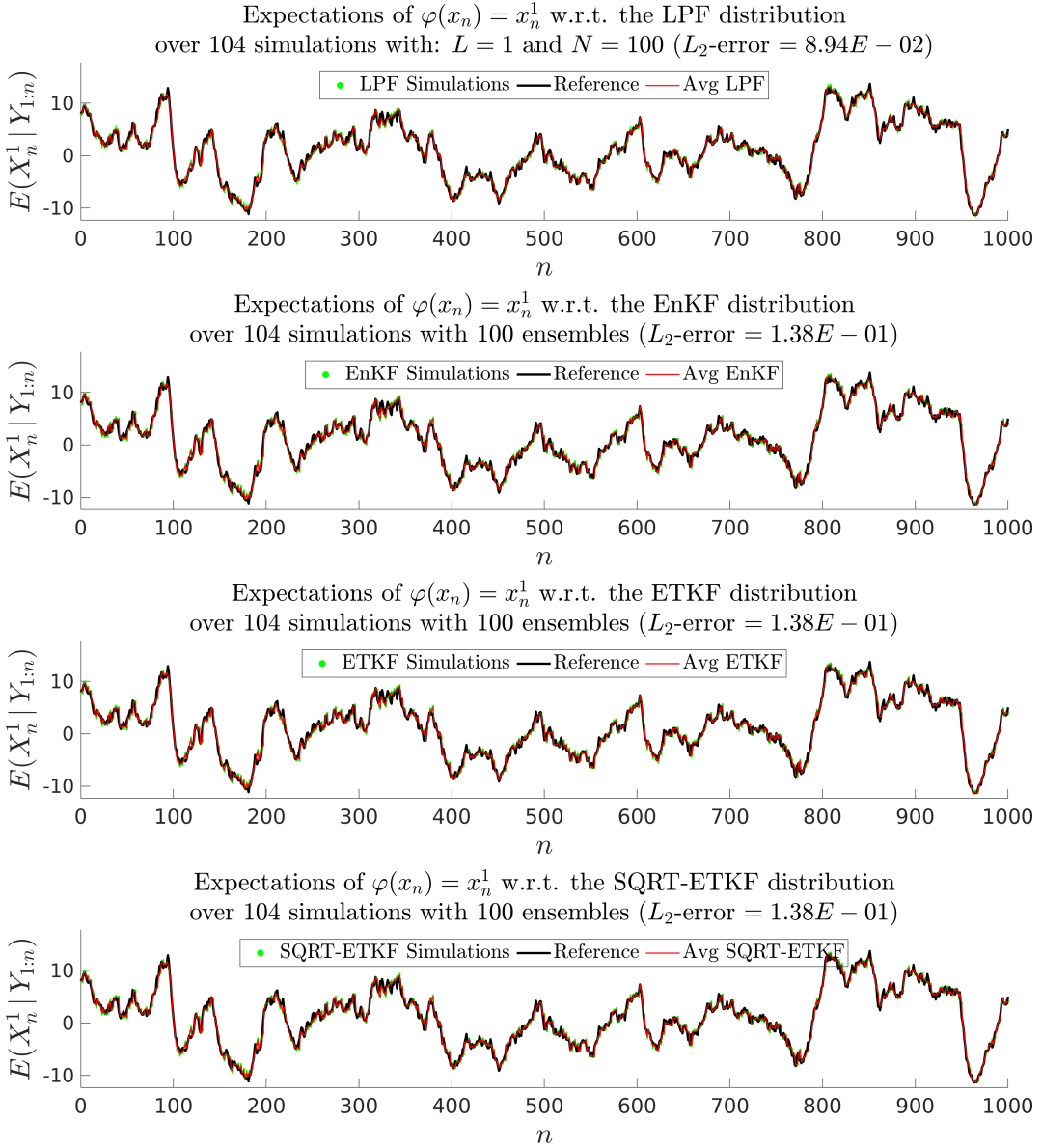


Figure 7: (Lorenz 96 Model) Approximation clouds of $\pi_n(\varphi)$ against n with $\varphi(x_{1:n}) = x_n^1$ for each method. Details as in Figure 3 except that in the black curve the true signal is plotted as the reference.

$(d_g + 2)^2$. Then,

$$U_{i,j}^{n+1} = U_{i,j}^n - \frac{\Delta t}{\Delta x} (F_{i+\frac{1}{2},j}^n - F_{i-\frac{1}{2},j}^n) - \frac{\Delta t}{\Delta y} (G_{i,j+\frac{1}{2}}^n - G_{i,j-\frac{1}{2}}^n),$$

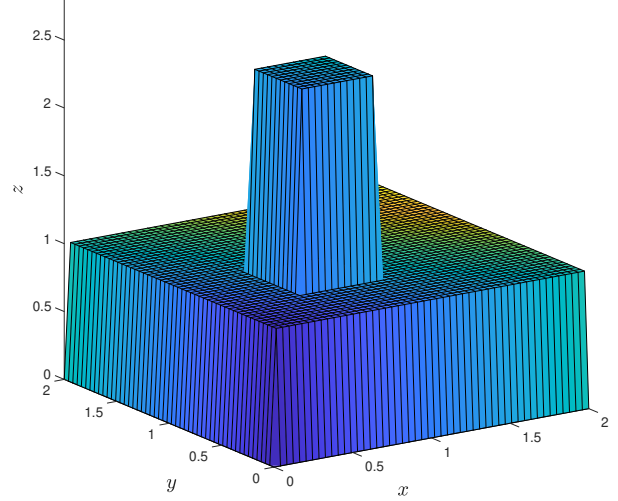
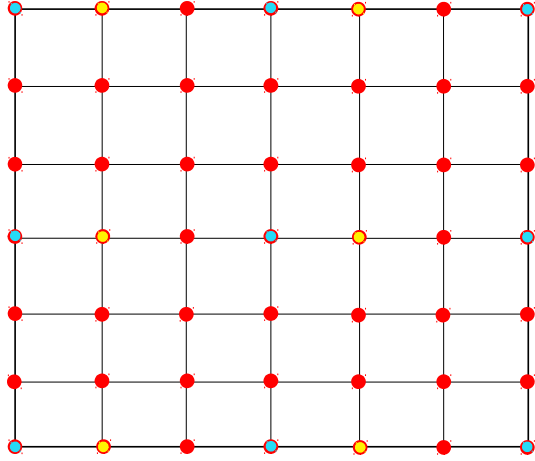


Figure 8: The figure on the left represents the physical grid where the red, cyan and yellow circles are the grid points at which h , u and v are observed, respectively. The figure on the right shows the initial height of the water.

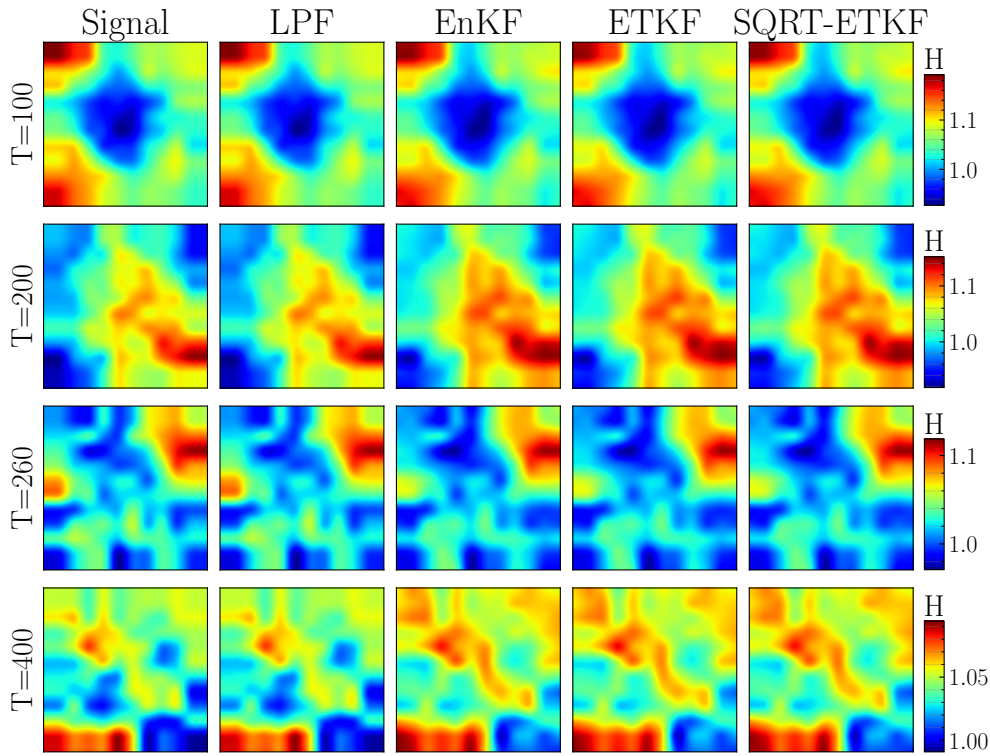


Figure 9: (SWE Model) Snapshots of the height h at four different times: 100 (first row), 200 (second row), 260 (third row) & 400 (fourth row). From left to right: the signal, the mean of 50 simulations of the LPF with 100 particles and the mean of 50 simulations of the ENKF, ETKF & ETKF-SQRT with 1000 ensembles, respectively.

