A Stochastic Covariance Shrinkage Approach to Particle Rejuvenation in the Ensemble Transform Particle Filter

Andrey A Popov[1], Amit N Subrahmanya[1], and Adrian Sandu[1]

[1]Computational Science Laboratory, Department of Computer Science, Virginia Tech, 2202 Kraft Drive, Blacksburg, VA, 24060, USA

Correspondence: Andrey A Popov (apopov@vt.edu)

Abstract. Rejuvenation in particle filters is necessary to prevent the collapse of the weights when the number of particles is insufficient to sample the high probability regions of the state space. Rejuvenation is often implemented in a heuristic manner by the addition of stochastic samples that widen the support of the ensemble. This work aims at improving canonical rejuvenation methodology by the introduction of additional prior information obtained from climatological samples; the dynamical

5   particles used for importance sampling are augmented with samples obtained from stochastic covariance shrinkage. The ensemble transport particle filter, and its second order variant, are extended with the proposed rejuvenation approach. Numerical experiments show that modified filters significantly improve the analyses for low dynamical ensemble sizes.

1   Introduction

Ensemble-based data assimilation (Asch et al., 2016; Law et al., 2015; Reich and Cotter, 2015) aims to estimate our uncertainty

10   about the state of some dynamical system through an ensemble of possible states. The assumed underlying distribution of these states is taken to be our uncertainty.

Oftentimes the ensemble cannot describe the distribution of our uncertainty to sufficient accuracy. The curse of dimensionality (Tan et al., 2018) ensures that true descriptions of general high-dimensional distributions stay out of our reach. Several techniques such as the principal of maximum entropy (Jaynes, 2003) exist in order to attempt to alleviate some of the burden

15   by prescribing a distribution to a set of data by constraining the choice with known or estimated quantities and qualities. For instance, the ensemble Kalman filter (Burgers et al., 1998; Evensen, 1994, 2009), could be thought of as an abuse of the principle of maximum entropy: by discarding information about the underlying dynamical system, and assuming that the ensemble only gives information about a mean (that lives in $\mathbb{R}^n$), and a covariance, the underlying distribution of our uncertainty is taken to be normal. In such a way, any application of Bayes' rule has to transform our assumed prior normal distribution into our

20   assumed posterior normal distribution.

Previous work (Popov et al., 2020) has focused on augmenting the information represented by the ensemble with information derived from covariance shrinkage through a surrogate ensemble in the ensemble transport Kalman filter. In this paper, we extend this idea to the ensemble transport particle filter (ETPF) (Reich, 2013). The ETPF attempts to use importance sampling (Liu, 2008), not for a solution to the problem of Bayesian inference, but to simply transport a given ensemble into one

25 that is equally sampled from some distribution whose moments, in the ensemble limit, approach the moments of the correct posterior distribution. This means that the underlying distribution from which our posterior is sampled, could be potentially very far, in information distance, to the optimal posterior distribution, for a finite ensemble.

This work explores a new approach to particle rejuvenation, which is necessary to prevent weight collapse in particle filters. Instead of heuristics, the approach makes use of prior information to enrich the ensemble subspace. Our contributions are as follows: we introduce an alternative way of performing particle rejuvenation in ETPF by incorporating climatological 30 covariance information. We accomplish this by augmenting the dynamical (model) ensemble with synthetic anomalies with optimal scaling, accompanied by a statistically correct estimator. We show that this method of performing particle rejuvenation significantly improves the analysis RMSE for low dynamical ensemble sizes.

This paper is organized as follows. Section 2 reviews the concept of Bayesian inference with the addition of prior informa-35 tion, and its use in importance sampling. Section 3 introduces the ensemble transform particle filter and its canonical rejuvenation heuristic. The concept of stochastic covariance shrinkage is proposed in Section 4, and ETPF is extended to make use of this shrinkage. Numerical experiment results are reported in Section 5. Finally, concluding remarks are drawn in Section 6.

2 Optimal coupling with prior information and the ensemble transform particle filter

Bayesian inference (Jaynes, 2003) aims at transforming prior information about the state of a system (represented by the 40 distribution of a random variable X^{f}), additional qualitative and quantitative information (P), which we will use to stand for , and information obtained by observing the system (Y), into combined posterior information (X^{a}):

$$\pi_{X^{\mathrm{a}}}(X \mid Y, P) = \frac{\pi_{Y \mid X}(Y \mid X, P)\,\pi_{X^{\mathrm{f}}}(X \mid P)}{\pi_Y(Y \mid P)}, \tag{1}$$

where $\pi_{X^{\mathrm{f}}}(x \mid p)$ represents the prior state probability density conditioned by all other relevant information, and $\pi_{Y \mid X^{\mathrm{f}}}(y \mid x, p)$ is the observational likelihood conditioned by the the forecast and the prior information. Here we consider the finite dimensional 45 case where $X^{\mathrm{f}}, X^{\mathrm{a}} \in \mathbb{R}^n$, $Y \in \mathbb{R}^m$, and the supports of the probability densities $\pi_{X^{\mathrm{f}}}$ and $\pi_{X^{\mathrm{a}}}$ are subsets of the respective spaces.

