

# Ensemble Gaussian Mixture Filtering with Particle-localized Covariances

Andrey A Popov  
The Oden Institute for  
Computational Engineering & Sciences  
The University of Texas at Austin  
Austin, Texas  
andrey.a.popov@utexas.edu

Renato Zanetti  
Dept. of Aerospace Engineering and  
Engineering Mechanics  
The University of Texas at Austin  
Austin, Texas  
renato@utexas.edu

**Abstract**—The ensemble Gaussian mixture filter (EnGMF) is a powerful filter for highly non-Gaussian and non-linear models that has practical utility in the case of a small number of samples, and theoretical convergence to full Bayesian inference in the ensemble limit. We aim to increase the utility of the EnGMF by introducing a particle-local notion of covariance into the Gaussian mixture estimate of the prior distribution. We show on a simple bivariate problem that each particle having its own local estimate of the covariance both has nice qualitative and quantitative properties, and significantly improves the estimate of the prior and posterior distributions for all ensemble sizes. We additionally show the utility of the proposed methodology for sequential filtering for the Lorenz '63 equations, achieving a significant reduction in error in the low ensemble size regime.

**Index Terms**—ensemble Gaussian mixture filter, localization, sequential filtering, data assimilation

## I. INTRODUCTION

The sample covariance is a good measure of the global relationship between state variables, but fails to capture relationships between variables for distributions whose curvature is highly non-Gaussian. In this work we present a way to more accurately capture these relationships for state estimation in the ensemble Gaussian mixture filter (EnGMF).

The first law of geography [1] states that “everything is related to everything else, but near things are more related than distant things.” In the context of state estimation and data assimilation for geophysical systems [2], [3] this law begets heuristics known as covariance localization. Common approaches to covariance localization rely on spatial distances between state and observation variables [2]. The two predominant approaches are B-localization, where the prior covariance is scaled in a way such that far away state variables do not influence each other, and R-localization, where the observation covariances are scaled in a way such that state variables that are far away from an observation are not influenced by it. These approaches fundamentally rely on a spatial understanding of state variables, and are not fit for general problems where the state variables might not have a spatial structure. A new more general approach to localization is therefore required.

This work was sponsored in part by DARPA (Defense Advanced Research Projects Agency) under STTR contract number W31P4Q-21-C-0032.

We propose a novel approach to localization that does not require an innate spatial structure in the state variables, but instead relies on inter-particle distance. This idea is loosely based on the use of inter-particle distance in the ensemble transform particle filter (ETPF) [4], though in the ETPF the inter-particle distance is used to solve an optimal transport problem. Continuing the naming trend, we call our approach E-localization, standing for ensemble localization. In this work we apply this methodology to the EnGMF to beget the E-localized EnGMF (EEnGMF).

We show that the proposed approach is able to more accurately describe the prior distribution with less samples, and that this description of the prior leads to a better posterior. We show this both in a qualitative and quantitative manner on a simple test problem. We additionally show that this approach begets better performance in the sequential filtering setting.

This paper is organized as follows: we first provide some background on the filtering problem and introduce the EnGMF in section II. We next introduce the E-localization methodology in section III. We illustrate the difference between the classic EnGMF and the E-localized EnGMF for a simple bivariate example in section IV. Penultimately, we showcase the EEnGMF on the Lorenz '63 equations in section V. Finally, we have some concluding remarks in section VI.

## II. ENSEMBLE GAUSSIAN MIXTURE FILTER

We first provide a brief overview of the state estimation problem/data assimilation problem. Assume that we wish to estimate the true state  $x^t$  of some physical process, but only have access to noisy non-linear observations of it,

$$\mathbf{y} = h(\mathbf{x}^t) + \boldsymbol{\eta}, \quad (1)$$

where  $\boldsymbol{\eta}$  is unbiased additive Gaussian noise,

$$\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), \quad (2)$$

and  $\mathbf{R}$  is the observation error covariance. Given a random variable  $\mathbf{x}^-$  that describes our prior knowledge about the state, we use Bayesian inference to find a random variable

$$\mathbf{x}^+ = \mathbf{x}^- | \mathbf{y}, \quad (3)$$

that describes our posterior knowledge of the state, given the newfound observation eq. (1).