where F and G are the numerical Lax-Friedrichs fluxes given by

$$\begin{aligned}
 F_{i+\frac{1}{2},j}^n &= \frac{1}{2}[A(U_{i,j}^n) + A(U_{i+1,j}^n)] - \frac{1}{2}\lambda_{i+\frac{1}{2},j,\max}^x [U_{i+1,j}^n - U_{i,j}^n] \\
 F_{i-\frac{1}{2},j}^n &= \frac{1}{2}[A(U_{i,j}^n) + A(U_{i-1,j}^n)] - \frac{1}{2}\lambda_{i-\frac{1}{2},j,\max}^x [U_{i,j}^n - U_{i-1,j}^n] \\
 G_{i,j+\frac{1}{2}}^n &= \frac{1}{2}[B(U_{i,j}^n) + B(U_{i,j+1}^n)] - \frac{1}{2}\lambda_{i,j+\frac{1}{2},\max}^y [U_{i,j+1}^n - U_{i,j}^n] \\
 G_{i,j-\frac{1}{2}}^n &= \frac{1}{2}[B(U_{i,j}^n) + B(U_{i,j-1}^n)] - \frac{1}{2}\lambda_{i,j-\frac{1}{2},\max}^y [U_{i,j}^n - U_{i,j-1}^n],
 \end{aligned}$$

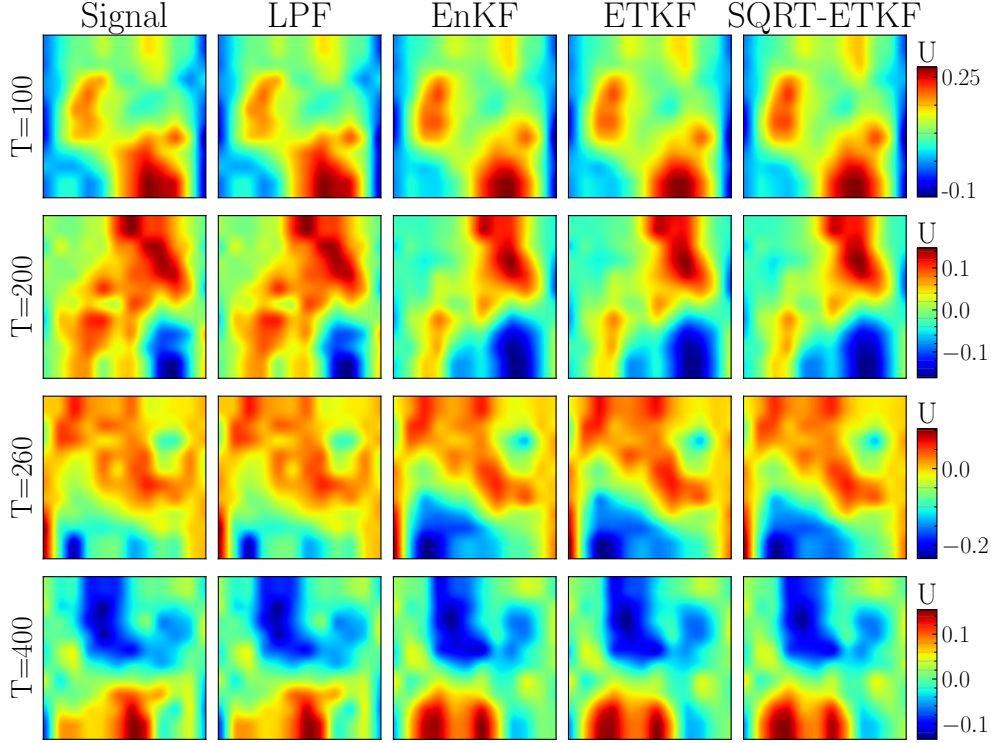


Figure 10: (SWE Model) Snapshots of the horizontal x -axis velocity u at four different times: 100 (first row), 200 (second row), 260 (third row) & 400 (fourth row). From left to right: the signal, the mean of 50 simulations of the LPF with 100 particles and the mean of 50 simulations of the ENKF, ETKF & ETKF-SQRT with 1000 ensembles, respectively.

where $\lambda_{i^*,j^*,\max}^x$ is the maximum eigenvalue of the Jacobian matrix $\partial A(U)/\partial U$ evaluated at U_{i^*,j^*}^n . The eigenvalues are $\{u_{i^*,j^*} \pm \sqrt{gh_{i^*,j^*}}, u_{i^*,j^*}\}$. We set $\lambda_{i^*,j^*,\max}^x = |u_{i^*,j^*}| + \sqrt{gh_{i^*,j^*}}$. Similarly, $\lambda_{i^*,j^*,\max}^y$ is the maximum eigenvalue of the Jacobian matrix $\partial B(U)/\partial U$ evaluated at U_{i^*,j^*}^n and we take it to be $|v_{i^*,j^*}| + \sqrt{gh_{i^*,j^*}}$. The initial value conditions are $h(0, x, y) = h_0(x, y)$ and $v(0, x, y) = u(0, x, y) = 0$ for $(x, y) \in \Omega$. We use reflective boundary conditions with $u(t, 0, y) = u(t, a, y) = 0$ and $v(t, x, 0) = v(t, x, a) = 0$. The dimension of the state vector in (9) is $d = 3d_g^2$.

We write $(h_{i,j})_{1 \leq i,j \leq d_g}$, $(u_{i,j})_{1 \leq i,j \leq d_g}$ and $(v_{i,j})_{1 \leq i,j \leq d_g}$ in a vector representation as $(h_i)_{1 \leq i \leq d_g^2}$, $(u_i)_{1 \leq i \leq d_g^2}$ and $(v_i)_{1 \leq i \leq d_g^2}$. Then we write the state vector $X_n = (X_n^i)_{1 \leq i \leq d}$ in (9) as

$$X_n = [(h_i)_{1 \leq i \leq d_g^2}, (u_i)_{1 \leq i \leq d_g^2}, (v_i)_{1 \leq i \leq d_g^2}]^T \in \mathbb{R}^d.$$

We take $a = 2$, $d_g = 35$ (so $d = 3675$) and $h_0(x, y) = 2.5$ for $0.5 \leq x, y \leq 1$ and one elsewhere. The spatial frequency at which height is observed is 1, i.e., it is observed at all grid points, while the spatial frequency at which u and v are observed is 3. However, the first component that is observed in u is $X_n^{d_g^2+1}$ and the first component that is observed in v is $X_n^{2d_g^2+2}$ (see Figure 8). Therefore we take the matrix C as

$$C = \left[\begin{array}{cccc|cccc} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_{d_g^2} & \mathbf{e}_{d_g^2+1} & \mathbf{e}_{d_g^2+4} & \mathbf{e}_{d_g^2+7} & \cdots \\ \mathbf{e}_{2d_g^2+2} & \mathbf{e}_{2d_g^2+5} & \mathbf{e}_{2d_g^2+8} & \cdots & & & & \end{array} \right]^T,$$

where \mathbf{e}_ℓ is a vector in \mathbb{R}^d with one at position ℓ and zeros everywhere else. We set $N = 100$,

