

Reducing the Bias in Blocked Particle Filtering for High-Dimensional Systems *

Francesco Bertoli[†]

Adrian N. Bishop[‡]

Abstract

Particle filtering is a powerful approximation method that applies to state estimation in nonlinear and non-Gaussian dynamical state-space models. Unfortunately, the approximation error depends exponentially on the system dimension. This means that an incredibly large number of particles may be needed to appropriately control the error in very large scale filtering problems. The computational burden required is often prohibitive in practice. Rebeschini and Van Handel (2013) analyse a new approach for particle filtering in large-scale dynamic random fields. Through a suitable localisation operation they reduce the dependence of the error to the size of local sets, each of which may be considerably smaller than the dimension of the original system. The drawback is that this localisation operation introduces a bias. In this work, we propose a modified version of Rebeschini and Van Handel's blocked particle filter. We introduce a new degree of freedom allowing us to reduce the bias. We do this by enlarging the space during the update phase and thus reducing the amount of dependent information thrown away due to localisation. By designing an appropriate tradeoff between the various tuning parameters it is possible to reduce the total error bound via allowing a temporary enlargement of the update operator without really increasing the overall computational burden.

1 Introduction

Recursive Bayesian estimation (or filtering) is a technique for recursively estimating the state of a random process observed via noisy measurements. If the underlying dynamical model is linear and Gaussian we have the celebrated Kalman filter [12] which is an exact solution to the Bayesian filtering problem. Unfortunately, in many practical scenarios of interest, the Bayes filter is not exactly computable. Therefore, we seek techniques to approximate this ideal filter. The Kalman filter can be applied in more general settings [9, 11] as an approximation. Particle filtering is a more general approximation method that is easily applied to nonlinear and non-Gaussian state-space models. The particle filter approximates the Bayesian filter via Monte Carlo simulation/sampling. The samples (or particles) are propagated through a sequential importance sampling mechanism that attempts to capture the dynamics of the unobservable process and the likelihood of the observations available. Other approximations exist such as Gaussian mixture filters etc. [9, 1].

The particle filter has been widely studied in theory and in countless practical applications [8, 13, 7]. In [6] the authors prove that the error can be controlled uniformly in time, thus providing a solid mathematical support for application of the filter in numerous fields. Unfortunately, the particle filter computation is strongly dependent on the dimension of the underlying estimation problem. Specifically, the error bound grows exponentially with the system's dimension, making the filter infeasible in most high-dimensional applications. This problem is known as the *curse of dimensionality* [14]. A heuristic explanation of this phenomenon for a particular case can be found in [17]. In [5, 14] the authors give a precise relation between the dimension of the system and the number of particle required to avoid weight degeneracy [7]. The fact that the approximation error is exponential in the dimension and only inversely controlled by the sample size implies that an incredibly large number of particles are required when dealing with a high dimensional system if we want to control the error at a reasonable level. Obviously, a large number of particles means a heavy computational burden, that is often simply prohibitive.

Recent studies [3, 4, 15, 16] however suggest that high-dimensional particle filtering may be feasible in particular applications and/or if one is willing to accept a degree of systematic bias. In [3], the particle filter is applied in a static setting where the objective is to sample from some high-dimensional target distribution. In this case, through a sequence of intermediate and simpler distributions, it is shown that the particle filter

*This work was supported by AFOSR/AOARD via AOARD-144042.

[†]F Bertoli is with the Australian National University and NICTA. He is supported by NICTA.

[‡]A.N. Bishop is with the Australian National University and NICTA. He is supported by NICTA and the Australian Research Council (ARC) via a Discovery Early Career Researcher Award (DE-120102873).

will converge to a sampled representation of the target distribution with a typical Monte Carlo error (inverse in the number of particles) given a complexity on the order of the dimension squared. Although [3] deals only, in essence, with a static problem of sampling from a fixed target distribution, the analysis introduces a novel way of thinking about high-dimensional particle filtering which may carry over to dynamic filtering problems. Related work appears in [4].

1.1 Background: The Motivating Paper

In [15] the authors consider particle filtering in large-scale dynamic random fields. They assume the dynamics of the underlying process are localised to a neighbourhood of the field and the observations are local to each site. They exploit this idea by localising the algorithm during the update phase. They argue that the difficulty in high dimensional particle filtering is due largely to the dimension of the observation and the nonlinearity of the update operation. Therefore, they partition the field into independent blocks and correct every marginalised block separately. The posterior is simply the product of the blocked marginals. The real contribution of [15] is a descriptive and technical analysis that shows the error introduced due to the localisation procedure can be readily controlled if the dynamics of the random field at each site are only locally dependent on those sites within close proximity. The standard sampling approximation error is shown to be exponential in only the size of the individual blocks. The number of samples/particles controls the sampling approximation error at the typical rate while the error due to the localisation process is a systematic bias that can only be controlled through an increase in the block size. Since each block is updated independently, parallel implementation is readily applicable and the computational burden may be alleviated, albeit this remains to be seen in practice. While the results of [15] are at the proof-of-concept stage, the idea is incredibly powerful.