We now describe the inner workings of the EnGMF [5]–[8], in order to understand some of its deficiencies.

Given an ensemble of  $N$  independently and identically distributed (iid) samples from the prior distribution,  $\mathbf{X}^- = [\mathbf{x}_1^-, \dots, \mathbf{x}_N^-]$ , one can use Kernel density estimation techniques to construct an approximation of the prior distribution as a Gaussian mixture model (GMM),

$$p^-(\mathbf{x}) = \frac{1}{N} \sum_{j=1}^N \mathcal{N}(\mathbf{x} | \mathbf{x}_j^-, \mathbf{B}_j^-), \quad (4)$$

where each mean is centered at one of the ensemble members (particles) and the covariances  $\mathbf{B}_j^-$  are chosen in a way that attempts to make the distribution estimate more accurate.

The EnGMF formulas, used to construct the posterior GMM estimate, are

$$\begin{aligned} \mathbf{x}_j^\sim &= \mathbf{x}_j^- - \mathbf{G}_j (h(\mathbf{x}_j^-) - \mathbf{y}), \\ \mathbf{B}_j^+ &= (\mathbf{I} - \mathbf{G}_j \mathbf{H}_j^T) \mathbf{B}_j^-, \\ \mathbf{G}_j &= \mathbf{B}_j^- \mathbf{H}_j^T (\mathbf{H}_j \mathbf{B}_j^- \mathbf{H}_j^T + \mathbf{R}_i)^{-1}, \\ v_j &= \frac{\mathcal{N}(\mathbf{y} | h(\mathbf{x}_j^-), \mathbf{H}_j \mathbf{B}_j^- \mathbf{H}_j^T + \mathbf{R}_i)}{\sum_{j=1}^N \mathcal{N}(\mathbf{y} | h(\mathbf{x}_j^-), \mathbf{H}_j \mathbf{B}_j^- \mathbf{H}_j^T + \mathbf{R}_i)}, \\ \mathbf{H}_j &= \left. \frac{dh}{dx} \right|_{x=\mathbf{x}_j^-}, \end{aligned} \quad (5)$$

where  $\mathbf{x}_j^\sim$  are the means of the posterior GMM (but are not samples from the posterior GMM),  $\mathbf{B}_j^+$  are the posterior covariances,  $\mathbf{G}_j$  are the particle gain matrices,  $v_j$  are the weights of the posterior GMM modes, and  $\mathbf{H}_j$  are the linearizations of the observation operator eq. (1) about the ensemble members (particles).

Through the use of eq. (5), the weighted estimate of the posterior GMM can be constructed,

$$p^+(\mathbf{x}) = \sum_{j=1}^N v_j \mathcal{N}(\mathbf{x} | \mathbf{x}_j^\sim, \mathbf{B}_j^+), \quad (6)$$

In the final step of the EnGMF, an ensemble of  $N$  iid samples from the posterior GMM eq. (6) is computed, resulting in  $\mathbf{X}^+ = [\mathbf{x}_1^+, \dots, \mathbf{x}_N^+]$ . There are many ways by which a Gaussian mixture model can be sampled [6], [9]. In this work we focus on the naive approach. First sample  $N$  random variables from the distribution defined by the weights  $v_j$  from eq. (5). For each random variable, take its corresponding Gaussian mixture mode in eq. (6) and sample a random variable from it, by the canonical procedure. For details on implementation of the resampling procedure for Gaussian mixture models see [6].

The choice of covariance in the prior GMM eq. (4) dictates the convergence of the EnGMF algorithm. In the canonical EnGMF [5] the prior covariances are chosen to all be the

same scaled empirical covariance. The unbiased estimate of the empirical covariance at time index  $i$  is given by,

$$\mathbf{P}^- = \frac{1}{N-1} \mathbf{X} \left( \mathbf{I}_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^T \right) \mathbf{X}^T, \quad (7)$$

which in this work we refer to as the *global covariance* as it captures the global linear modes (in the eigenvector sense) of the data. The scaling of the covariances,

$$\mathbf{B}_j = \beta^2 \mathbf{P}^-, \quad 1 \leq j \leq N, \quad (8)$$

is performed by the choice of bandwidth factor  $\beta^2$ , which can be either determined empirically or through some prior knowledge about the behavior of the data.