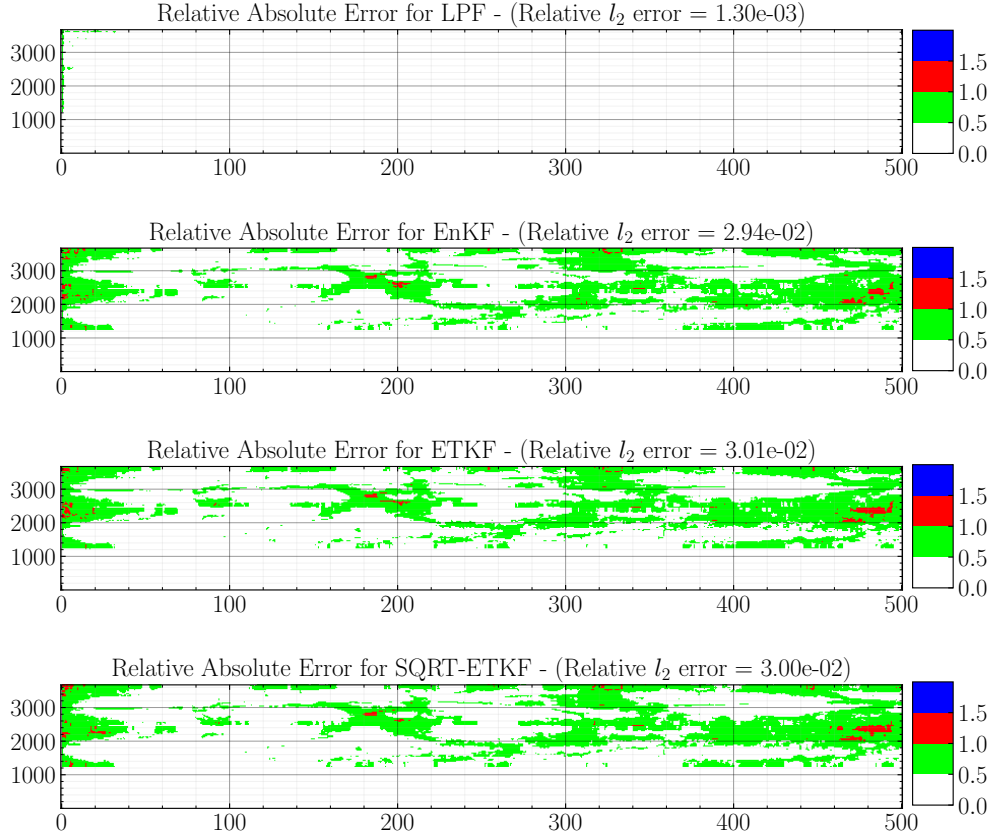


Figure 11: (SWE Model) The relative absolute errors. The horizontal axis represents the time parameter n , while the vertical axis represents the signal's coordinates. We also include the relative l_2 -norm in the titles.

$N^* = 0.5 N$, $T = 500$, $\hat{k} = 1$ (i.e. $M = 500$), $R_1^{1/2} = 0.01I_d$, $R_2^{1/2} = 0.01I_{d_y}$ and

$$\Delta_t = \frac{\text{CFL} \times \Delta_x}{\max \left\{ \max_{0 \leq i, j \leq d_g+1} \left(\lambda_{i \pm \frac{1}{2}, j, \max}^x \right), \max_{0 \leq i, j \leq d_g+1} \left(\lambda_{i, j \pm \frac{1}{2}, \max}^y \right) \right\}},$$

where CFL is the Courant-Friedrich-Levy number which we set to 0.5.

Figure 9 and Figure 10 display, respectively, snapshots of the perturbed and filtered water's height and water's horizontal x -axis velocity at times 100, 200, 260 & 400 using the different methods. We ran 50 independent simulations of each of the EnKF, ETKF and ETKF-SQRT methods with 1000 ensembles because they provided unreliable estimates for lower ensemble sizes. We ran 50 simulations of the LPF with only 100 particles. Figure 11 shows the relative absolute errors for each filter. In Figure 12 a histogram of the relative absolute errors is provided. From the histograms we note that around 91% of the relative absolute errors of the LPF are less than 0.01 whereas in the other methods they are around 41%. Finally, Figure 13 and Figure 14 show respectively cloud plots of the approximations of $\mathbb{E}(x_n^1 | y_{1:n})$ and $\mathbb{E}(x_n^{d_g^2+1} | y_{1:n})$ for $1 \leq n \leq T$. As in the linear Gaussian example the variability is much less for the LPF. All plots show that the LPF with only 100 particles outperforms the other methods in terms of accuracy and this is illustrated by the one order of magnitude improvement in terms of relative L_2 errors.

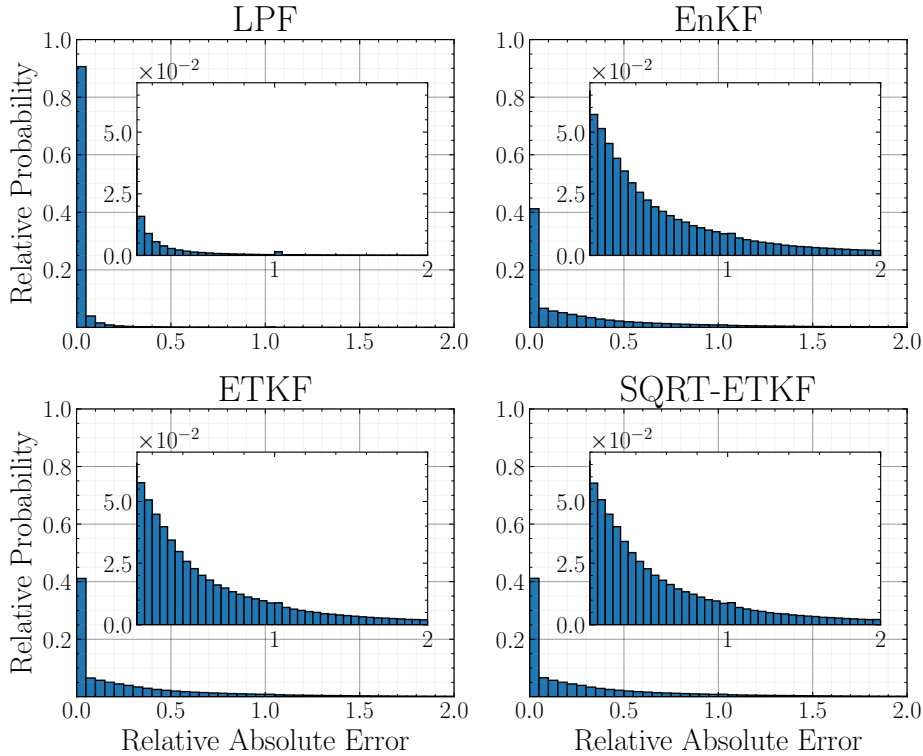


Figure 12: (SWE Model) A histogram of the relative absolute errors for each filter. The relative probability here is defined as the number of elements in the bin divided by the total number of elements, which is $d \times (T + 1)$. For readability, we zoom in the region $[0, 2] \times [0, 0.1]$ and include it in each histogram.

5 Discussion

In the literature fixed-lag methods have been mainly in the context of improving smoothing approximations. The original motivation was to alleviate path degeneracy and justify biased approximations that resample only paths within a fixed lag to the current time, [18, 23]. Later these ideas have been complemented with MCMC for the purpose of parameter estimation [20]. In addition, SMC sampling techniques for a fixed lag have been used without bias to enhance particle filtering [11]. All these contributions are very close to the spirit of this paper, but to the best of our knowledge this paper is the first to use fixed lag sampling ideas for high dimensional SSMs.

We have proposed a fixed-lag particle filter that under certain conditions on the model has provably stable errors as the dimension of the model grows. Our proposed algorithm also incorporates adaptive tempering steps, which have been crucial for other successful SMC approaches for high-dimensional problems [21, 7]. Tempering was also crucial in the theory of [1, 2], which was originally developed for independent states. The conditions used in this paper are general enough to go beyond this and allow for dependency between some coordinates of the states or observations and at the same time be able to apply the results in [1, 2]. As mentioned earlier, our algorithm inherently contains a bias as other similar fixed-lag approaches, but we have shown this bias to be bounded and controlled by the lag parameter L and the number of particles N . Regarding performance, that of the proposed method was found to be superior to commonly used methods such as the EnKF, ETKF in our numerical examples. In particular the results in the challenging