Classical particle filtering (Reich and Cotter, 2015) represents state distributions by collections of particles, i.e., ensembles of samples. Specifically, consider an ensemble of N^{f} particles $\mathbf{X}^{\mathrm{f}} = [\mathbf{X}_1^{\mathrm{f}}, \dots \mathbf{X}_{N^{\mathrm{f}}}^{\mathrm{f}}] \in \mathbb{R}^{n \times N^{\mathrm{f}}}$. The prior distribution density $\pi_{X^{\mathrm{f}}}$ is approximated weakly by the corresponding empirical measure

$$50 \quad \pi_{\hat{X}^{\mathrm{f}}}(X \mid P) = \sum_{i=j}^{N^{\mathrm{f}}} w_j^{\mathrm{f}} \delta_{X - \mathbf{X}_j^{\mathrm{f}}}, \tag{2}$$

where w_j^{f} for $1 \le j \le N^{\mathrm{f}}$ are the prior importance weights associated with each particle. Similarly, the posterior density is approximated weakly by an empirical measure based on the same sample values (particle states) but with different posterior importance weights w_j^{a} for $1 \le j \le N^{\mathrm{f}}$:

$$\pi_{\hat{X}^{\mathrm{a}}}(x \mid y, p) = \sum_{i=j}^{N^{\mathrm{f}}} w_j^{\mathrm{a}} \delta_{x - \mathbf{X}_j^{\mathrm{f}}}, \tag{3}$$

The posterior importance sampling weights are obtained from eq. (1):

$$w_j^{\mathrm{a}} \propto \pi_Y(y \mid X_j^{\mathrm{f}}, P)\, \pi_P(P \mid X_j^{\mathrm{f}}) = w_j^{\mathrm{f}} \pi_Y(y \mid X_j^{\mathrm{f}}, P). \tag{4}$$

The ensemble of weights is denoted by $\mathbf{w} = [w_1, \ldots w_{N^{\mathrm{f}}}]^T$, and $\mathbf{w}^{\mathrm{f}}$ and $\mathbf{w}^{\mathrm{a}}$ refer to the forecast and analysis weights respectively. Using (3) and (4) empirical estimates of the posterior mean and covariance,

$$\bar{\mathbf{x}}^{\mathrm{a}} = \sum_{j=1}^{N^{\mathrm{f}}} w_j^{\mathrm{a}} \mathbf{X}_j^{\mathrm{f}},$$

$$\boldsymbol{\Sigma}_{X^{\mathrm{a}}} = \frac{N^{\mathrm{f}}}{N^{\mathrm{f}}-1} \mathbf{X}_j^{\mathrm{f}} \left(\mathrm{diag}\left(\mathbf{w}^{\mathrm{a}}\right) - \mathbf{w}^{\mathrm{a}} \mathbf{w}^{\mathrm{a},T} \right) \mathbf{X}_j^{\mathrm{f},T}, \tag{5}$$

respectively, are obtained by the importance sampling approach (Liu, 2008).

The goal of particle filtering (with resampling) is to find an ensemble $\mathbf{X}^{\mathrm{a}} \in \mathbb{R}^{n \times N^{\mathrm{a}}}$ of N^{a} realizations of the random variable X^{a} that represents the posterior distribution $\pi_{X^{\mathrm{a}}}$ with equal weights. Specifically, the the posterior density is approximated weakly by the empirical measure

$$\pi_{\hat{X}^{\mathrm{a}}}(x \mid p) = \sum_{j=1}^{N^{\mathrm{a}}} \frac{1}{N^{\mathrm{a}}} \delta_{x - \mathbf{x}_j^{\mathrm{a}}}, \tag{6}$$

where the importance sampling weights are uniform and equal to $1/N^{\mathrm{a}}$ (so as to be equally likely). We impose that the empirical mean (5) is preserved by (6):

$$\bar{\mathbf{x}}^{\mathrm{a}} = \sum_{j=1}^{N^{\mathrm{a}}} \frac{1}{N^{\mathrm{a}}} \mathbf{X}_j^{\mathrm{a}} = \sum_{j=1}^{N} w_j^{\mathrm{a}} \mathbf{X}_j^{\mathrm{f}} = \mathbf{X}^{\mathrm{f}} \mathbf{w}^{\mathrm{a}}. \tag{7}$$

The optimal coupling (McCann et al., 1995; Reich and Cotter, 2015) between the prior empirical distribution eq. (2) and the posterior empirical distribution eq. (6), can be defined as an ensemble transformation,

$$\mathbf{X}^{\mathrm{a}} = \mathbf{X}^{\mathrm{f}} \mathbf{T}^*, \tag{8}$$

where $\mathbf{T}^* \in \mathbb{R}^{N^{\mathrm{f}} \times N^{\mathrm{a}}}$ is the solution to the optimal transport Monge-Kantorovich problem (Villani, 2003). The discrete optimal transportation problem is

$$\mathbf{T}^* = \arg\min_{\mathbf{T}} \sum_{\substack{1 \leq j \leq N^{\mathrm{f}} \\ 1 \leq k \leq N^{\mathrm{a}}}} \mathbf{T}_{j,k} \left\| \mathbf{X}_j^{\mathrm{f}} - \mathbf{X}_k^{\mathrm{f}} \right\|_2^2 \tag{9}$$

subject to $\quad \mathbf{T} \mathbf{1}_{N^{\mathrm{a}}} = N^{\mathrm{a}} \mathbf{w}^{\mathrm{a}}, \mathbf{T}^T \mathbf{1}_{N^{\mathrm{f}}} = \mathbf{1}_{N^{\mathrm{f}}}, \mathbf{T}_{i,j} \geq 0,$

where the distance measure of squared Euclidean distance is taken for a provably unique solution to the Monge-Kantorovich problem to exist (McCann and Guillen, 2011). The vector of ones of size q is represented by $\mathbf{1}_q$. The problem eq. (9) is a linear programming problem.

The discrete optimal transport eq. (8) formulation begets a mapping $\mathbf{X}^\mathrm{a} = \Psi_{N^\mathrm{f}, N^\mathrm{a}}(\mathbf{X}^\mathrm{f})$, that, in the limit of ensemble size $(N^\mathrm{f} = N^\mathrm{a} \to \infty)$, converges weakly to a mapping Ψ, such that $X^\mathrm{a} = \Psi(X^\mathrm{f})$, which has the exact desired distribution given by eq. (1) (Reich and Cotter, 2015, Theorem 5.19). We believe, but don't prove, that this is likely when N^f and N^a, are not equal.