If the prior distribution is Gaussian, the bandwidth factor,

$$\beta_{\text{sil}}^2 = \left( \frac{4}{N(n+2)} \right)^{\frac{2}{n+4}}, \quad (9)$$

known as Silverman’s rule of thumb [10], is optimal in terms of minimizing squared integral error. The bandwidth estimate eq. (9) is known to typically be too large in the case of highly non-Gaussian data [10], [11]. To account for this deficiency, in this work we choose to scale Silverman’s rule of thumb bandwidth by a constant scaling factor,

$$\beta^2 = s_\beta \beta_{\text{sil}}^2, \quad (10)$$

typically with  $0 < s_\beta \leq 1$ .

In the vast majority of interesting cases, things are non-linear and non-Gaussian, thus merely looking at the global linear estimate of the behavior of the data eq. (7) is likely insufficient. Additionally the mean and covariance updates in eq. (5) behave in a similar fashion to that of the extended Kalman filter. While this ensures that the EnGMF converges [8], in the case of a non-linear model and non-Gaussian distribution, a poor description of the prior would lead to an even worse description of the posterior.

Though the EnGMF contains the word “ensemble”, it is really a particle filter, though the distinction is not discussed further in this work.

### III. ESTIMATING LOCAL COVARIANCES

We now introduce the novel E-localization methodology. As stated previously, the global estimate of the covariance eq. (7) might not capture the local curvature of the distribution of interest, thus setting the covariance of each Gaussian mixture in the prior distribution estimate eq. (4) to have the same covariance eq. (8) can induce said distribution estimate to be of low accuracy.

To this end, we wish to find local covariance estimates for each particle,

$$\mathbf{B}_j = \beta^2 \mathbf{P}_j^-, \quad 1 \leq j \leq N, \quad (11)$$

where each  $\mathbf{P}_j^-$  is a unique estimate of the local curvature of the prior distribution. It is evident that each particle having its own covariance would enable a good estimate of the prior distribution eq. (4) that is much more accurate than through the use of the global covariance eq. (7) alone.

We now formalize our approach. Similar to other localization techniques [2], we make use of a decorrelation function to diminish the influence of far away particles on the local particle of interest. Given weights,

$$w_{j,k} = \frac{\ell(d(j,k)/r_j)}{\sum_{m=1}^N \ell(d(j,m)/r_j)}, \quad (12)$$

where  $d(j,k)$  is the distance between the  $j$ th and  $k$ th particles (in this work we use Euclidean distance, but this could be some other distance metric),  $r_j$  is the localization radius whose function is to scale the influence of the particles on each other, and  $\ell$  is a decorrelation function. We additionally assign the vector of all weights for the  $j$ th particle,

$$\mathbf{w}_j := [w_{j,1}, \dots, w_{j,N}]^T. \quad (13)$$

The localization radius,  $r_j$ , is a parameter that has to be carefully chosen. If the radius of influence is too large, the weights eq. (12) would effectively be uniform and would degenerate to the unlocalized case. If the radius of influence is too small, poor sampling would result in spurious modes being introduced to the covariance estimate. For localization in the geosciences, state variables typically do not change distance, but inter-particle distance changes at every step in the sequential filtering case, thus a constant choice of radius for each particle simply does not make sense.

In order to solve this newfound problem, we again take inspiration from Kernel density estimation literature [10] in determining the localization radius. For the  $j$ th particle, take the distance to the  $\sqrt{N}$ th (rounded) nearest neighbor according to the distance function  $d(j,k)$ , and call it  $d_j$ . We take the localization radius to be a scaling of this distance,

$$r_j = s_r d_j, \quad (14)$$

where  $s_r$  is the radius scaling factor.

Care must be taken when calculating the inter-particle distances in eq. (12), as a naive calculation can lead to an  $\mathcal{O}(N^2)$  cost, which is undesirable. In sequential filtering, it would be possible to keep track of a spatial index of particles and approximate the  $\sqrt{N}$ th nearest-neighbor in only  $\mathcal{O}(N)$  computations. This idea is not further explored in this work.

Note that as the ensemble size tends towards infinity,  $N \rightarrow \infty$ , the scaled radius eq. (14) tends towards zero, as the  $\sqrt{N}$ th nearest neighbor of any particle should tend closer and closer towards it if the region of support of the prior distribution is connected.