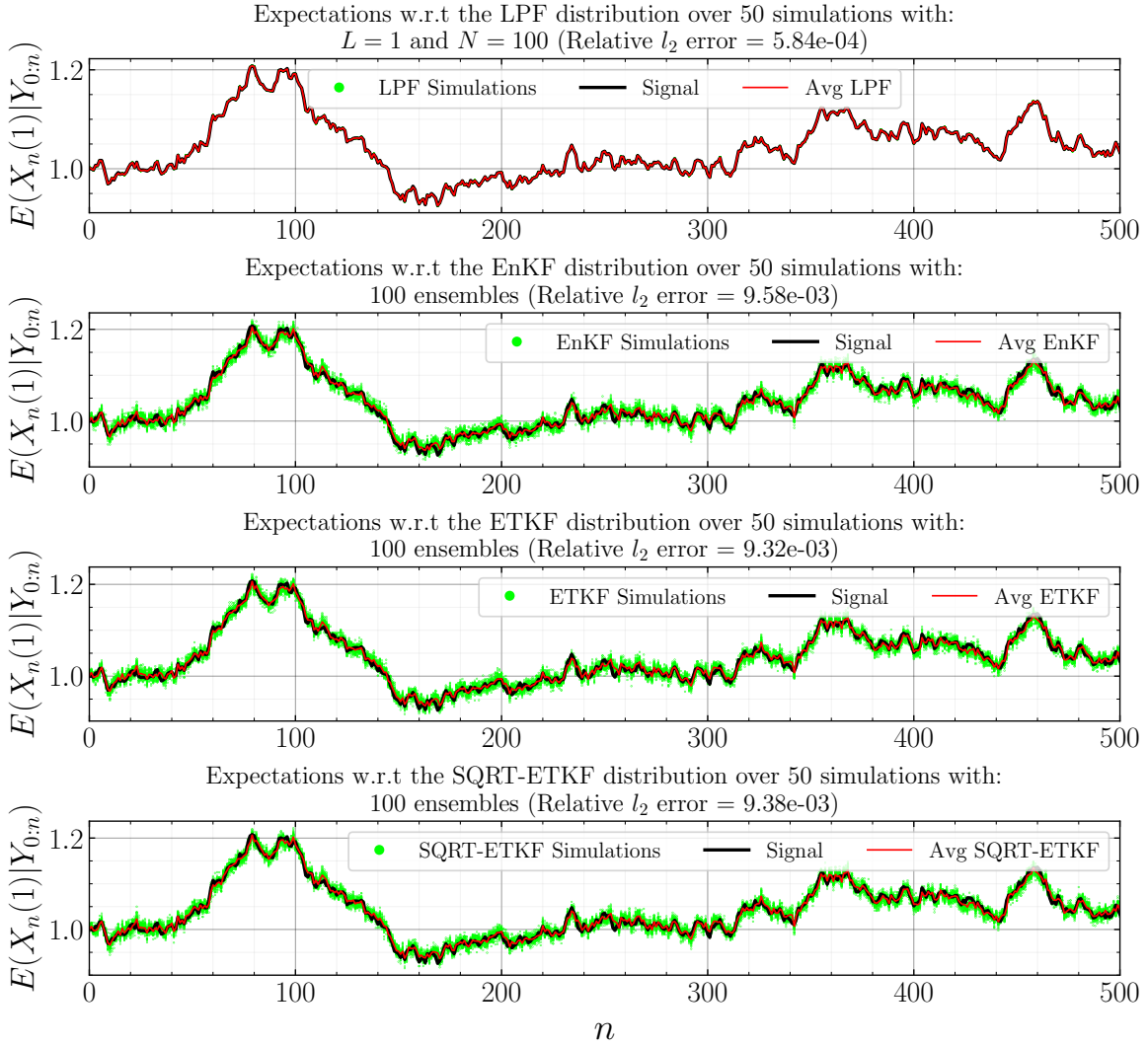


Figure 13: (SWE Model) A comparison of the expectation of $\varphi(x_{1:n}) = x_n^1$ as found by LPF, EnKF, ETKF and ETKF-SQRT. The green circles represent the different simulations of each filter. The black curve is the reference (in this example it is the signal) and the red curve is the average of the simulations.

shallow water equation showed impressive improvement and accuracy. Future work can include more case studies in even more challenging applications and comparisons with current state-of-art methods.

Acknowledgements

AJ & HR were supported by KAUST baseline funding. The work of DC has been partially supported by European Research Council (ERC) Synergy grant STUOD-DLV-8564. NK was supported by a J.P. Morgan A.I. Research Award.

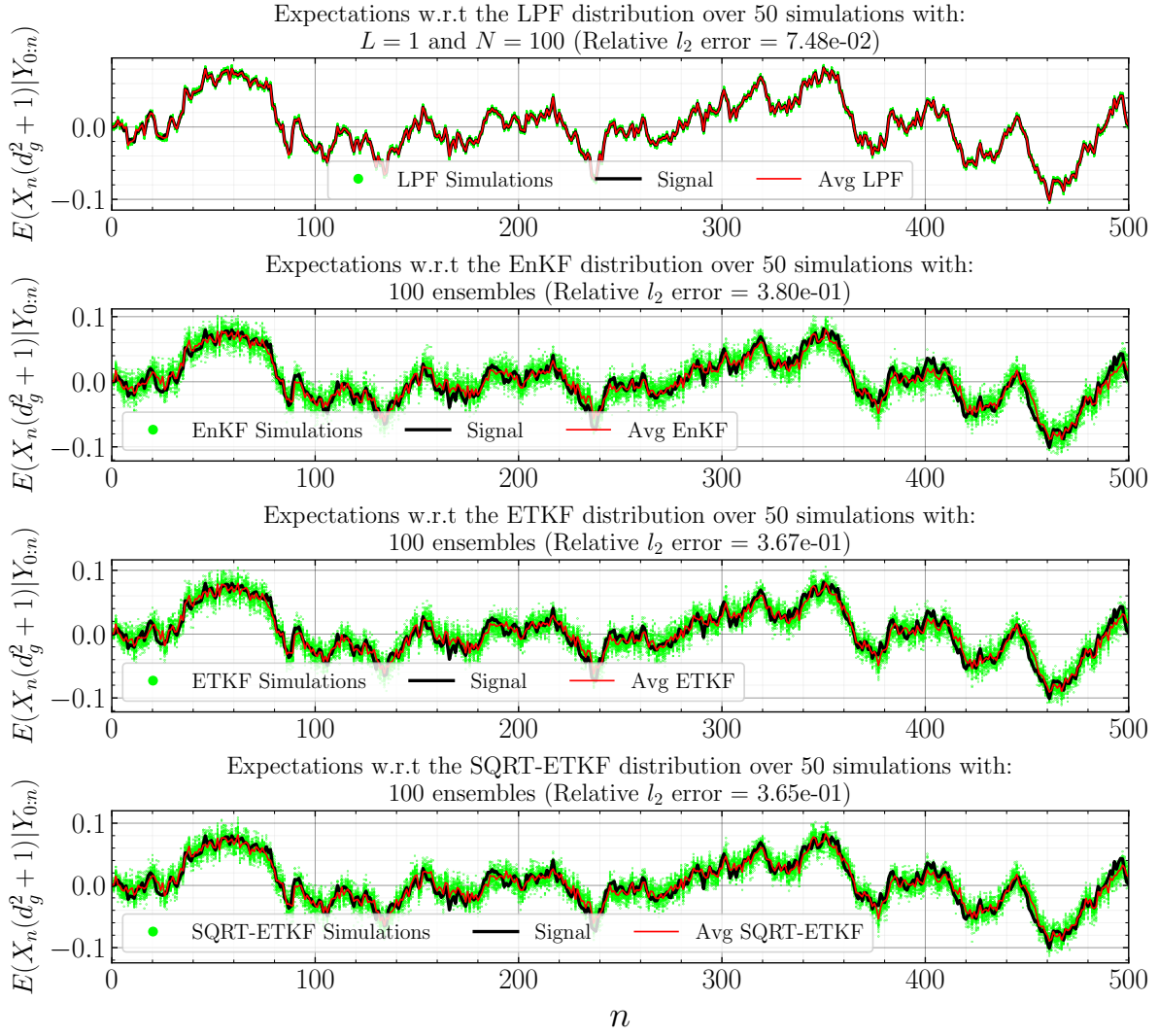


Figure 14: (SWE Model) A comparison of the expectation of $\varphi(x_{1:n}) = x_n^{d_g^2+1}$ as found by LPF, EnKF, ETKF and ETKF-SQRT. The green circles represent the different simulations of each filter. The black curve is the signal and the red curve is the average of the simulations.