The authors in [15] show that although the total approximation error can be controlled uniformly in time, it suffers from a spatial inhomogeneity. Specifically, the nodes close to the block boundaries display a larger error than those far removed from the boundaries (as one might expect). A simple approach to average this spatial inhomogeneity is given in [2] where adaptive partitioning of the field is employed.

1.2 Contribution

In this paper we consider again the idea proposed in [15] and propose a modified particle filtering algorithm that displays an additional degree of freedom. The idea proposed herein is to enlarge the blocks during the update phase, allowing for more observations to be employed during the correction at each block. The main contribution is the addition of a new parameter that captures how much we enlarge each block prior to the update. Obviously, by enlarging each block prior to updating we reduce the bias error but we increase the complexity involved in updating each (enlarged) block. By designing an appropriate tradeoff between the various tuning parameters it is possible to reduce the total error bound via allowing a temporary enlargement of the update operator without increasing the overall computational burden.

2 Problem Setup and Applications of the Blocked Filter

We borrow the problem setup and notation directly from [15].

Consider a Markov chain $(X_n)_{n \geq 0}$ defined on a Polish state space \mathbb{X} with transition density $p : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$ with respect to a reference measure ψ . Moreover consider a process $(Y_n)_{n \geq 0}$, defined on a Polish space \mathbb{Y} , conditionally independent given $(X_n)_{n \geq 0}$, with a transition density $g : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}$ with respect to a measure φ . The process $(X_n)_{n \geq 0}$ is observed via the process $(Y_n)_{n \geq 0}$. Our aim is to estimate the probability of the state X_n given the measurements up to that time and the initial condition μ . Therefore we introduce the filter

$$\pi_n^\mu := \mathbf{P}_\mu[X_n \in \cdot | Y_1, \dots, Y_n]$$

It can be easily seen, using Bayes rule, that the filter can be written in a recursive way

$$\pi_0^\mu = \mu, \quad \pi_n^\mu = F_n \pi_{n-1}^\mu$$

where the operator F_n is defined as follows

$$(F_n \rho)(A) := \frac{\int \mathbf{1}_A(x) p(x_0, x) g(x, Y_n) \psi(dx) \rho(dx_0)}{\int p(x_0, x) g(x, Y_n) \psi(dx) \rho(dx_0)}$$

Moreover, the above operator is typically split into two sub-steps $F_n = C_n P$ where

$$(P\rho)(A) := \int \mathbf{1}_A(x) p(x_0, x) \psi(dx) \rho(dx_0)$$

is a prediction step, and

$$(C_n \rho)(A) := \frac{\int \mathbf{1}_A(x) g(x, Y_n) \rho(dx)}{\int g(x, Y_n) \rho(dx)}$$

is a correction (or update) step. In the prediction step, the measure is transformed according to the density $p(\cdot, \cdot)$, while in the update step we use the new information Y_n to correct the predicted measure. We then write the recursion as follows

$$\pi_{n-1}^\mu \xrightarrow{\text{prediction}} \pi_{n-}^\mu := P \pi_{n-1}^\mu \xrightarrow{\text{correction}} \pi_n^\mu = C_n \pi_{n-}^\mu$$

The classic bootstrap particle filter uses N particles (or samples) to approximate the measure π_n^μ . Given a sampled approximation of π_{n-1}^μ , the particles are first moved according to the transition $p(\cdot, \cdot)$ in order to approximate a sampled representation of the prediction. The update then computes a weighted posterior empirical measure via $g(\cdot, \cdot)$. Eventually, a resample step is added in order to avoid weight degeneracy [7]. More formally, denoting the bootstrap filter by $\hat{\pi}_n^\mu$, we have $\hat{\pi}_n^\mu = \hat{F}_n \hat{\pi}_{n-1}^\mu$ where $\hat{F}_n = C_n S^N P$ and S^N represents the sampling operator here defined

$$S^N \rho := \frac{1}{N} \sum_{i=1}^N \delta_{x(i)}, \quad x(i) \text{ is i.i.d. } \sim \rho$$

It is possible to prove that

$$\sup_{|f| \leq 1} \mathbf{E}[\pi_n^\mu(f) - \hat{\pi}_n^\mu(f)] \leq a_0 / \sqrt{N}$$

with a_0 independent of time. Unfortunately, the constant c typically depends (exponentially) on the dimension of the underlying problem. Intuition for this exponential dependence is given in [15, 17].