Similar to other localization techniques [2], we take the decorrelation function to be a Gaussian,

$$\ell(v) = e^{-\frac{1}{2}v^2} + \epsilon, \quad (15)$$

where we take  $\epsilon = 1e-4$  to be a factor that ensures no weights eq. (12) tend towards zero.

Given the weights vector  $\mathbf{w}_j$  defined by eq. (12) and eq. (13), the local sample covariance for the  $j$ th particle can be defined by,

$$\mathbf{P}_j^\sim = \frac{1}{1 - \sum_{k=1}^N w_{j,k}^2} \mathbf{X} (\boxtimes \mathbf{w}_j - \mathbf{w}_j \mathbf{w}_j^T) \mathbf{X}^T, \quad (16)$$

where the symbol  $\boxtimes$  is used to represent the matrix consisting of the subsequent on its diagonal and is zero otherwise.

The covariance estimate eq. (16) should tend towards a good representation of the local modes (in the eigenvector sense) of a given probability distribution, but is not necessarily scaled in a desirable fashion. Observe that as the scaling factor  $s_r$  in eq. (14) tends towards zero, the weights  $w_{j,k}$  in eq. (12) degenerate towards only one weight dominating, meaning that the local covariance estimate in eq. (16) tends towards the zero matrix. This is not desirable behavior. In order to fix this flaw, we enforce the local covariance estimates to have the same mean trace-generalized-variance as the global covariance estimate eq. (7) by linear scaling,

$$\mathbf{P}_j^- = \frac{\text{tr} \mathbf{P}^-}{\sum_{j=1}^N N^{-1} \text{tr} \mathbf{P}_j^\sim} \mathbf{P}_j^\sim, \quad (17)$$

which we refer to as the normalized local covariance estimate.

Note that the trace of a symmetric positive definite covariance estimate matrix can be computed by only using the ensemble anomalies without explicitly computing the matrix itself [12]. This is especially useful in the case when the ensemble size is significantly less than the state dimension  $N \ll n$ . Additionally, it is important to note for the practitioner that the computations of the local covariance estimates in eq. (16) are completely independent of each other, and thus can be performed in an embarrassingly parallel fashion.

Additionally, it is of note that the normalization correction in eq. (17) is purely heuristic. It is of independent interest to provide a more robust covariance correction, for instance one based on more rigorous kernel density estimation theory.

Each local normalized covariance in eq. (17) can then be used as the particle-localized covariance in eq. (11). This begets the E-localized approximation to the prior distribution in eq. (4). By utilizing this approximation to the prior in the EnGMF eq. (5), we beget the E-localized EnGMF, which in this work we refer to as the ELENGMF.

#### IV. ILLUSTRATION WITH A SIMPLE BIVARIATE DISTRIBUTION

We use a simple bivariate example to highlight the virtues of our proposed methodology.

We first construct a bivariate (in terms of two state variables  $\mathbf{x} = [x_1 \ x_2]^T$ ) Gaussian mixture model consisting of two constituent terms such that the local covariance of each constituent term in the mixture is significantly different than the global covariance of the total mixture. Each constituent term has the same local covariance,

$$\mathbf{C}_{\text{local}} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \quad (18)$$

where  $\rho$  is the correlation coefficient. We separate the two Gaussian terms in the  $x_2$  direction by a constant half-height  $v$ , meaning that the means of the two constituent terms are,