A Technical Results

For completeness, we describe briefly a result that expresses in a more general formulation findings from the works in [1, 3]. Consider a sequence of distributions on the space \mathbb{E}^{L^d} , with $\mathbb{E} \subset \mathbb{R}$ assumed compact and a fixed $L \in \mathbb{N}$. We define the bridging sequence of densities:

$$\Pi_k := \tilde{\Pi}^{1-\phi_k} \times \Pi^{\phi_k}, \quad \phi_1 = 0, \quad \phi_{k+1} - \phi_k = \frac{1}{d}, \quad k = 1, 2, \dots, d+1,$$

The initial and target densities, $\tilde{\Pi} : \mathbb{E}^{Ld} \rightarrow \mathbb{R}$ and $\Pi : \mathbb{E}^{Ld} \rightarrow \mathbb{R}$ respectively, admit the factorisation, for fixed $m \in \mathbb{N}$, $\tilde{\Pi}^{(m)}, \Pi^{(m)} \in \mathcal{P}(\mathbb{E}^{Lm})$, and $\tilde{\Pi}^1, \Pi^1 \in \mathcal{P}(\mathbb{E}^L)$:

$$\begin{aligned}\tilde{\Pi}(x_{1:L}) &= \tilde{\Pi}^{(m)}(x_{1:L}^{1:m}) \times \prod_{j=m+1}^d \tilde{\Pi}^1(x_{1:L}^j), \\ \Pi(x_{1:L}) &= \Pi^{(m)}(x_{1:L}^{1:m}) \times \prod_{j=m+1}^d \Pi^1(x_{1:L}^j).\end{aligned}$$

Note that we have indexed the Ld co-ordinates in an array format, using two indicators, one subscript for the ‘block’, $l = 1, \dots, L$, and one superscript for placement in a block, $j = 1, \dots, d$. Consider related Markov kernels $\mathsf{K}_k : \mathbb{E}^{Ld} \times \mathcal{B}(\mathbb{E}^{Ld}) \rightarrow [0, 1]$, $k = 1, \dots, d+1$, such that $\Pi_k \mathsf{K}_k = \Pi_k$. Also, kernels admit a factorisation, for $\mathsf{K}_k^{(m)} : \mathbb{E}^{Lm} \times \mathcal{B}(\mathbb{E}^{Lm}) \rightarrow [0, 1]$ and $\mathsf{K}_k^1 : \mathbb{E}^L \times \mathcal{B}(\mathbb{E}^L) \rightarrow [0, 1]$:

$$\mathsf{K}_k(x_{1:L}, dz_{1:L}) := \mathsf{K}_k^{(m)}(x_{1:L}^{1:m}, dz_{1:L}^{1:m}) \times \prod_{j=m+1}^d \mathsf{K}_k^1(x_{1:L}^j, dz_{1:L}^j),$$

so that $\Pi_k^{(m)} \mathsf{K}_k^{(m)} = \Pi_k^{(m)}$, $\Pi_k^1 \mathsf{K}_k^1 = \Pi_k^1$, with $\Pi_k^{(m)}, \Pi_k^1$ defined in an obvious way. One now applies the SMC sampler in [Algorithm 1](#), in the setting we have formulated herein, under the more general consideration that the initial positions from the N particles are chosen arbitrarily from within the state space \mathbb{E}^{Ld} . The sequence of importance sampling weights are given as:

$$W_{k+1}^{(i)} = W_k^{(i)} \times \left(\frac{\Pi}{\tilde{\Pi}}\right)(x_{1:L,k}^{(i)})^{\phi_{k+1} - \phi_k}, \quad k = 2, \dots, d, \quad i = 1, 2, \dots, N.$$

We denote here by $x_{l,k}^{(i)}$ the l th ‘block’, $l = 1, \dots, L$, of i th particle after the application of kernels $\mathsf{K}_2, \dots, \mathsf{K}_k$, $k = 2, \dots, d+1$. Simple calculations give that:

$$\begin{aligned}\log W_{d+1}^{(i)} &= \frac{1}{d} \sum_{k=1}^d \log \left(\left(\frac{\Pi}{\tilde{\Pi}}\right)(x_{1:L,k}^{(i)}) \right) \\ &= \frac{1}{d} \sum_{k=1}^d \log \left(\left(\frac{\Pi^{(m)}}{\tilde{\Pi}^{(m)}}\right)(x_{1:L,k}^{(i),1:m}) \right) + \frac{1}{d} \sum_{k=1}^d \sum_{j=m+1}^d \log \left(\left(\frac{\Pi^1}{\tilde{\Pi}^1}\right)(x_{1:L,k}^{(i),j}) \right).\end{aligned}$$

Notice the two above sums involve two *independent* Markov sequences of random variables.

The work in [\[1\]](#) can now provide the following results. We adopt the notation for a *continuum* of involved distributions and Markov kernels, by using subscripts $s, t \in [0, 1]$ below, at the same position that we have so far used $n \in \mathbb{N}$. Also, the σ -algebra \mathcal{F}_0 used below corresponds to the information about the initial position of the Markov chains.

Assumption A.1. *Consider the following assumptions:*

(A1) *There exist $\theta \in (0, 1)$, $(\zeta^{(m)}, \zeta) \in \mathcal{P}(\mathcal{B}(\mathbb{E}^{mL})) \times \mathcal{P}(\mathcal{B}(\mathbb{E}^L))$, such that for each $s \in [0, 1]$:*

$$\mathsf{K}_s^{(m)}(x_{1:L}^{1:m}, A) \geq \theta \zeta^{(m)}(A), \quad \text{for all } (x_{1:L}^{1:m}, A) \in \mathbb{E}^{mL} \times \mathcal{B}(\mathbb{E}^{mL}),$$

$$\mathsf{K}_s(x_{1:L}, A) \geq \theta \zeta(A), \quad \text{for all } (x_{1:L}, A) \in \mathbb{E}^L \times \mathcal{B}(\mathbb{E}^L).$$

(A2) There exists $C < \infty$ such that for any $s, t \in [0, 1]$ we have

$$\begin{aligned} \sup_{x_{1:L}^{1:m} \in \mathbb{E}^{mL}} \left\| \mathbb{K}_s^{(m)}(x_{1:L}^{1:m}, \cdot) - \mathbb{K}_t^{(m)}(x_{1:L}^{1:m}, \cdot) \right\|_{\text{TV}} &\leq C|s - t|, \\ \sup_{x_{1:L} \in \mathbb{E}^L} \left\| \mathbb{K}_s(x_{1:L}, \cdot) - \mathbb{K}_t(x_{1:L}, \cdot) \right\|_{\text{TV}} &\leq C|s - t|. \end{aligned}$$

(A3) There exists a $C < \infty$ such that:

$$\max \left\{ \sup_{x_{1:L}^{1:m} \in \mathbb{E}^{mL}} \left| \log \left(\frac{\Pi^{(m)}}{\widetilde{\Pi}^{(m)}} \right) (x_{1:L}^{1:m}) \right|, \sup_{x_{1:L} \in \mathbb{E}^L} \left| \log \left(\frac{\Pi_1^1}{\widetilde{\Pi}_1} \right) (x_{1:L}) \right| \right\} \leq C.$$

Theorem A.1. Under *Assumption A.1*:

(i) We have $\log W_{d+1}^{(i)} = o(1) + \overline{W}_{d+1}^{(i)}$, where $o(1)$ denotes here a term that converges weakly to 0 as $d \rightarrow \infty$, and we have defined:

$$\log \overline{W}_{d+1}^{(i)} = \frac{1}{d} \sum_{k=1}^d \sum_{j=m+1}^d \left\{ \log \left(\left(\frac{\Pi^1}{\widetilde{\Pi}^1} \right) (x_{1:L,k}^{(i),j}) \right) - \mathbb{E} \left[\log \left(\left(\frac{\Pi^1}{\widetilde{\Pi}^1} \right) (x_{1:L,k}^{(i),j}) \right) \mid \mathcal{F}_0 \right] \right\}.$$

The following weak limit holds, as $d \rightarrow \infty$:

$$\log \overline{W}_{d+1}^{(i)} \rightarrow_D \mathcal{N}(0, \sigma^2),$$

for variance

$$\sigma^2 = \int_0^1 \left\{ \Pi_s^1((\widehat{\varphi}_s)^2) - (\Pi_s^1(\widehat{\varphi}_s))^2 \right\} ds,$$

where $\widehat{\varphi}_s$ is the solution to the Poisson equation:

$$\varphi_s - \Pi_s^1(\varphi_s) = \widehat{\varphi}_s - \mathbb{K}_s^1(\widehat{\varphi}_s),$$

for $\varphi_s = \varphi_s(x_1^j, \dots, x_L^j) = s \cdot \log \left(\left(\frac{\Pi^1}{\widetilde{\Pi}^1} \right) (x_{1:L}^j) \right)$.

(ii) We have the weak limit, as $d \rightarrow \infty$, for any fixed $j^* \geq m + 1$:

$$x_{1:L,d+1}^{(i),j^*} \rightarrow_D \Pi^1,$$

and the one:

$$x_{1:L,d+1}^{(i),1:m} \rightarrow_D \Pi^{(m)}.$$

The weak limits above are jointly independent over particle index $i = 1, \dots, N$, and over the set of co-ordinates $\{1 : m\}, m + 1, \dots, M^*$, for any fixed $M^* > m$.