80   The standard ETPF (Reich, 2013) makes the assumption that the prior and posterior ensemble sizes are the same, $N \coloneqq N^\mathrm{a} = N^\mathrm{f}$ in (9). A second order extension to the ETPF (which we will call ETPF2 here) (Acevedo et al., 2017) modifies the optimal transport equation (8) as follows:

$$\mathbf{X}^\mathrm{a} = \mathbf{X}^\mathrm{f}(\mathbf{T}^* + \mathbf{D}), \tag{10}$$

where the additional term $\mathbf{D}$ is a matrix that ensures that the empirical covariance estimate $\mathbf{\Sigma}_{X^\mathrm{a}}$ from (5) is preserved by (6).

85 3   Particle Rejuvenation in ETPF

Particle and ensemble-based filters often underrepresent uncertainty (Asch et al., 2016) due to the relatively small number of samples when compared to the dimension of the state and data spaces. Over several data assimilation cycles multiple particle start carrying either unimportant or redundant information, which leads to weight collapse or to ensemble degeneracy (Strogatz, 2018). To alleviate these effects, methods such as inflation (Anderson, 2001; Popov and Sandu, 2020), rejuvenation (Reich, 90   2013), and resampling (Reich and Cotter, 2015) have been developed.

In order to avoid ensemble collapse, ETPF employs a particle rejuvenation approach (Acevedo et al., 2017; Reich, 2013; Chustagulprom et al., 2016) that perturbs the analysis ensemble by a random sampling from the ensemble of prior anomalies,

$$\mathbf{X}^\mathrm{a} \leftarrow \mathbf{X}^\mathrm{a} + \sqrt{\frac{\tau}{N-1}} \mathbf{A}^\mathrm{f} \boldsymbol{\eta} \left(\mathbf{I}_N - N^{-1}\mathbf{1}_N\mathbf{1}_N^T\right), \tag{11}$$

where $\boldsymbol{\eta} \sim (\mathcal{N}(0,1))_{N \times N}$ is a matrix of i.i.d. samples from the standard normal distribution of size N, the factor τ is treated 95   as a hyperparameter that controls the magnitude of the correction, and the ensemble anomalies

$$\mathbf{A}^\mathrm{f} = \mathbf{X}^\mathrm{f}\left(\mathbf{I}_N - N^{-1}\mathbf{1}_N\mathbf{1}_N^T\right), \tag{12}$$

are defined as the ensemble of deviations from the sample mean. Of note is the fact that the extra term $\left(\mathbf{I}_N - N^{-1}\mathbf{1}_N\mathbf{1}_N^T\right)$ in (11) ensures that the introduction of the random matrix $\boldsymbol{\eta}$ does not modify the mean of $\mathbf{X}^\mathrm{a}$. This is due to the fact that,

$$\left(\mathbf{I}_N - N^{-1}\mathbf{1}_N\mathbf{1}_N^T\right)\mathbf{1}_N = \mathbf{0}_N,$$
$$\mathbf{1}_N^T\left(\mathbf{I}_N - N^{-1}\mathbf{1}_N\mathbf{1}_N^T\right) = \mathbf{0}_N^T. \tag{13}$$

100   Notice that if we define the matrix,

$$\mathbf{B} \coloneqq \sqrt{\frac{\tau}{N-1}}\left(\mathbf{I}_N - N^{-1}\mathbf{1}_N\mathbf{1}_N^T\right)\boldsymbol{\eta}\left(\mathbf{I}_N - N^{-1}\mathbf{1}_N\mathbf{1}_N^T\right), \tag{14}$$

it is possible to write the ETPF with rejuvenation (11) as,

$$
\begin{aligned}
\mathbf{X}^{\mathrm{a}} &= \mathbf{X}^{\mathrm{f}}\mathbf{T}^{*} + \sqrt{\frac{\tau}{N-1}}\mathbf{A}^{\mathrm{f}}\boldsymbol{\eta}\left(\mathbf{I}_N - N^{-1}\mathbf{1}_N\mathbf{1}_N^T\right) \\
&= \mathbf{X}^{\mathrm{f}}\mathbf{T}^{*} + \sqrt{\frac{\tau}{N-1}}\mathbf{X}^{\mathrm{f}}\left(\mathbf{I}_N - N^{-1}\mathbf{1}_N\mathbf{1}_N^T\right)\boldsymbol{\eta}\left(\mathbf{I}_N - N^{-1}\mathbf{1}_N\mathbf{1}_N^T\right) \\
&= \mathbf{X}^{\mathrm{f}}\left[\mathbf{T}^{*} + \sqrt{\frac{\tau}{N-1}}\left(\mathbf{I}_N - N^{-1}\mathbf{1}_N\mathbf{1}_N^T\right)\boldsymbol{\eta}\left(\mathbf{I}_N - N^{-1}\mathbf{1}_N\mathbf{1}_N^T\right)\right] \\
&= \mathbf{X}^{\mathrm{f}}\widetilde{\mathbf{T}} \quad \text{with} \quad \widetilde{\mathbf{T}} := \mathbf{T}^{*} + \mathbf{B},
\end{aligned}
\tag{15}
$$

with the matrix $\mathbf{B}$ acting as a stochastic perturbation of the optimal transport operator $\mathbf{T}^{*}$, such that $\mathbf{B}$ preserves the constraints $\mathbf{1}_N^T\mathbf{B} = \mathbf{0}_N^T$ and $\mathbf{B}\mathbf{1}_N = \mathbf{0}_N$ in eq. (9) because of the results in (13). This, of course, immediately calls into question the optimality of the transport for a finite ensemble, as adding this type of $\mathbf{B}$ matrix is perturbing the transport mapping $\widetilde{\mathbf{T}}$ away from the optimum $\mathbf{T}^{*}$.

4   Particle Rejuvenation Through Stochastic Shrinkage

In the context of ensemble methods, covariance shrinkage (Nino-Ruiz and Sandu, 2018, 2015; Ruiz et al., 2014) is used, similar to other canonical covariance tapering techniques such as inflation (Anderson, 2001; Popov and Sandu, 2020), localization (Anderson, 2012; Hunt et al., 2007; Nino-Ruiz and Sandu, 2017; Nino-Ruiz et al., 2015; Petrie, 2008; Zhang et al., 2010), to enrich the information represented by an undersampled covariance matrix.