We now consider the pair (X_n, Y_n) as a random field $(X_n^v, Y_n^v)_{v \in V}$ indexed on a finite undirected graph $G = (V, E)$. The vertex set V will represent the collection of sites and the edge set E the spatial relationships between them. The cardinality of V captures, in some sense, the dimension of interest. More formally, the spaces \mathbb{X} and \mathbb{Y} are defined as products $\mathbb{X} := \prod_{v \in V} \mathbb{X}^v$, $\mathbb{Y} := \prod_{v \in V} \mathbb{Y}^v$. The reference measures are products $\psi := \otimes_{v \in V} \psi^v$, $\varphi := \otimes_{v \in V} \varphi^v$, where ψ^v and φ^v are reference measures on \mathbb{X}^v and \mathbb{Y}^v respectively. The transition densities are defined as

$$p(x, z) := \prod_{v \in V} p^v(x, z^v), \quad g(x, y) := \prod_{v \in V} g^v(x^v, y^v)$$

where $p^v : \mathbb{X} \times \mathbb{X}^v \rightarrow \mathbb{R}$ and $g^v : \mathbb{X}^v \times \mathbb{Y}^v \rightarrow \mathbb{R}$ are densities with respect to the reference measures ψ^v and φ^v . From the definition we can see that the observations Y_n are assumed to be completely local, in the sense that Y_n^v depends uniquely on the value assumed by X_n^v . The process $(X_n)_{n \geq 0}$ is local in the sense that the state at a site v depends only on the state at nearby sites. We state this formally. Consider the graph G equipped with the distance $d(v, v')$ defined by the number of hops along the shortest path connecting v and v' . We can define the neighbourhood of a site v as

$$N(v) := \{v' \in V : d(v, v') \leq r\}$$

where r represents the range of interaction. Then we assume

$$p^v(x_1, z^v) = p^v(x_2, z^v) \quad \text{whenever} \quad x_1^{N(v)} = x_2^{N(v)}$$

where we write for $I \subseteq V$, $x^I = (x^i)_{i \in I}$. In other words, the random field $(X_n)_{n \geq 0}$ is local in the sense that given X_0, \dots, X_{n-1} the present state X_n^v depends only on $X_{n-1}^{N(v)}$.

2.1 Blocked Particle Filter

In [15] the authors propose an application of the blocked filter algorithm to the field model just explained, exploiting the local dynamic dependencies. We briefly illustrate this algorithm. Consider a partition $\mathcal{K} = \{K_i\}_i$ of V into non-overlapping blocks with a union equal to V . The idea is to create independence across blocks on V by marginalising after the prediction step. We then update each block separately and finally we form π_n^μ via the product of the independent (updated) blocked marginals. More formally, consider the block operator B on the space $\mathcal{M}(\mathbb{X})$ of measures on \mathbb{X} , defined by

$$B\rho := \bigotimes_{K \in \mathcal{K}} B^K \rho$$

where $B^K \rho$ is the marginal of the measure ρ on the subset $K \subseteq V$. Then the proposed block filter can be written as a recursion $\hat{\pi}_0^\mu = \mu$, $\hat{\pi}_n^\mu = \hat{F}_n \hat{\pi}_{n-1}^\mu$ where the operator $\hat{F}_n = C_n B S^N P$ consists of four steps

$$\hat{\pi}_{n-1}^\mu \xrightarrow{\text{prediction/sampling}} \hat{\pi}_{n-}^\mu = S^N P \hat{\pi}_{n-1}^\mu \xrightarrow{\text{blocking/correction}} \hat{\pi}_n^\mu = C_n B \hat{\pi}_{n-}^\mu$$

We make the following definition.

Definition 1. Given $\mu, \nu \in \mathcal{M}(\mathbb{X})$ and a subset $I \subseteq V$ we define a distance of the marginals on I as follows

$$\|\mu - \nu\|_I := \sup_{f \in M(\mathbb{X}^I): |f| \leq 1} \mathbf{E}[|\mu(f) - \nu(f)|^2]^{\frac{1}{2}}$$

where the expectation is taken with respect to the random sampling and $M(\mathbb{X}^I)$ is the class of measurable function on \mathbb{X} that depends only on the values on I , that is $f(x) = f(y)$ when $x^I = y^I$. If $I = V$ we omit the subscript and write $\|\mu - \nu\|$.

With no expectation it follows that $\|\cdot\|$ is equivalent to the total variation which we write as $\|\cdot\|$. The two norms are interchangeable when no sampling occurs.

Now, given a set $I \subseteq V$ we define the boundary and the interior

$$\partial I := \{v \in I \mid N(v) \not\subseteq I\}, \quad \text{int}(I) := I \setminus \partial I$$

and given a partition \mathcal{K} , we define the following quantities

$$\begin{aligned} \Delta &:= \max_{v \in V} |N(v)| \\ |\mathcal{K}|_\infty &:= \max_{K \in \mathcal{K}} |K| \\ \Delta_{\mathcal{K}} &:= \max_{K \in \mathcal{K}} |\{K' \in \mathcal{K} : d(K, K') \leq r\}| \end{aligned}$$

where the first quantity is independent of the partition. The result proven in [15] is the following.