Proof. The result is contained in Appendix C of [1]. The crux of the proof for part (i) is the use of a CLT for triangular martingale arrays (notice that $\{\log \overline{W}_k^{(i)}\}_{k=1}^{d+1}$ is constructed to be a martingale process, for any $d \in \mathbb{N}$) and cumbersome but otherwise straightforward control of remainder terms. The term in the weights that involve the first m co-ordinates disappear in the limit $d \rightarrow \infty$, as there is only a finite number of them. Thus, the limiting values for the ‘standardised’ weights $\overline{W}_{d+1}^{(i)}$ involve only CLT obtained via co-ordinates $m + 1, \dots$. Result (ii) is obtained also in Appendix C of [1]. Under the dynamics of the inhomogeneous Markov chain $\{\mathbb{K}_k\}$, the particles carry out enough steps to reach the correct distribution after the execution of d steps, as $d \rightarrow \infty$. \square

B Proof of Proposition 3.1

Proposition 3.1 is a corollary of Theorem A.1. In particular, it suffices to apply Theorem A.1 for each time index $n \in \mathbb{N}$ of interest, under the choices:

$$\begin{aligned} \mathsf{L} &= L + 1, \quad \mathsf{\Pi} = \widehat{\pi}_n(dx_{(n-L):n}), \quad \widetilde{\mathsf{\Pi}} = \widehat{\pi}_{n-1}(dx_{(n-L):(n-1)}) \times f(x_{n-1}, dx_n), \\ \mathsf{K}_s^{(m)} &= \overline{K}_{s,n}^{(m)}, \quad \mathsf{K}_s = \overline{K}_{s,n}. \end{aligned}$$

In the above expressions $\widehat{\pi}_n(dx_{(n-L):n})$ denotes the marginal of $\widehat{\pi}_n$ on components $x_{(n-L):n}$; and similarly for $\widehat{\pi}_{n-1}(dx_{(n-L):(n-1)})$. The particular upper bounds in the L_2 -norms quoted in Proposition 3.1 follow immediately from the weak limits obtained via Theorem A.1 via relatively simple calculations that make use of the exchangeability of the limiting laws over the particle index. For the exact calculation, see Appendix A.2 of [2].

C Algorithm with Adaptive Resampling and Tempering

Algorithm 3 : Lagged Particle Filter with Adaptive Resampling

- **Initialization** $n = 1$:

We are given the annealing parameter $\phi_{1,1} \geq 0$, N^* (the resampling threshold), the number of MCMC steps S , the lag $L + 1$, and x_0 . We begin by sampling particles $\tilde{x}_1^{(i)}$, $i \in \{1, \dots, N\}$, from the proposal $\eta_1(x_1) = f(x_0, x_1)$ via (9). We will sequentially sample from

$$\tilde{\pi}_1^k(x_1) \propto \tilde{\gamma}_1^k(x_1) = f(x_0, x_1) g(x_1, y_1)^{\phi_{k,1}}$$

starting at $\tilde{\pi}_1^1$ ending at $\tilde{\pi}_1^{K_1} = \pi_1$ for some $K_1 \in \mathbb{N}$. For $i \in \{1, \dots, N\}$, compute the weights

$$W_{1,1}^{(i)} = \frac{\tilde{\gamma}_1^1(\tilde{x}_1^{(i)})}{\eta_1(\tilde{x}_1^{(i)})} = g(\tilde{x}_1^{(i)}, y_1)^{\phi_{1,1}}$$

and normalize $\bar{W}_{1,1}^{(i)} = W_{1,1}^{(i)} / \sum_{j=1}^N W_{1,1}^{(j)}$. Calculate the effective sample size (ESS)

$$ESS = \frac{1}{\sum_{i=1}^N (\bar{W}_{1,1}^{(i)})^2}.$$

If $ESS = ESS(\phi_{1,1}) \leq N^*$, resample $\{\tilde{x}_1^{(i)}\}_{i=1}^N$ and set $\bar{W}_{1,1}^{(i)} = \frac{1}{N}$, $i \in \{1, \dots, N\}$.

Set $k = 1$ and denote the particles by $\{x_{k,1}^{(i)}\}$ and the weights by $\{\bar{W}_{k,1}^{(i)}\}$. Set $flag = 0$. Perform the following SMC sampler:

While $flag = 0$, do:

1. Find $\delta \in [0, 1 - \phi_{k,1}]$ so that $ESS(\delta) = N^*$, where $ESS(\delta)$ is as above except now we have the unknown δ instead of $\phi_{1,1}$ in the weights. If $\delta = 1 - \phi_{k,1}$, set $\phi_{k+1,1} = 1$, $K_1 = k + 1$ and $flag = 1$, else set $\phi_{k+1,1} = \phi_{k,1} + \delta$.

2. Compute the weights

$$W_{k+1,1}^{(i)} = W_{k,1}^{(i)} \cdot \frac{\tilde{\gamma}_1^{k+1}(x_{k,1}^{(i)})}{\tilde{\gamma}_1^k(x_{k,1}^{(i)})} = W_{k,1}^{(i)} \cdot g(x_{k,1}^{(i)}, y_1)^\delta,$$

normalize $\bar{W}_{k+1,1}^{(i)} = W_{k+1,1}^{(i)} / \sum_{i=1}^N W_{k+1,1}^{(i)}$. If $ESS \leq N^*$, resample $\{x_{k,1}^{(i)}\}$, set $\bar{W}_{k+1,1}^{(i)} = \frac{1}{N}$. Denote the samples $\{\hat{x}_{k,1}^{(i)}\}$.

3. (MCMC) For $m \in \{1, \dots, S\}$, do:

– For $i \in \{1, \dots, N\}$, sample $x_{k+1,1}^{(i)} | \hat{x}_{k,1}^{(i)}$ from a Markov kernel that preserves

$$\tilde{\pi}_1^{k+1}(x_1) \propto f(x_0, x_1) g(x_1, y_1)^{\phi_{k+1,1}}.$$

E.g., one can use a random walk with proposal

$$x_1'^{(i)} = \hat{x}_{k,1}^{(i)} + Z^{(i)}, \quad Z^{(i)} \sim \mathcal{N}(0, \Sigma_m),$$

for some covariance matrix Σ_m , so that the acceptance probability is

$$a_m^{(i)} = 1 \wedge \frac{\tilde{\pi}_1^{k+1}(x_1'^{(i)})}{\tilde{\pi}_1^{k+1}(\hat{x}_{k,1}^{(i)})} = 1 \wedge \frac{g(x_1'^{(i)}, y_1)^{\phi_{k+1,1}} f(x_0, x_1'^{(i)})}{g(\hat{x}_{k,1}^{(i)}, y_1)^{\phi_{k+1,1}} f(x_0, \hat{x}_{k,1}^{(i)})}.$$

With probability $a_m^{(i)}$, set $\hat{x}_{k,1}^{(i)} = x_1'^{(i)}$, $i \in \{1, \dots, N\}$.

4. Set $k = k + 1$ and $x_{k,1}^{(i)} = \hat{x}_{k-1,1}^{(i)}$, $i \in \{1, \dots, N\}$.

Calculate ESS for the last weights $\{\bar{W}_{K_1,1}^{(i)}\}$. If $ESS \leq N^*$, resample $\{x_{K_1,1}^{(i)}\}$, set $\bar{W}_{K_1,1}^{(i)} = \frac{1}{N}$. Denote the final particles by $\{x_1^{(i)}\}$ and the final weights $\{\bar{W}_1^{(i)}\}$.

• **Iterations** $n = 2, 3, \dots, L$:

We have N paths $\{x_{1:n-1}^{(i)}\}$ with weights $\{\bar{W}_{n-1}^{(i)}\}$. We sample $\tilde{x}_n^{(i)}$ from the proposal distribution $\eta_n(x_{1:n}) = \pi_{n-1}(x_{1:n-1}) f(x_{n-1}, x_n)$, via (9). We will sequentially sample from

$$\tilde{\pi}_n^k(x_{1:n}) \propto \tilde{\gamma}_n^k(x_{1:n}) := \left(\prod_{j=1}^{n-1} g(x_j, y_j) f(x_{j-1}, x_j) \right) f(x_{n-1}, x_n) g(x_n, y_n)^{\phi_{k,n}}$$

starting at $\tilde{\pi}_n^1$ and ending at $\tilde{\pi}_n^{K_n} = \pi_n$ for some $K_n \in \mathbb{N}$. Compute the weights

$$W_{1,n}^{(i)} = \bar{W}_{n-1}^{(i)} \cdot \frac{\tilde{\gamma}_n^1(x_{1:n-1}, \tilde{x}_n^{(i)})}{\eta_n(x_{1:n-1}, \tilde{x}_n^{(i)})} = \bar{W}_{n-1}^{(i)} \cdot g(\tilde{x}_n^{(i)}, y_n)^{\phi_{1,n}},$$

normalize $\bar{W}_{1,n}^{(i)} = W_{1,n}^{(i)} / \sum_{j=1}^N W_{1,n}^{(j)}$. If $ESS \leq N^*$, resample $\{\tilde{x}_{1,n}^{(i)}\}$, set $\bar{W}_{1,n}^{(i)} = \frac{1}{N}$.