From a Bayesian perspective, covariance shrinkage seeks to incorporate additional prior information on error correlations into the analysis, in order to enhance the inference. In many data assimilation models, climatological covariance information is often available, i.e., it is known prior information. Climatological covariances are typically precomputed or derived from climatological models and are often employed in variational data assimilation Lorenc et al. (2015).

Following (Popov et al., 2020), we describe the stochastic covariance shrinkage technique. Instead of perturbing the transform matrix as in (15), we instead consider enhancing the dynamic ensemble $\mathbf{X}^{\mathrm{f}} \in \mathbb{R}^{n \times N}$ with an M member synthetic ensemble $\mathcal{X}^{\mathrm{f}} \in \mathbb{R}^{n \times M}$ of samples independent of the dynamical ensemble. This leads to the total $N + M$ members ensemble:

$$
\mathfrak{X}^{\mathrm{f}} = \begin{bmatrix} \mathbf{X}^{\mathrm{f}} & \mathcal{X}^{\mathrm{f}} \end{bmatrix} \in \mathbb{R}^{n \times (N+M)}
\tag{16}
$$

with weights $\mathfrak{w}^{\mathrm{f}} = [w_1^{\mathrm{f}}, \ldots w_{N+M}^{\mathrm{f}}]$. We make the ansatz that each ensemble member is distributed as

$$
\mathfrak{X}_{:,i}^{\mathrm{f}} \sim \pi_{X^{\mathrm{f}}}(X \mid P),
\tag{17}
$$

which is the full distribution of the forecast conditioned by the prior information that we have provided to the algorithm. Note that (17) is not the empirical measure distribution (2), that only has information from the ensemble members, but rather the 'exact' distribution that is assumed to contain all the information from the forecast. THis is because we are now incorporating more prior information P in the form of climatological information.

Taking (12) to be the anomalies of the dynamic ensemble, and

$$\mathcal{A}^{\mathrm{f}} = \mathcal{X}^{\mathrm{f}}\left(\mathbf{I}_M - M^{-1}\mathbf{1}_M\mathbf{1}_M^T\right), \tag{18}$$

to be the anomalies of the synthetic ensemble, the total empirical (unbiased) covariance can be written as,

130 $$\mathbf{\Sigma}_{\mathfrak{X}^{\mathrm{f}}} = \mathbf{\Sigma}_{\mathbf{X}^{\mathrm{f}}} + \mathbf{\Sigma}_{\mathcal{X}^{\mathrm{f}}}, \tag{19}$$

where the constituent covariances are defined in terms of the weights

$$\mathbf{\Sigma}_{\mathbf{X}^{\mathrm{f}}} = \sum_{i=1}^{N} w_i^{\mathrm{f}} \frac{N}{N-1} \mathbf{A}_{:,i}^{\mathrm{f}} \mathbf{A}_{:,i}^{\mathrm{f},T}, \quad \mathbf{\Sigma}_{\mathcal{X}^{\mathrm{f}}} = \sum_{i=1}^{M} w_{N+i}^{\mathrm{f}} \frac{M}{M-1} \mathcal{A}_{:,i}^{\mathrm{f}} \mathcal{A}_{:,i}^{\mathrm{f},T}. \tag{20}$$

In the covariance shrinkage approach, the sample mean of the synthetic ensemble is assumed to be the sample mean of the dynamic ensemble,

135 $$\mathcal{X}^{\mathrm{f}}\mathbf{1}_M = \mathbf{X}^{\mathrm{f}}\mathbf{1}_N, \tag{21}$$

by construction and (13), thus requiring that only the synthetic ensemble anomalies need to be determined. Taking a covariance matrix $\mathcal{P}$ that represents a climatological estimate of the covariance, we sample the anomalies from some unbiased distribution with a scaling, by a factor μ, of said covariance. In the Gaussian case

$$\mathcal{A}_{:,i}^{\mathrm{f}} \sim \mathcal{N}(\mathbf{0}_n, \mu\mathcal{P}), \tag{22}$$

140 where μ is a scaling factor defined later. An alternate choice of distribution that we explore is the symmetric Laplace distribution (Kozubowski et al., 2013),

$$\mathcal{A}_{:,i}^{\mathrm{f}} \sim \mathcal{L}(\mathbf{0}_n, \mu\mathcal{P}), \tag{23}$$

which is described by the pdf

$$\pi(\mathbf{x}) = \frac{2}{[(2\pi)^n \mu\mathcal{P}]^{\frac{1}{2}}} \left(\frac{\mathbf{x}^T \mathcal{P}^{-1}\mathbf{x}}{2\mu}\right)^{\frac{2-n}{4}} K_{\frac{2-n}{2}}\left(\sqrt{\frac{2\mathbf{x}^T \mathcal{P}^{-1}\mathbf{x}}{\mu}}\right), \tag{24}$$

145 where K is the modified Bessel function of the second kind (Olver et al., 2010). The choice of Laplace distribution is motivated by robust statistics techniques (Rao et al., 2017). The resulting sampled covariance would therefore be an estimate of the scaled climatological covariance,

$$\mathbf{\Sigma}_{\mathcal{X}^{\mathrm{f}}} \approx \mu\mathcal{P}. \tag{25}$$

Remark 1. In order to stay consistent with the mean estimate, the anomalies are replaced with their sample mean zero coun-
150 terparts

$$\mathcal{A}^{\mathrm{f}} \leftarrow \mathcal{A}^{\mathrm{f}}\left(\mathbf{I}_M - M^{-1}\mathbf{1}_M\mathbf{1}_M^T\right). \tag{26}$$

The weights $\mathfrak{w}^{\mathrm{f}}$ are divided into two classes: those that are associated with the dynamic ensemble, and those that are associated with the synthetic ensemble,

$$
w_i^{\mathrm{f}} = \begin{cases} 1-\gamma & 1 \leq i \leq N \\ \gamma & N+1 \leq i \leq N+M \end{cases}, \tag{27}
$$

where the parameter γ is known as the covariance shrinkage factor.