Theorem 1 (Blocked Particle Filter [15]). *There exists a constant $0 < \varepsilon_0 < 1$, depending only on the quantities $\Delta, \Delta_{\mathcal{K}}$ such that if there exists $\varepsilon_0 < \varepsilon < 1$ and $0 < \kappa < 1$ such that*

$$\varepsilon < p^v(x, z^v) < \varepsilon^{-1}, \quad \kappa < g^v(x^v, y^v) < \kappa^{-1} \quad \forall x, z \in \mathbb{X}, y \in \mathbb{Y}, v \in V$$

then for every $x \in \mathbb{X}$, $n \geq 0$, $K \in \mathcal{K}$ and $I \subseteq K$ we have

$$\|\pi_n^\mu - \hat{\pi}_n^\mu\|_I \leq \alpha \left[e^{-\beta_1 d(I, \partial K)} + \frac{e^{\beta_2 |\mathcal{K}|_\infty}}{\sqrt{N}} \right]$$

where the constants α, β_1, β_2 are positive, finite and dependent only on $\Delta, \Delta_{\mathcal{K}}, \varepsilon, \kappa$ and r .

The intuition is that the algorithm approximation error is exponential in $|K|$ rather than in $|V|$ but that the error at some individual locations increases with the proximity of those locations to the border of the blocks. This leads to a spatial inhomogeneity as seen in the first term of the bound.

2.2 Adaptively Blocked Particle Filter

A first attempt to achieve a spatially homogeneous error bound can be found in [2]. The idea is to consider a finite number m of partitions \mathcal{K}_i and to apply them cyclically. Clearly we have to choose the partitions in such a way there is no node that is consistently close to a border. This condition is expressed by a bound on the average, or exponential average, of the border distance. Given $\beta > 0$ write

$$\theta \leq \theta_m(v) = \frac{1}{m} \sum_{j=0}^{m-1} d(v, \partial K_j(v)) \quad 0 < \phi_m(v) = \frac{1}{m} \sum_{j=0}^{m-1} e^{-\beta d(v, \partial K_j(v))} \leq \phi$$

Clearly θ and ϕ represent how well balanced the collection of partitions are. Define $\Delta_d(v) := \max_s d(v, \partial K_s)$ and $\nabla_d(v) := \min_s d(v, \partial K_s)$.

Theorem 2 ([2]). *There exists a constant $0 < \varepsilon_0 < 1$, depending only on the quantities $\Delta, \Delta_{\mathcal{K}}$ such that if there exists $\varepsilon_0 < \varepsilon < 1$ and $0 < \kappa < 1$ such that*

$$\varepsilon < p^v(x, z^v) < \varepsilon^{-1}, \quad \kappa < g^v(x^v, y^v) < \kappa^{-1} \quad \forall x, z \in \mathbb{X}, y \in \mathbb{Y}, v \in V$$

then for every $x \in \mathbb{X}$, $n \geq 0$ and $v \in V$ we have

$$\begin{aligned} \frac{1}{m} \sum_{k=0}^{m-1} \|\pi_{n-k}^\mu - \hat{\pi}_{n-k}^\mu\|_v &\leq \alpha \left(\phi_m(v) + \frac{|\mathcal{K}|_\infty e^{\beta |\mathcal{K}|_\infty}}{\sqrt{N}} \right) \\ &\leq \alpha \left(e^{-\beta e^{-\beta(\Delta_d(v) - \nabla_d(v))} \frac{1}{m} \theta_m(v)} + \frac{|\mathcal{K}|_\infty e^{\beta |\mathcal{K}|_\infty}}{\sqrt{N}} \right) \end{aligned}$$

where $0 < \alpha, \beta < \infty$ depend only on $\varepsilon, \kappa, r, \Delta$ and $|\mathcal{K}|_\infty := \max_s \max_{K \in \mathcal{K}_s} |K|$ in this case.

If $\theta = \theta_m(\nu) = \frac{1}{m} \sum_{j=0}^{m-1} d(\nu, \partial K_j(\nu))$ where $K_j(\nu) \in \mathcal{K}_j$ for all $\nu \in \mathcal{V}$, then the bound is completely spatially invariant. See [2] for further discussion on this method.

3 Enlarged Blocked Particle Filtering

Suppose now we are given a partition \mathcal{K} over V but it turns out we are interested only in estimating the marginal of π_n^μ on a particular block $K \in \mathcal{K}$. We could first redefine the partition with a larger block encompassing K and a bunch of single site blocks (to speed up the overall computation). It is of course not possible to define a partition in this manner for multiple blocks of interest. However, the idea proposed here is based on extending the state space by creating multiple independent copies of the measurements (and states) that are then used in different (and independent) enlarged blocks.

We introduce some new notation. Consider a parameter $b \geq 0$, that we will consider fixed throughout the rest of the paper. Then define, for any $K \in \mathcal{K}$, an enlarged block

$$\bar{K} := \{\nu \in V \mid d(\nu, K) \leq b\}$$

Now define the enlarged spaces

$$\mathbb{X}^E := \prod_{K \in \mathcal{K}} \prod_{\nu \in \bar{K}} \mathbb{X}^\nu, \quad \mathbb{Y}^E := \prod_{K \in \mathcal{K}} \prod_{\nu \in \bar{K}} \mathbb{Y}^\nu$$