$$\mu_1 = [0 \ v]^T, \quad \mu_2 = [0 \ -v]^T. \quad (19)$$

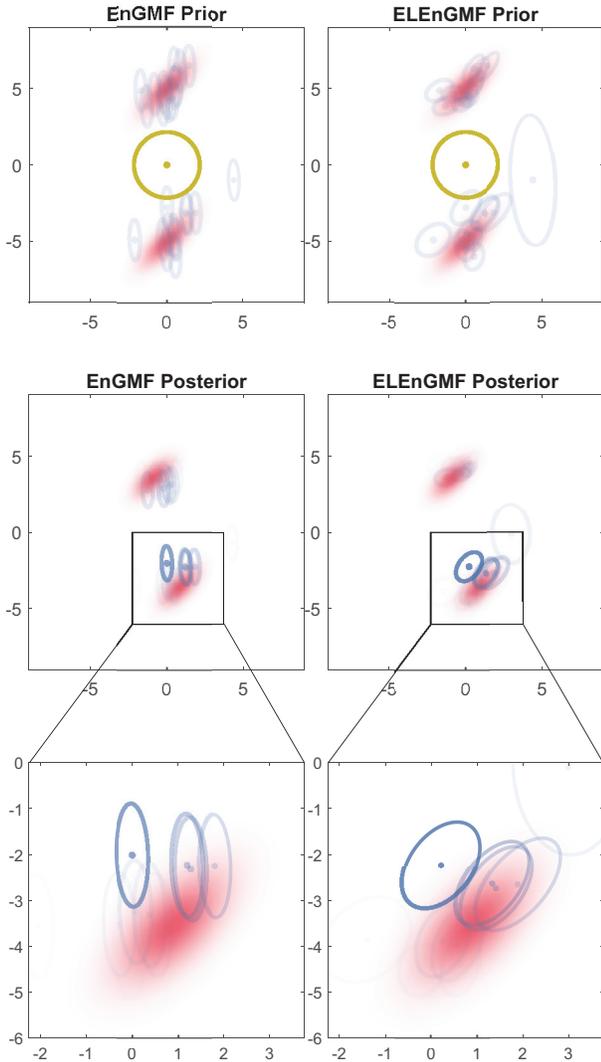


Fig. 1. A qualitative look at the differences between the EnGMF and the EEnGMF for a bivariate two mode Gaussian mixture. The top left panel shows the true prior distribution (red heat map), the observation with one standard deviation (yellow dot and circle), and the sample-based EnGMF Gaussian mixture means (blue dots) and one standard deviation (blue ellipses). The top right panel is similar to the previous, except the blue ellipses now represent the EEnGMF Gaussian mixture standard deviations. The middle left and right panels represent the EnGMF and EEnGMF posterior distributions respectively, with the opacity of the blue mixture modes representing their weights. The bottom two panels focus on the bottom mode of the true posterior to better showcase the differences in the covariances of the EEnGMF versus the EnGMF.

Explicitly, the Gaussian mixture distribution that we analyze is,

$$p(x) = \frac{1}{2} [\mathcal{N}(x|\mu_1, \mathbf{C}_{\text{local}}) + \mathcal{N}(x|\mu_2, \mathbf{C}_{\text{local}})], \quad (20)$$

where each term is equally weighted.

By a simple application of known formulas [13] the global covariance of eq. (20) is given by,

$$\mathbf{C}_{\text{global}} = \begin{bmatrix} 1 & \rho \\ \rho & 1 + v^2 \end{bmatrix}, \quad (21)$$

meaning that as we increase the separation half-height  $v$ , the second state variable starts to dominate the global covariance.

We can formalize this intuition by looking at the dominant eigenvectors of the covariances. When the correlation coefficient  $\rho$  is positive, the dominant eigenvector of the local covariance eq. (18) is,

$$\begin{bmatrix} 1 & 1 \end{bmatrix}^T, \quad (22)$$

which is not dependent on either the correlation coefficient  $\rho$  or the separation half-height  $v$ . The dominant eigenvector of the global covariance eq. (21), on the other hand, is,

$$\begin{bmatrix} \frac{-v^2 + \sqrt{v^4 + 4\rho^2}}{2\rho} & 1 \end{bmatrix}^T, \quad (23)$$

which tends towards an eigenvector of,

$$\begin{bmatrix} 0 & 1 \end{bmatrix}^T, \quad (24)$$

as the separation half-height  $v$  tends towards infinity.

This mismatch between the dominant modes eqs. (22) and (23) of the two covariances eqs. (18) and (21) is taken advantage of by our proposed E-localization technique.

With our newfound intuition, we are ready to perform filtering experiments. For our prior distribution we take the Gaussian mixture eq. (20) with correlation coefficient  $\rho = 0.75$  and vertical separation radius of  $v = 5$ . For our observation operator eq. (1), we take the identity function  $h = \text{id}$ , meaning that we observe both variables. For the observation error covariance, we take the scaled identity  $\mathbf{R} = 2\mathbf{I}_2$ . We assume that we received an observation exactly at the origin,  $[0 \ 0]^T$ , thus tending both modes towards it.