One choice to calculate γ is the Rao-Blackwell Ledoit-Wolf (RBLW) estimator (Chen et al., 2009) (Nino-Ruiz and Sandu, 2017, equation (9)),

$$
\gamma_{\mathrm{RBLW}} = \left[\frac{N-2}{N(N+2)} + \frac{(n+1)N-2}{\hat{U}(\mathcal{P}, \boldsymbol{\Sigma}_{\mathbf{X}^{\mathrm{f}}})N(N+2)(n-1)} \right] \wedge 1, \tag{28}
$$

where the sphericity factor,

$$
\hat{U}(\mathcal{P}, \boldsymbol{\Sigma}_{\mathbf{X}^{\mathrm{f}}}) := \frac{1}{n-1} \left(\frac{n \operatorname{tr}(\mathcal{C}_i^2)}{\operatorname{tr}^2(\mathcal{C}_i)} - 1 \right), \quad \text{with} \quad \mathcal{C} := \mathcal{P}^{-\frac{1}{2}} \boldsymbol{\Sigma}_{\mathbf{X}^{\mathrm{f}}} \mathcal{P}^{-\frac{1}{2}}, \tag{29}
$$

represents the mismatch between the climatological covariance (called the "target" in statistical literature) and sample covariance matrices. Note that if $\mathcal{P} = \boldsymbol{\Sigma}_{\mathbf{X}^{\mathrm{f}}}$, then $\hat{U}(\mathcal{P}, \boldsymbol{\Sigma}_{\mathbf{X}^{\mathrm{f}}}) = 0$. In such a framework the scaling parameter for the climatological covariance is defined to be

$$
\mu = \frac{\operatorname{tr}(\mathcal{C})}{n}. \tag{30}
$$

Remark 2. The RBLW estimator (28) makes the assumption that the underlying distribution of the dynamic ensemble is Gaussian. Typically this assumption is violated for dynamical systems of interest.

Remark 3. In statistical literature, the target covariance is often taken to be the identity, $\mathcal{P} = \mathbf{I}$, which implies that $\mathcal{C} = \boldsymbol{\Sigma}_{\mathbf{X}^{\mathrm{f}}}$ in (29). The assumption that the target is a climatological covariance is natural generalization in the specific context of sequential data assimilation.

Remark 4. The scaling of the target matrix $\mathcal{P}$ is not of any consequence. Let $\widetilde{\mathcal{P}} = \beta \mathcal{P}$ be a scalar scaling of the target matrix, then $\widetilde{\mathcal{C}}_i = \frac{1}{\beta} \mathcal{C}_i$, implying that $\widetilde{\mu}_i = \frac{1}{\beta} \mu_i$, rendering the matrix scaling inconsequential.

The resulting analysis ensemble based on prior states and importance sampling weights of $N+M$ states is transported into an equally weighted posterior ensemble of N states through the transformation

$$
\mathbf{X}^{\mathrm{a}} = \mathfrak{X}^{\mathrm{f}} \mathbf{T}^*, \tag{31}
$$

where the optimal transport matrix $\mathbf{T}^* \in \mathbf{R}^{(N+M) \times N}$ is computed by solving (9).

Recall that in the traditional method of rejuvenation (15), the optimal transport matrix is perturbed randomly into a nearby transport matrix; no new prior information is introduced. We take a fundamentally different approach by incorporating "unseen"

prior information derived from a climatological covariance. To this end, we augment of the empirical measure distribution (2) with samples from the climatological distribution, to accommodate the total ensemble (16),

$$\pi_{\hat{X}^{\mathrm{f}}}(x \mid p) = \sum_{i=j}^{N^{\mathrm{f}}} w_j^{\mathrm{f}} \, \delta_{x - \mathfrak{X}_j^{\mathrm{f}}}, \tag{32}$$

before the Monge-Kantorovich problem (9) is solved. In effect we are able to avoid ensemble collapse by enhancing the empirical measure distribution (32) with new prior information, as opposed to a reweighing of the old prior information. We will denote our method as FETPF, standing for 'foresight' ETPF, as we believe including prior information in the analysis procedure is a type of foresight.

4.1   Multiple Climatological Covariance Matrices

It is conceivable that multiple climatological models give rise to multiple climatological covariances, or alternatively multiple candidates for the most 'common' behavior of the model is to be chosen.

Given a collection of target covariances, $\{\mathcal{P}_j\}_{j \in \mathcal{J}}$, we must choose the appropriate covariance from which to sample. We consider the sphericity of the mismatch between the target and forecast covariances eq. (29). Based on authors' numerical experience, we select the target covariance that corresponds to the highest sphericity of the mismatch:

$$\mathcal{P}^* = \underset{\mathcal{P}_j}{\arg\max} \; \hat{U}(\mathcal{P}_j, \boldsymbol{\Sigma}_{\mathbf{X}^{\mathrm{f}}}), \tag{33}$$

We can justify this choice by realizing that the smaller the sphericity, the closer our samples are to that of canonical rejuvenation techniques. The aim of climatological shrinkage is to introduce unknown information into our inference procedure, thus the target covariance with the highest mismatch introduces the highest amount of outside information.

Remark 5. It is also possible to construct 'multi-target' shrinkage estimators (Lancewicki and Aladjem, 2014) that consider all target matrices simultaneously.