Consider the collection $\bar{\mathcal{K}} = \{\bar{K} : K \in \mathcal{K}\}$. This is no longer a partition of V . However, $\bar{\mathcal{K}}$ is a partition on \mathbb{X}^E , and here we can apply the blocking and updating operators associated with $\bar{\mathcal{K}}$. We use the superscript E to note enlarged objects. The measures ψ^E and φ^E are defined straightforwardly. The block operator becomes

$$\mathbb{B}^E : \mathcal{M}(\mathbb{X}) \rightarrow \mathcal{M}(\mathbb{X}^E), \quad \mathbb{B}^E(\rho) := \bigotimes_{K \in \mathcal{K}} \mathbb{B}^{\bar{K}} \rho$$

To update, we need the same operator C_n redefined on the new space $\mathcal{M}(\mathbb{X}^E)$,

$$(C_n^E \rho)(A) := \frac{\int \mathbf{1}_A(x) \prod_{\nu \in \mathbb{X}^E} g^\nu(x^\nu, Y_n^\nu) \rho(dx)}{\int \prod_{\nu \in \mathbb{X}^E} g^\nu(x^\nu, Y_n^\nu) \rho(dx)}$$

We also define

$$\mathbb{B}^{-1} : \mathcal{M}(\mathbb{X}^E) \rightarrow \mathcal{M}(\mathbb{X}), \quad \mathbb{B}^{-1}(\rho) := \bigotimes_{K \in \mathcal{K}} \mathbb{B}^K \rho$$

Now we can write the enlarged blocked filter algorithm as a recursion

$$\hat{\pi}_0^\mu = \mu, \quad \hat{\pi}_n^\mu = \hat{F}_n^E \hat{\pi}_{n-1}^\mu \quad (n \geq 1)$$

where $\hat{F}_n^E := \mathbb{B}^{-1} C_n^E \mathbb{B}^E S^N P$. Now we have five steps. Skipping the prediction/sample steps, graphically we have

$$\hat{\pi}_{n-1}^\mu = S^N P \hat{\pi}_{n-1}^\mu \xrightarrow{\text{enlarging/blocking}} \mathbb{B}^E \hat{\pi}_{n-1}^\mu \xrightarrow{\text{updating}} C_n^E \mathbb{B}^E \hat{\pi}_{n-1}^\mu \xrightarrow{\text{marginalizing}} \hat{\pi}_n^\mu$$

To write out the explicit expression of the filter we note that

$$(\mathbb{B}^{-1} C_n^E \mathbb{B}^E P \nu)(A) = (C_n^E \mathbb{B}^E P \nu)(A^E)$$

where $A^E := A \times (\mathbb{X}^E \setminus \mathbb{X})$. Therefore, splitting a variable $z \in \mathbb{X}^E$ in $z = (x, z_E)$ with $x \in \mathbb{X}$ and $z_E \in \mathbb{X}^E \setminus \mathbb{X}$ (where now we put E as subscript just for notational simplicity) and an enlarged block $\bar{K} = (K, K^E)$ where $K^E = \{\nu \in \bar{K} : \nu \notin K\}$, we can write

$$\begin{aligned} (\hat{F}_s^E \nu)(A) &= \frac{\int \mathbf{1}_{A^E}(z) \prod_{K' \in \mathcal{K}} \left[\prod_{w \in \bar{K}'} p^w(x_0, z^w) g^w(z^w, Y_s^w) \nu(dx_0) \psi^{\bar{K}'}(dz^{\bar{K}'}) \right]}{\int \prod_{K' \in \mathcal{K}} \left[\prod_{w \in \bar{K}'} p^w(x_0, x^w) g^w(x^w, Y_s^w) \nu(dx_0) \psi^{\bar{K}'}(dz^{\bar{K}'}) \right]} \\ &= \frac{\int \mathbf{1}_A(x) \prod_{K' \in \mathcal{K}} \left[\prod_{w \in K'} p^w(x_0, x^w) g^w(x^w, Y_s^w) \prod_{w \in K'^E} p^w(x_0, z_E^w) g^w(z_E^w, Y_s^w) \nu(dx_0) \psi^{K'}(dx^{K'}) \psi^{K'^E}(dz_E^{K'^E}) \right]}{\int \prod_{K' \in \mathcal{K}} \left[\prod_{w \in K'} p^w(x_0, x^w) g^w(x^w, Y_s^w) \prod_{w \in K'^E} p^w(x_0, z_E^w) g^w(z_E^w, Y_s^w) \nu(dx_0) \psi^{K'}(dx^{K'}) \psi^{K'^E}(dz_E^{K'^E}) \right]} \end{aligned}$$

where for $J \subseteq V$ we write $\psi^J(dx^J) = \prod_{\nu \in J} \psi^\nu(dx^\nu)$.

4 Main Results and Discussion

Define an ideal enlarged blocked filter $\tilde{\pi}_n^\mu = \tilde{F}_n \dots \tilde{F}_1 \mu$ where $\tilde{F}_s := B^{-1} C_s^E B^E P$. Fix $I \subseteq V$. We then use the triangle inequality to decompose the error according to

$$\|\pi_n^\mu - \hat{\pi}_n^\mu\|_I \leq \|\pi_n^\mu - \tilde{\pi}_n^\mu\|_I + \|\tilde{\pi}_n^\mu - \hat{\pi}_n^\mu\|_I$$

where we refer to the first and second decomposed terms as the bias and variance respectively. The bias represents the error introduced solely as a result of the blocking operation. In the standard bootstrap filter, this bias term vanishes and the typical analysis considers only the variance term.