Set $k = 1$ and denote the particles by $\{x_{k,n}^{(i)}\}$ and the weights $\{\bar{W}_{k,n}^{(i)}\}$. Set $flag = 0$. Perform the following SMC sampler:

While $flag = 0$, do:

1. Find $\delta \in [0, 1 - \phi_{k,n}]$ so that $ESS(\delta) = N^*$. If $\delta = 1 - \phi_{k,n}$, set $\phi_{k+1,n} = 1$, $K_n = k + 1$ and $flag = 1$, else set $\phi_{k+1,n} = \phi_{k,n} + \delta$.
2. Compute the weights

$$W_{k+1,n}^{(i)} = \bar{W}_{k,n}^{(i)} \cdot \frac{\tilde{\gamma}_n^{k+1}(x_{1:n-1}^{(i)}, x_{k,n}^{(i)})}{\tilde{\gamma}_n^k(x_{1:n-1}^{(i)}, x_{k,n}^{(i)})} = \bar{W}_{k,n}^{(i)} \cdot g(x_{k,n}^{(i)}, y_n)^\delta,$$

normalize $\bar{W}_{k+1,n}^{(i)} = W_{k+1,n}^{(i)} / \sum_{i=1}^N W_{k+1,n}^{(i)}$. If $ESS \leq N^*$, resample $\{x_{k,n}^{(i)}\}$, set $W_{k+1,n}^{(i)} = \frac{1}{N}$. Denote the particles by $\{\hat{x}_{k,n}^{(i)}\}$. Attach these to the samples $\{x_{1:n-1}^{(i)}\}$ from previous time steps and denote the complete paths $\{\hat{x}_{k,1:n}^{(i)}\}$.

3. (MCMC) For $m \in \{1, \dots, S\}$, do:

- For $i \in \{1, \dots, N\}$, sample $x_{k+1,1:n}^{(i)} | \hat{x}_{k,1:n}^{(i)}$ from a Markov kernel that preserves

$$\tilde{\pi}_n^{k+1}(x_{1:n}) \propto \left(\prod_{j=1}^{n-1} g(x_j, y_j) f(x_{j-1}, x_j) \right) f(x_{n-1}, x_n) g(x_n, y_n)^{\phi_{k+1,n}}.$$

Again, one can use a random walk

$$x_{1:n}'^{(i)} = \hat{x}_{k,1:n}^{(i)} + Z^{(i)}, \quad Z^{(i)} \sim \mathcal{N}(0, \Sigma_m)$$

for some covariance matrix Σ_m , so that the acceptance probability is

$$\begin{aligned} \alpha_m^{(i)} &= 1 \wedge \frac{\tilde{\pi}_n^{k+1}(x_{1:n}')^{(i)}}{\tilde{\pi}_n^{k+1}(\hat{x}_{k,1:n}^{(i)})} \\ &= 1 \wedge \frac{\prod_{j=1}^{n-1} g(x_j', y_j) f(x_{j-1}', x_j') f(x_{n-1}', x_n') g(x_n', y_n)^{\phi_{k+1,n}}}{\prod_{j=1}^{n-1} g(\hat{x}_{k,j}^{(i)}, y_j) f(\hat{x}_{k,j-1}^{(i)}, \hat{x}_{k,j}^{(i)}) f(x_{n-1}^{(i)}, \hat{x}_{k,n}^{(i)}) g(\hat{x}_{k,n}^{(i)}, y_n)^{\phi_{k+1,n}}}. \end{aligned}$$

With probability $\alpha_m^{(i)}$, set $\hat{x}_{k,1:n}^{(i)} = x_{1:n}'^{(i)}$.

4. Set $k = k + 1$ and $x_{k,n}^{(i)} = \hat{x}_{k-1,n}^{(i)}$, $i \in \{1, \dots, N\}$.

Calculate ESS for the last weights $\{\bar{W}_{K_n,n}^{(i)}\}$. If $ESS \leq N^*$, resample $\{x_{0:n-1}^{(i)}, x_{K_n,n}^{(i)}\}$ and set $\bar{W}_{K_n,n}^{(i)} = \frac{1}{N}$.

Set $x_{0:n}^{(i)} = \{x_{0:n-1}^{(i)}, x_{K_n,n}^{(i)}\}$ and $\bar{W}_n^{(i)} = \bar{W}_{K_n,n}^{(i)}$, $i \in \{1, \dots, N\}$.

• **Iterations** $n \geq L + 1$:

We have N paths $\{x_{1:n-1}^{(i)}\}$ with weights $\{\bar{W}_{n-1}^{(i)}\}$. We begin by sampling particles $\tilde{x}_n^{(i)}$ from proposal $\eta_n(x_{1:n}) = f(x_{n-1}, x_n) \hat{\pi}_{n-1}(x_{1:n-1})$ (recalling that $\hat{\pi}_L = \pi_L$), $i \in \{1, \dots, N\}$, via (9). Setting $\mu_0(x_1) = f(x_0, x_1)$, we will sample sequentially from the distributions

$$\begin{aligned} \tilde{\pi}_n^k(x_{1:n}) &\propto \left(\hat{\pi}_{n-1}(x_{1:n-1}) f(x_{n-1}, x_n) \right)^{1-\phi_{k,n}} \hat{\pi}_n(x_{1:n})^{\phi_{k,n}} \\ &= \left(\frac{\mu_{n-L}(x_{n-L+1}) g(x_n, y_n)}{f(x_{n-L}, x_{n-L+1})} \right)^{\phi_{k,n}} \left[\prod_{j=0}^{n-L-1} \mu_j(x_{j+1}) \prod_{j=n-L+1}^n f(x_{j-1}, x_j) \prod_{j=1}^{n-1} g(x_j, y_j) \right] \end{aligned}$$

starting at $\tilde{\pi}_n^1$ and ending at $\tilde{\pi}_n^{K_n} = \hat{\pi}_n$ for some $K_n \in \mathbb{N}$. Compute the weights

$$W_{1,n}^{(i)} = \bar{W}_{n-1}^{(i)} \cdot \frac{\tilde{\gamma}_n^1(x_{1:n-1}^{(i)}, \tilde{x}_n^{(i)})}{\eta_n(x_{1:n-1}^{(i)}, \tilde{x}_n^{(i)})} = \bar{W}_{n-1}^{(i)} \left[\frac{g(x_n^{(i)}, y_n) \mu_{n-L}(x_{n-L+1}^{(i)})}{f(x_{n-L}^{(i)}, x_{n-L+1}^{(i)})} \right]^{\phi_{1,n}},$$

normalize $\bar{W}_{1,n}^{(i)} = W_{1,n}^{(i)} / \sum_{i=1}^N W_{1,n}^{(i)}$. If $ESS \leq N^*$, resample $\{\tilde{x}_n^{(i)}\}$ and set $\bar{W}_{1,n}^{(i)} = \frac{1}{N}$.

Set $k = 1$ and denote the particles by $\{x_{k,n}^{(i)}\}$ and the weights $\{\bar{W}_{k,n}^{(i)}\}$. Set $flag = 0$.

Perform the following SMC sampler:

While $flag = 0$, do:

1. Find $\delta \in [0, 1 - \phi_{k,n}]$ so that $ESS(\delta) = N^*$. If $\delta = 1 - \phi_{k,n}$, set $\phi_{k+1,n} = 1$, $K_n = k + 1$ and $flag = 1$, else set $\phi_{k+1,n} = \phi_{k,n} + \delta$.
2. Compute the weights

$$W_{k+1,n}^{(i)} = \bar{W}_{k,n}^{(i)} \cdot \frac{\tilde{\gamma}_n^{k+1}(x_{1:n-1}^{(i)}, x_{k,n}^{(i)})}{\tilde{\gamma}_n^k(x_{1:n-1}^{(i)}, x_{k,n}^{(i)})} = \bar{W}_{k,n}^{(i)} \cdot g(y_n | x_{k,n}^{(i)})^\delta,$$

normalize $\bar{W}_{k+1,n}^{(i)} = W_{k+1,n}^{(i)} / \sum_{i=1}^N W_{k+1,n}^{(i)}$. If $ESS \leq N^*$, resample $\{x_{k,n}^{(i)}\}$, set $\bar{W}_{k+1,n}^{(i)} = \frac{1}{N}$. Denote the resulting samples $\{\hat{x}_{k,n}^{(i)}\}$. Attach these to the samples $x_{n-L:n-1}^{(i)}$ from the previous L time steps and denote the new samples by $\{\hat{x}_{k,n-L:n}^{(i)}\}$.