5   Numerical Experiments

For all our experiments we use the Lorenz '63 system (Lorenz, 1963):

$$
\begin{aligned}
x' &= \sigma(y - x), \\
y' &= x(\rho - z) - y, \\
z' &= xy - \beta z,
\end{aligned}
\tag{34}
$$

with canonical parameter values $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$. We observe the first component, with Gaussian unbiased observation error, with a variance of $\mathbf{R} = 8$. The implementation is from our test problem suite (Computational Science Laboratory, 2020; Roberts et al., 2019).
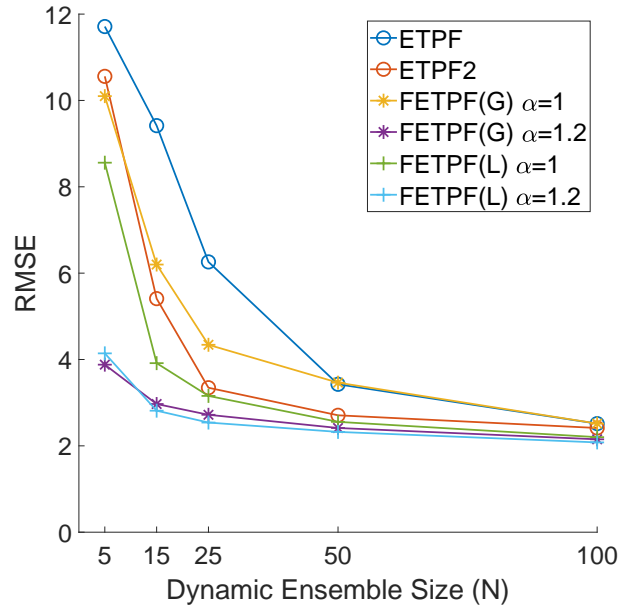
Figure 1. Analysis RMSE versus dynamic ensemble size (N) of the Gaussian (G) and Laplace (L) covariance shrinkage approaches (FETPF) to particle rejuvenation for a synthetic ensemble size of $M = 100$ and for various values of synthetic anomaly inflation, α, from (37), with respect to a canonical particle rejuvenation approach for first order ETPF and second order ETPF (denoted ETPF2) for an optimal rejuvenation factor $\tau = 0.04$. The target covariance is taken to be (35).

As discussed previously, the canonical choice for the shrinkage covariance is the identity matrix. It has been the authors'
205 experience that for most dynamical systems the choice is poor. Moreover, the sequential data assimilation problem typically
provides ways to calculate climatological approximations to the covariance. We take advantage of such techniques in this paper.

The first type of climatological covariance that we investigate is that of the distribution over that of the whole manifold of
the dynamics. The (trace-state normalized) matrix that is obtained by taking the temporal covariance of 50000 sample points
on the attractor of the canonical Lorenz '63 model is:

$$\mathcal{P} = \begin{bmatrix} 0.8616 & 0.8618 & -0.0148 \\ 0.8618 & 1.1149 & -0.0035 \\ -0.0148 & -0.0035 & 1.0234 \end{bmatrix}, \tag{35}$$

210 with condition number 15.88.

For testing multiple covariances, we run an ETPF with $N = 100$ with 20000 evenly spaced samples over a time interval
of 2400 time units and calculate the trace-state normalized forecast covariances. Under a square Frobenius norm distance, we
cluster the empirical covariance matrices of the same ensemble at different times using the k-means algorithm (Tan et al., 2018)
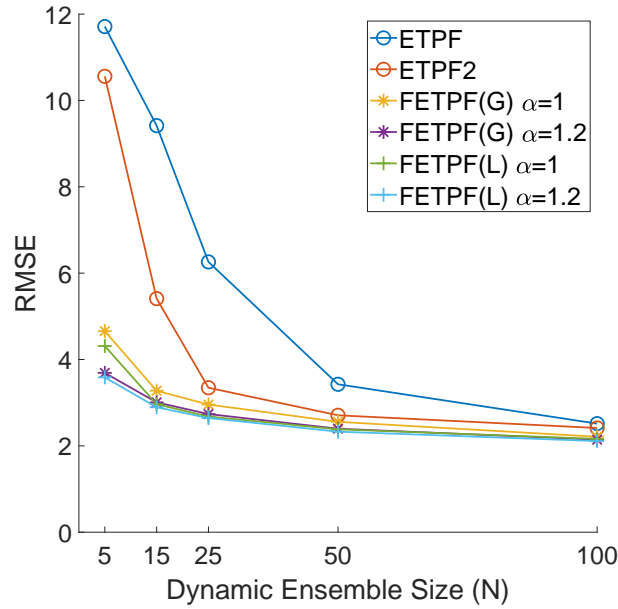
Figure 2. Analysis RMSE versus Dynamic ensemble size (N) of the Gaussian (G) and Laplace (L) covariance shrinkage approaches (FETPF) to particle rejuvenation with the multi-target covariances (36) for a synthetic ensemble size of $M = 100$ and for various values of synthetic anomaly inflation (α) with respect to a canonical particle rejuvenation approach for first order ETPF and second order ETPF (denoted ETPF2) for an optimal rejuvenation factor $\tau = 0.04$.

into two clusters. The collection of climatological covariances for the Lorenz '63 thus consists of the centroids of each cluster,

$$215 \quad \mathcal{P}_1 = \begin{bmatrix} 0.5017 & 0.5524 & -0.4587 \\ 0.5524 & 1.0731 & -0.6723 \\ -0.4587 & -0.6723 & 1.4252 \end{bmatrix}, \quad \mathcal{P}_2 = \begin{bmatrix} 0.5443 & 0.6830 & 0.4330 \\ 0.6830 & 1.2748 & 0.6318 \\ 0.4330 & 0.6318 & 1.1808 \end{bmatrix}, \quad (36)$$

with condition numbers 13.68 and 16.98 respectively. As can be seen, the clusters are mainly split by the correlation factors of z with respect to the other variables being positive or negative.