Going forward, we consider bounding both the bias and the variance. We stress however, that the bias is fundamentally more interesting as it pertains directly to the localisation idea considered herein. Indeed, the sampling operation that leads to the variance term could be replaced with other approximation techniques with no loss of generality (albeit a different approximation error than detailed subsequently).

For sake of completeness/clarity we firstly state a result that includes both a bias and a variance bound.

Theorem 3 (Main result). *Suppose there exists a constant $0 < \varepsilon_0 < 1$, depending only on Δ and $\Delta_{\mathcal{K}}$ and assume*

$$\varepsilon < p^v(x, z^v) < \varepsilon^{-1}, \quad \kappa < g^v(x^v, y^v) < \kappa^{-1} \quad \forall x, z \in \mathbb{X}, y \in \mathbb{Y}, v \in V$$

Then for every time $n \geq 0$, $x \in \mathbb{X}$, $K \in \mathcal{K}$ and $I \subseteq K$ we have

$$\|\pi_n^\mu - \hat{\pi}_n^\mu\|_I \leq \alpha \left[e^{-\beta_1 d(I, \delta \bar{K})} + \frac{e^{\beta_2 |\overline{\mathcal{K}}|_\infty}}{\sqrt{N}} \right]$$

where the constants $0 < \alpha, \beta_1, \beta_2 < \infty$ depend only on $\varepsilon, \kappa, r, \Delta, \Delta_{\mathcal{K}}, \Delta_{\overline{\mathcal{K}}}$.

This single (total error) bound is derived in practice as two separate bounds which we now explicitly state.

Theorem 4 (Bounding the Bias). *Assume there exists $0 < \varepsilon < 1$ such that*

$$\varepsilon < p^v(x, z^v) < \varepsilon^{-1} \quad \text{for all } v \in V, x, z \in \mathbb{X}$$

and such that

$$\varepsilon > \varepsilon_0 = \left(1 - \frac{1}{18\Delta^2} \right)^{1/2\Delta}$$

Let $\beta = -(2r)^{-1} \log 18\Delta^2(1 - \varepsilon^{2\Delta}) > 0$. Then for every $n \geq 0$ we have

$$\|\pi_n^\mu - \tilde{\pi}_n^\mu\|_I \leq \frac{8e^{-\beta}}{1 - e^{-\beta}} (1 - \varepsilon^{2\Delta}) |I| e^{-\beta d(I, \delta \bar{K})}$$

for every $x \in \mathbb{X}$, $K \in \mathcal{K}$ and $I \subseteq K$.

The only difference between this bias bound and the bias bound in [15] is the presence of $|\overline{\mathcal{K}}|_\infty$ in place of $|\mathcal{K}|_\infty$. For a given partition \mathcal{K} any enlargement of the blocks in \mathcal{K} yielding $\overline{\mathcal{K}}$ results in a tighter bias bound as expected.

Theorem 5 (Bounding the Variance). *Assume there exists $0 < \kappa, \varepsilon < 1$ such that*

$$\varepsilon < p^v(x, z^v) < \varepsilon^{-1}, \quad \kappa < g^v(x^v, y^v) < \kappa^{-1} \quad \text{for all } v \in V, x, z \in \mathbb{X}, y \in \mathbb{Y}$$

and such that

$$\varepsilon > \varepsilon_0 = \left(1 - \frac{1}{6\Delta_{\overline{\mathcal{K}}}\Delta^2} \right)^{1/2\Delta}$$

Let $\beta = -\log 6\Delta_{\overline{\mathcal{K}}}\Delta^2(1 - \varepsilon^{2\Delta}) > 0$ where $\Delta_{\overline{\mathcal{K}}} := \max\{|K \in \mathcal{K} : d(K, \bar{K}) \leq r\}|$. Then for every $n \geq 0$ we have

$$\|\tilde{\pi}_n^\mu - \hat{\pi}_n^\mu\|_I \leq |I| \frac{64\Delta_{\overline{\mathcal{K}}}}{1 - e^{-\beta}} \frac{\varepsilon^{-4|\overline{\mathcal{K}}|_\infty} \kappa^{-4|\overline{\mathcal{K}}|_\infty \Delta_{\mathcal{K}}}}{\sqrt{N}}$$

for every $x \in \mathbb{X}$, $K \in \mathcal{K}$ and $I \subseteq K$.

Again, the only significant difference between this variance bound and the variance bound in [15] is the presence of $|\overline{\mathcal{K}}|_\infty$ in place of $|\mathcal{K}|_\infty$. The variance depends inversely on the number of samples and exponentially in the size of the enlarged blocks.