3. (MCMC) For $m \in \{1, \dots, S\}$:

– For $i \in \{1, \dots, N\}$, sample $x_{k+1,n-L:n}^{(i)} | \hat{x}_{k,n-L:n}^{(i)}$ from a Markov kernel that preserves

$$\begin{aligned} \tilde{\pi}_{k+1,n}(x_{1:n}) &= \left(\frac{g(x_n, y_n) \mu_{n-L}(x_{n-L+1})}{f(x_{n-L}, x_{n-L+1})} \right)^{\phi_{k+1,n}} \times \\ &\quad \left[\prod_{j=0}^{n-L-1} \mu_j(x_{j+1}) \prod_{j=n-L+1}^n f(x_{j-1}, x_j) \prod_{j=1}^{n-1} g(x_j, y_j) \right] \end{aligned}$$

Again, one can use a random walk

$$x_{n-L:n}^{\prime(i)} = \hat{x}_{k,n-L:n}^{(i)} + Z^{(i)}, \quad Z^{(i)} \sim \mathcal{N}(0, \Sigma_m)$$

for some covariance matrix Σ_m , so that the acceptance probability is

$$\begin{aligned} a_m^{(i)} &= 1 \wedge \frac{\tilde{\pi}_{k+1,n}(x_{n-L:n}^{\prime(i)})}{\tilde{\pi}_n^{k+1}(\hat{x}_{k,n-L:n}^{(i)})} \\ &= 1 \wedge \left[\left(\frac{g(x_n^{\prime(i)}, y_n) \mu_{n-L}(x_{n-L+1}^{\prime(i)})}{f(x_{n-L}^{\prime(i)}, x_{n-L+1}^{\prime(i)})} \frac{f(\hat{x}_{k,n-L}^{(i)}, \hat{x}_{k,n-L+1}^{(i)})}{g(\hat{x}_{k,n-L+1}^{(i)}, y_n) \mu_{n-L}(\hat{x}_{k,n-L+1}^{(i)})} \right)^{\phi_{k+1,n}} \times \right. \\ &\quad \left. \frac{\mu_{n-L-1}(x_{n-L}^{\prime(i)}) \prod_{j=n-L}^{n-1} g(x_j^{\prime(i)}, y_j) \prod_{j=n-L+1}^n f(x_{j-1}^{\prime(i)}, x_j^{\prime(i)})}{\mu_{n-L-1}(\hat{x}_{k,n-L}^{(i)}) \prod_{j=n-L}^{n-1} g(\hat{x}_{k,j}^{(i)}, y_j) \prod_{j=n-L+1}^n f(\hat{x}_{k,j-1}^{(i)}, \hat{x}_{k,j}^{(i)})} \right]. \end{aligned}$$

With probability $\alpha_m^{(i)}$, set $\hat{x}_{k,n-L:n}^{(i)} = x_{n-L:n}^{\prime(i)}$.

4. Set $k = k + 1$ and $x_{k,n}^{(i)} = \hat{x}_{k-1,n}^{(i)}$, $i \in \{1, \dots, N\}$.

If $ESS \leq N^*$, resample $\{x_{0:n-1}^{(i)}, x_{K_n,n}^{(i)}\}$ and set $\bar{W}_{K_n,n}^{(i)} = \frac{1}{N}$.

Set $x_{0:n}^{(i)} = \{x_{0:n-1}^{(i)}, x_{K_n,n}^{(i)}\}$ and $\bar{W}_n^{(i)} = \bar{W}_{K_n,n}^{(i)}$, $i \in \{1, \dots, N\}$.

References

- [1] BESKOS, A., CRISAN, D. & JASRA, A. (2014). On the stability of sequential Monte Carlo methods in high dimensions. *Ann. Appl. Probab.*, **24**, 1396–1445.
- [2] BESKOS, A., CRISAN, D., JASRA, A. & WHITELEY, N. (2014). Error bounds and normalizing constants for sequential Monte Carlo samplers in high-dimensions. *Adv. Appl. Probab.*, **46**, 279–306.
- [3] BESKOS, A., CRISAN, D., JASRA, A., KAMATANI, K., & ZHOU, Y. (2017). A stable particle filter for a class of high-dimensional state-space models. *Adv. Appl. Probab.* **49**, 24–48.
- [4] BISHOP, C. H., ETHERTON, B. J. & MAJUMDAR S. J. (2001). Adaptive sampling with the ensemble transform Kalman filter. *part i: Theoretical aspects. Monthly Weather Review*, **129**.
- [5] CAPPÉ, O., RYDEN, T., & MOULINES, É. (2005). *Inference in Hidden Markov Models*. Springer: New York.
- [6] CHATTERJEE, S., & DIACONIS, P. (2018). The sample size required in importance sampling. *Ann. Appl. Probab.* **28**, 1099–1135.
- [7] COTTER, C., CRISAN, D., HOLM, D., PAN, W., & SHEVCHENKO, I. (2020). A Particle Filter for Stochastic Advection by Lie Transport: A Case Study for the Damped and Forced Incompressible Two-Dimensional Euler Equation. *SIAM/ASA J. Uncert. Quant.*, **8**, 1446–1492.
- [8] CRISAN, D. & ROZOVSKII, B. (2011). *The Oxford Handbook of Nonlinear Filtering*. Oxford: OUP.
- [9] DEL MORAL, P. (2004). *Feynman-Kac Formulae*. Springer.
- [10] DEL MORAL, P., DOUCET, A. & JASRA, A. (2006). Sequential Monte Carlo samplers. *J. R. Statist. Soc. B*, **68**, 411–436.
- [11] DOUCET, A., BRIERS, M., & SENECAI, S. (2006). Efficient block sampling strategies for sequential Monte Carlo methods. *Journal of Computational and Graphical Statistics*, **15**(3), 693-711.
- [12] EVENSEN, G. (1994). Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, **99**(C5):10,143–10,162.
- [13] EVENSEN, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, **53**:343–367.
- [14] GODSILL, S. & CLAPP, T. (2001). Improvement strategies for Monte Carlo particle filters. In *Sequential Monte Carlo Methods in Practice* (A. Doucet, N. de Freitas and N. Gordon, eds.) 139–158. New York: Springer.
- [15] JACOB, P. E., MURRAY, L. M., & RUBENTHALER, S. (2015). Path storage in the particle filter. *Statistics and Computing*, **25**(2), 487-496.

- [16] KANTAS, N., BESKOS, A., & JASRA, A. (2014). Sequential Monte Carlo for inverse problems: a case study for the Navier Stokes equation. *SIAM/ASA JUQ*, **2**, 464–489.
- [17] KANTAS, N., DOUCET, A., SINGH, S. S., MACIEJOWSKI, J., & CHOPIN, N. (2015). On particles methods for parameter estimation in general state-space models. *Stat. Sci.*, **30**, 328–351.
- [18] KITAGAWA, G. & SATO, S. (2001). Monte Carlo smoothing and self-organising state-space model. In *Sequential Monte Carlo Methods in Practice* (A. Doucet, N. de Freitas and N. Gordon, eds.) 178–195. New York: Springer.
- [19] KLAAS, M., DE FREITAS, N., & DOUCET, A. (2005). Toward practical N² Monte Carlo: the marginal particle filter. In *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*, 308–315.
- [20] POLSON, N. G., STROUD, J. R., & MULLER, P. (2008). Practical filtering with sequential parameter learning. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **70**(2), 413–428.
- [21] PONS LLOPIS, F., KANTAS, N., BESKOS, A. & JASRA, A. (2018). Particle filtering for stochastic Navier-Stokes signals observed with linear additive noise. *SIAM J. Sci. Comp.* **40**, A1544–A1565.
- [22] NERGER, L., JANJIĆ, T., SCHRÖTER, J., & HILLER, W. (2012). A Unification of Ensemble Square Root Kalman Filters, *Mon. Wea. Rev.*, **140**(7), 2335–2345.
- [23] OLSSON, J., CAPPÉ, O, DOUC, R. & MOULINES, E. (2008). Sequential Monte Carlo smoothing with application to parameter estimation in nonlinear state space models. *Bernoulli*, **14**, 155–179.
- [24] REBESCHINI, P. & VAN HANDEL, R. (2015). Can local particle filters beat the curse of dimensionality? *Ann. Appl. Probab.* **25**, 2809–2866.
- [25] SAKOV, P., & OKE, P. R. (2008): Implications of the form of the ensemble transformation in the ensemble square root filters. *Mon. Wea. Rev.*, **136**, 1042–1053
- [26] SNYDER, C., BENGTSSON, T., BICKEL, P., & ANDERSON, J. (2008). Obstacles to high-dimensional particle filtering. *Mon. Wea. Rev.*, **136**(12), 4629–4640.
- [27] LORENZ, E. (1996). Predictability - A problem partly solved. *Seminar on Predictability*, Vol. I, ECMWF.
- [28] VREUGDENHIL, C.B. (1986). Numerical Methods for Shallow-Water Flow. *Water Science and Technology Library*. **13**. Springer, Dordrecht.