In order to stay in line with other particle rejuvenation techniques, a heuristic that we use is inflation on the synthetic ensemble, so that it is possible to overcome deficiencies in its descriptive power:

$$220 \quad \mathcal{A}^{\mathrm{f}} \leftarrow \alpha \mathcal{A}^{\mathrm{f}}, \quad (37)$$

which is equivalent to assuming an inflated scaling factor μ in (30). We therefore have two parameters that we can configure in our rejuvenation technique: M, the size of our surrogate synthetic ensemble, and α, the inflation applied to its realizations. It is the authors' experience that inflation should only be applied to the synthetic ensemble, and not the dynamical ensemble, otherwise the shrinkage estimate (28) does not lead to a stable algorithm.
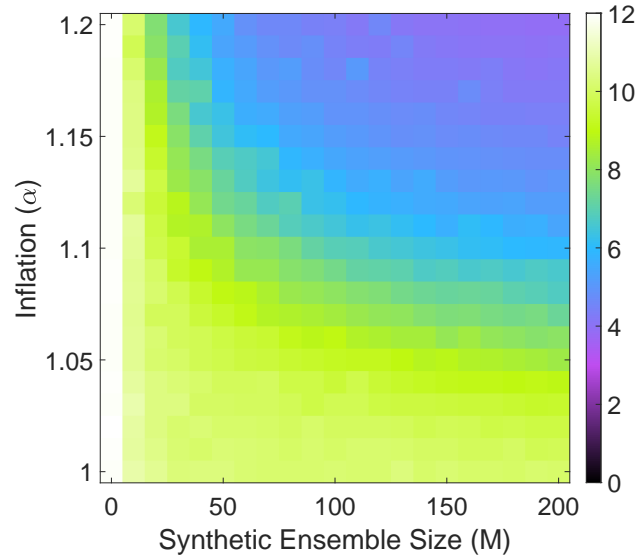
Figure 3. Analysis RMSE of the covariance shrinkage approach to particle rejuvenation (FETPF) for different values of the synthetic ensemble size M and synthetic inflation factor α, and for a dynamical ensemble size of $N = 5$.

In our experiments we report the error of the analysis with respect to the truth, measured by the spatio-temporal RMSE, given by the formula,

$$\text{RMSE}(\mathbf{x}^{\text{t}}, \bar{\mathbf{x}}^{\text{a}}) = \sqrt{\frac{1}{nK} \sum_{i=1}^{T} \sum_{j=1}^{n} \left([\mathbf{x}_i^{\text{t}}]_j - [\bar{\mathbf{x}}_i^{\text{a}}]_j \right)^2}, \tag{38}$$

where T stands for the relevant measured timeframe of the experiments.

Our first round of experiments reported in Figure 1 compares the canonical method of rejuvenation in ETPF (and second order extensions) with the optimal rejuvenation factor of $\tau = 0.04$ in (14) (computed by parameter search) to the stochastic covariance shrinkage technique for both Gaussian and Laplace samples. A dynamic ensemble size of $M = 100$, the inflation factors $\alpha = \{1.0, 1.2\}$. The target covariance (35) is used. We perform 10000 assimilation steps, but discard the first 1000 that are used for spinup. The time interval between successive observations is $\Delta t = 0.12$. We perform 20 independent runs and take the mean of the results to obtain an accurate estimate of the expected error.

Results in Figure 1 show that the stochastic covariance shrinkage technique converges to the same 'correct' RMSE in the case of a large dynamic ensemble of $N = 100$. Moreover, in the small ensemble case of $N = 5$, methods that inflate the synthetic ensemble with factor $\alpha = 1.2$ perform significantly better than those that do not. Laplace distributed synthetic samples also seem to slightly reduce the error.

Our second round of experiments reported in Figure 2 makes use of multiple values of the climatological covariance $\mathcal{P}$,
240 namely those in (36). The rest of the setup is identical to the previous experiment. Again, for a large value of dynamic ensemble size $N = 100$, the stochastic covariance shrinkage approach attains the correct error statistics. For a low ensemble size, however, there is virtually no difference between the Gaussian, Laplace, inflation, and no inflation stochastic covariance shrinkage methods as compared to the ETPF.

Our final round of experiments seeks to understand the effect of selecting the two free parameters, i.e., the synthetic ensemble
245 size M and the synthetic ensemble inflation factor α. Figure 3 shows the spatio-temporal RMSE of a small dynamic ensemble ($N = 5$) for various values of M and α, with Gaussian synthetic samples using the single target matrix (35). The results clearly show that in many operationally useful cases, it is necessary to have a sufficiently expressive synthetic ensemble, whose anomalies are sufficiently inflated.

The results of the first two experiments show that adding additional synthetic information during assimilation is more effec-
250 tive than randomly perturbing the ensemble post-assimilation. The authors hypothesize that the results point strongly towards the need of intelligently, and adaptively choosing the target covariance matrices, and to the need for better operational calculation of the covariance shrinkage factor γ.

6 Conclusions

This paper introduces a stochastic covariance shrinkage-based particle rejuvenation technique for the ensemble transport par-
255 ticle filter. Instead of reweighing existing prior information, the approach incorporates additional prior information into the ensemble through the use of synthetic anomalies. These anomalies come from climatological covariance information. Numerical experiments show that the use of climatological prior information to perform rejuvenation leads to reduced analyses errors for significantly smaller dynamical ensemble sizes than the original rejuvenation approach.

We believe that the stochastic covariance shrinkage approach to importance sampling can be used not just for particle
260 rejuvenation in the ETPF, but in other particle filters as well.

Acknowledgements

This work was supported by awards NSF CDS&E-MSS–1953113, DOE ASCR DE–SC0021313, and by the Computational Science Laboratory at Virginia Tech.

References

265   Acevedo, W., de Wiljes, J., and Reich, S.: Second-order accurate ensemble transform particle filters, SIAM Journal on Scientific Computing, 39, A1834–A1850, 2017.

Anderson, J. L.: An ensemble adjustment Kalman filter for data assimilation, Monthly weather review, 129, 2884–2903, 2001.

Anderson, J. L.: Localization and sampling error correction in ensemble Kalman filter data assimilation, Monthly Weather Review, 140, 2359–2371, 2012.

270   Asch, M., Bocquet, M., and Nodet, M.: Data assimilation: methods, algorithms, and applications, SIAM, 2016.

Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, Monthly weather review, 126, 1719–1724, 1998.