4.1 How to Use the Enlarged Blocked Filter

Roughly, we now explain how one may implement the enlarged blocked filter to reduce the bias as compared with the algorithm proposed in [15] while maintaining a comparable variance and computational complexity.

Suppose firstly that one has a random field over $|V|$ sites and the computational power available (defining a bound on N) ensures that blocks of size $|V|/k$ can be readily handled for some $k > 0$. Then the complexity of the blocked particle filter proposed in [15] can, in a sense, be regarded as being of order $\mathcal{O}(kN)$. Really, one can imagine k particle filters running in parallel over each block and each with complexity on the order of $\mathcal{O}(N)$.

To exploit the enlarged blocked particle filter, one should start with a larger number $c > k$ of smaller blocks which when enlarged are mostly of the size $|V|/k$. Then, the complexity of the enlarged blocked particle filter proposed herein is on the order $\mathcal{O}(cN)$. One immediately sees that the variance of the enlarged blocked particle filter is mostly on the same order as that of the algorithm proposed in [15] and the computational complexity has only increased linearly. However, in almost all cases (and certainly with well-designed partitions) one will achieve a reduction in the bias at any given site in the random field.

4.2 Spatial Homogeneity

We consider a special but interesting case in which a spatial homogeneous total error bound is obtained, the bias bound is better (tighter) than in [15], and the computational requirements largely unchanged when compared with the algorithm in [15].

Corollary 1. *Assume the same hypothesis of Theorem 4. Consider the partition $\mathcal{K} = \{v\}_{v \in V}$ and suppose $b > r$. Then for every $n \geq 0$, $x \in \mathbb{X}$, and $v \in V$ we have*

$$\|\pi_n^\mu - \tilde{\pi}_n^\mu\|_v \leq \frac{8e^{-\beta}}{1 - e^{-\beta}} (1 - \epsilon^{2\Delta}) e^{-\beta(b-r)}$$

This bound is spatially homogeneous and with $b > r$ it is strictly less than the bias bound introduced in [15]. Note that while the bias bound here is spatially homogeneous, the actual bias may still be inhomogeneous since this result is potentially based on over bounding. On the other hand, it is possible to apply the adaptive scheme proposed in [2] with the enlarged blocked filter and potentially achieve true spatial homogeneity.

4.3 Discussion on the Enlarged Blocked Filter

The idea of the enlarged blocked particle filter is essentially based on the principle that larger blocks lead to a reduction in the bias introduced due to blocking.

So, why not just start with larger blocks?

- Well, irrespective of the size of the blocks, if one applies the standard blocked particle filter of [15] then there will always exist sites on the border of a block.
- If we extend (or enlarge) the blocks as proposed herein, we (typically) reduce the bias at each site (and particularly those sites that were on the border of a block in the original partition).
- If we increase the number of samples N with a fixed number of larger blocks (in the original partition) then while we can reduce the variance we have no effect on the bias for those sites on the border.
- If we start with small blocks in the original partition and then simultaneously enlarge the blocks along with the number of samples N then it may be possible maintain a given variance (or even reduce the variance) as compared to a partition with larger original block sizes but with a guaranteed smaller bias at each site.

The high-level point is that it is computationally more desirable to run a few extra parallel implementations of the particle filter (corresponding to more (enlarged) blocks) and obtain a tighter bias bound than it is to run a few less parallel implementations of the particle filter for the same variance bound but a larger bias bound. This is only possible through enlargement of the blocks as described herein.

Finally, we comment on the matter of consistency (as defined in say [10]) and observational double counting. Consider the partition $\mathcal{K} = \{v\}_{v \in V}$ and suppose $\bar{K} = V$ for each $K = v \in \mathcal{K}$. Practically, following the standard prediction step, the enlarged blocked filter is of the form $B^{-1}C_n^E B^E \rho$ which is mathematically equivalent to $BC_n \rho$. The point of this illustration is to highlight that even in this case, involving the most extreme enlargement possible, we are not double counting information or effectively applying measurements twice, and the enlarged blocked particle filter is consistent as per [10].

4.4 Proof Strategy

In this section we provide a summary of the proof strategy. Clearly the main result in Theorem 3 is immediately implied by Theorems 4 and 5. Much of the technical analysis required in the proof of Theorem 3 is similar to that originally detailed in [15].

In the case of the bias $\|\pi_n^\mu - \tilde{\pi}_n^\mu\|_J$, one first derives a local stability property for the filter π_n^μ which implies that the marginal over a local set $J \subseteq V$ of the initial state μ is forgotten exponentially fast. Such a property also implies that any approximation errors in, say, the initial state are also forgotten. It then follows that if one can bound the one-step approximation error $\|F_n \tilde{\pi}_{n-1}^\mu - \tilde{F}_n \tilde{\pi}_{n-1}^\mu\|_J$ at any time, then in conjunction with the local stability property one will obtain a time-uniform bound on the bias over a local region of the field.

In the case of the variance $\|\tilde{\pi}_n^\mu - \hat{\pi}_n^\mu\|_J$, a similar idea is used except one first establishes stability for the ideal enlarged blocked filter $\tilde{\pi}_n^\mu$. Then, one must bound the one-step approximation error $\|\tilde{F}_n \hat{\pi}_{n-1}^\mu - \hat{F}_n \hat{\pi}_{n-1}^\mu\|_J$ at any time. Putting the stability property and the bound on the one-step approximation together, one achieves the desired time-uniform bound on the variance of a block in the adaptively blocked filter.

We have obviously glossed over much of the intricacies involved in the proof in this summary. For example, in the case of the bias, the property introduced in [15] and referred to as the decay of correlations must be established to hold uniformly in time for the ideal block filter $\tilde{\pi}_n^\mu$. This property captures a notion of spatial stability where the state at some site in the random field is forgotten as one moves away from that site. Rebescini et al. provide a novel measure of this decay that allows them to establish local stability of the filter π_n^μ and to establish a bound on the one-step approximation error $\|F_n \tilde{\pi}_{n-1}^\mu - \tilde{F}_n \tilde{\pi}_{n-1}^\mu\|_J$. Conceptually, a property like the decay of correlations is necessary to establish such results.

Summarising, the steps needed to prove the bias bound are

1. Proving a (local) stability result for the ideal Bayesian filter;
2. Proving that a desired decay of correlation property holds uniformly in time for the measure $\tilde{\pi}_n^x$;
3. Controlling the one time-step error introduced by the new enlarged blocked filter;
4. Putting all these results together and finalising Theorem 4.

The variance analysis follows much the same path with the prime difficulty being establishment of local stability for the ideal enlarged blocked filter. Summarising the steps involved in proving the variance bound,

1. Proving a local stability result for the ideal enlarged blocked filter;
2. Controlling the one time-step error due to the sampling in the enlarged blocked particle filter;
3. Putting these results together and finalising Theorem 5

The proof details are omitted in this version of the work due to their similarity with those details presented in [15], but are available upon request.

5 Concluding Remarks

We have presented a modified version of the blocked particle filter originally proposed in [15]. The main feature of our algorithm is that we add a new parameter that can be tuned to decrease the bias as compared to [15]. The high-level argument for this approach is that it is computationally more desirable to run a few extra parallel implementations of the particle filter (corresponding to more (enlarged) blocks) and obtain a tighter bias bound than it is to run a few less parallel implementations of the particle filter for the same variance bound but a larger bias bound. This gain in bias reduction, with the same variance, and only a linear increase in the computational complexity, is only possible through enlargement of the blocks as described herein.

Finally, we also point out that the same adaptive approach to changing partitions proposed in [2] could be applied in the case of the enlarged blocked filter and this is an additional method for spatial smoothing and may be of interest in those cases in which the underlying model is time-varying.

References

- [1] B.D.O. Anderson and J.B. Moore. *Optimal Filtering*. Prentice Hall, Englewood Cliffs, N.J., 1979.
- [2] F Bertoli and A.N. Bishop. Adaptively blocked particle filtering with spatial smoothing in large-scale dynamic random fields. *arXiv e-print arXiv:1406.0136*, 2014.

- [3] A. Beskos, D. Crisan, and A. Jasra. On the stability of sequential Monte Carlo methods in high dimensions. *The Annals of Applied Probability*, 24(4):1396–1445, August 2014.
- [4] A. Beskos, D. Crisan, A. Jasra, and N. Whiteley. Error bounds and normalising constants for sequential Monte Carlo samplers in high dimensions. *Advances in Applied Probability*, 46(1):279–306, March 2014.
- [5] P. Bickel, B. Li, and T. Bengtsson. Sharp failure rates for the bootstrap particle filter in high dimensions. In *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, volume 3, pages 318–329. Institute of Mathematical Statistics, 2008.
- [6] P. Del Moral and A. Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. *Annales de l'Institut Henri Poincaré (B) Probability and Statistics*, 37(2):155–194, March 2001.
- [7] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer, New York, N.Y., 2001.
- [8] N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F (Radar and Signal Processing)*, 140(2):107–113, April 1993.
- [9] A.H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, N.Y., 1970.
- [10] S.J. Julier and J.K. Uhlmann. A non-divergent estimation algorithm in the presence of unknown correlations. In *Proceedings of the 1997 American Control Conference*, Albuquerque, New Mexico, June 1997.
- [11] S.J. Julier and J.K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, Mar 2004.
- [12] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering*, 82(Series D):35–45, 1960.
- [13] Genshiro Kitagawa. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25, 1996.
- [14] P.B. Quang, C. Musso, and F. Le Gland. An insight into the issue of dimensionality in particle filtering. In *Proceedings of the 13th International Conference on Information Fusion*, Edinburgh, Scotland, July 2010.
- [15] P. Rebeschini and R. van Handel. Can local particle filters beat the curse of dimensionality? *arXiv e-print arXiv:1301.6585*, 2013.
- [16] P. Rebeschini and R. van Handel. Comparison theorem for Gibbs measures. *arXiv e-print arXiv:1308.4117*, 2013.
- [17] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson. Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, 136(12):4629–4640, December 2008.