Chen, Y., Wiesel, A., and Hero, A. O.: Shrinkage estimation of high dimensional covariance matrices, in: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 2937–2940, IEEE, 2009.

275   Chustagulprom, N., Reich, S., and Reinhardt, M.: A hybrid ensemble transform particle filter for nonlinear and spatially extended dynamical systems, SIAM/ASA Journal on Uncertainty Quantification, 4, 592–608, 2016.

Computational Science Laboratory: ODE Test Problems, https://github.com/ComputationalScienceLaboratory/ODE-Test-Problems, 2020.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, Journal of Geophysical Research: Oceans, 99, 10 143–10 162, 1994.

280   Evensen, G.: Data assimilation: the ensemble Kalman filter, Springer Science & Business Media, 2009.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, Physica D: Nonlinear Phenomena, 230, 112–126, 2007.

Jaynes, E. T.: Probability theory: The logic of science, Cambridge university press, 2003.

Kozubowski, T. J., Podgórski, K., and Rychlik, I.: Multivariate generalized Laplace distribution and related random fields, Journal of Multi-
285   variate Analysis, 113, 59–72, 2013.

Lancewicki, T. and Aladjem, M.: Multi-target shrinkage estimation for covariance matrices, IEEE Transactions on Signal Processing, 62, 6380–6390, 2014.

Law, K., Stuart, A., and Zygalakis, K.: Data assimilation: a mathematical introduction, vol. 62, Springer, 2015.

Liu, J. S.: Monte Carlo strategies in scientific computing, Springer Science & Business Media, 2008.

290   Lorenc, A. C., Bowler, N. E., Clayton, A. M., Pring, S. R., and Fairbairn, D.: Comparison of hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP, Monthly Weather Review, 143, 212–229, 2015.

Lorenz, E. N.: Deterministic nonperiodic flow, Journal of the atmospheric sciences, 20, 130–141, 1963.

McCann, R. J. and Guillen, N.: Five lectures on optimal transportation: geometry, regularity and applications, Analysis and geometry of metric measure spaces: lecture notes of the séminaire de Mathématiques Supérieure (SMS) Montréal, pp. 145–180, 2011.

295   McCann, R. J. et al.: Existence and uniqueness of monotone measure-preserving maps, Duke Mathematical Journal, 80, 309–324, 1995.

Nino-Ruiz, E. D. and Sandu, A.: Ensemble Kalman filter implementations based on shrinkage covariance matrix estimation, Ocean Dynamics, 65, 1423–1439, https://doi.org/10.1007/s10236-015-0888-9, 2015.

Nino-Ruiz, E. D. and Sandu, A.: An ensemble Kalman filter implementation based on modified Cholesky decomposition for inverse covariance matrix estimation, SIAM Journal on Scientific Computing, submitted, https://arxiv.org/abs/1605.08875, 2017.

300   Nino-Ruiz, E. D. and Sandu, A.: Efficient parallel implementation of DDDAS inference using an ensemble Kalman filter with shrinkage covariance matrix estimation, Cluster Computing, pp. 1–11, https://link.springer.com/article/10.1007/s10586-017-1407-1, 2018.

Nino-Ruiz, E. D., Sandu, A., and Deng, X.: A parallel ensemble Kalman filter implementation based on modified Cholesky decomposition, in: Proceedings of the 6th Workshop on Latest Advances in Scalable Algorithms for Large-Scale Systems, vol. Supercomputing 2015 of *ScalA '15*, Austin, Texas, https://doi.org/10.1145/2832080.2832084, 2015.

305   Olver, F. W., Lozier, D. W., Boisvert, R. F., and Clark, C. W.: NIST handbook of mathematical functions hardback and CD-ROM, Cambridge university press, 2010.

Petrie, R.: Localization in the ensemble Kalman filter, MSc Atmosphere, Ocean and Climate University of Reading, 2008.

Popov, A. A. and Sandu, A.: An Explicit Probabilistic Derivation of Inflation in a Scalar Ensemble Kalman Filter for Finite Step, Finite Ensemble Convergence, arXiv preprint arXiv:2003.13162, 2020.

310   Popov, A. A., Sandu, A., Nino-Ruiz, E. D., and Evensen, G.: A Stochastic Covariance Shrinkage Approach in Ensemble Transform Kalman Filtering, 2020.

Rao, V., Sandu, A., Ng, M., and Nino-Ruiz, E. D.: Robust Data Assimilation using L_1 and Huber norms, SIAM Journal on Scientific Computing, 39, B548–B570, 2017.

Reich, S.: A nonparametric ensemble transform method for Bayesian inference, SIAM Journal on Scientific Computing, 35, A2013–A2024, 2013.

315   Reich, S. and Cotter, C.: Probabilistic forecasting and Bayesian data assimilation, Cambridge University Press, 2015.

Roberts, S., Popov, A. A., and Sandu, A.: ODE Test Problems: a MATLAB suite of initial value problems, CoRR, abs/1901.04098, http://arxiv.org/abs/1901.04098, 2019.

Ruiz, E. N., Sandu, A., and Anderson, J.: An efficient implementation of the ensemble Kalman filter based on an iterative Sherman-Morrison formula, Statistics and Computing, pp. 1–17, https://doi.org/10.1007/s11222-014-9454-4, 2014.

Strogatz, S. H.: Nonlinear dynamics and chaos with student solutions manual: With applications to physics, biology, chemistry, and engineering, CRC press, 2018.

Tan, P.-N., Steinbach, M., Karpatne, A., and Kumar, V.: Introduction to Data Mining (2nd Edition), Pearson, 2nd edn., 2018.

Villani, C.: Topics in optimal transportation, 58, American Mathematical Soc., 2003.

325   Zhang, Y., Liu, N., and Oliver, D. S.: Ensemble filter methods with perturbed observations applied to nonlinear problems, Computational Geosciences, 14, 249–261, 2010.