We first attempt to glean some qualitative information about the inner workings of the algorithms. Taking  $N = 25$  particles, the bandwidth scaling factor  $s_\beta = 0.1$ , and the E-localization radius scaling factor  $s_r = 1$ , fig. 1 looks at the efficacy of the EnGMF versus that of the EEnGMF. As can we observed the EnGMF utilizes the global covariance eq. (21) to attempt and match each of the faraway modes while the EEnGMF builds local estimates that more closely approximate the local covariance eq. (18). Of particular note is the outlier in the samples of the posterior to the right of the bottom mode. In the EnGMF, this mode is given the same weight as all other modes. In the EEnGMF, this mode's covariance is significantly increases as its nearest neighbor is comparatively far away. This likely means that the EEnGMF is much more likely to accurately describe long tail distributions.

We now take a more quantitative look at the efficacy of the EEnGMF versus that of the EnGMF. The KL divergence [14] of a distribution described by data,  $p$ , from a distribution describing the truth,  $q$ , can be written as the following expectation,

$$D_{\text{KL}}(p \parallel q) = \mathbb{E}_{x \sim p} \left[ \log \left( \frac{p(x)}{q(x)} \right) \right], \quad (25)$$

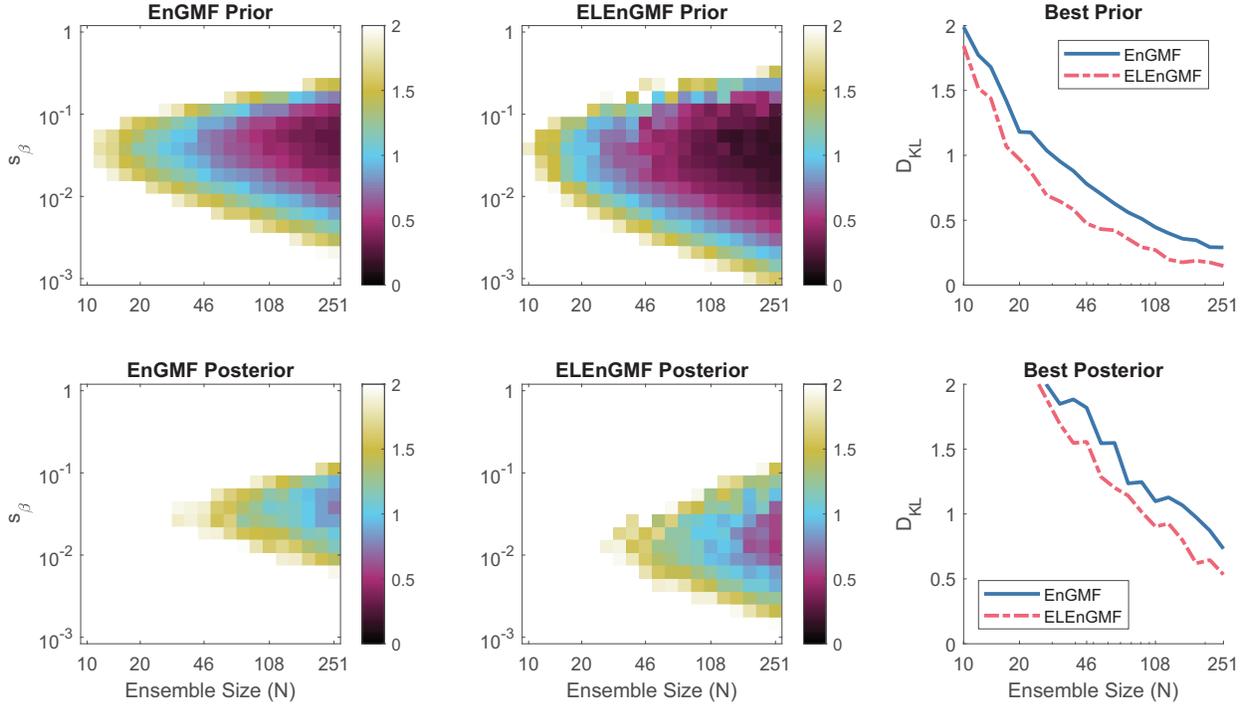


Fig. 2. KL divergence experiment for the simple bivariate distribution. The left four panels are all heatmaps with the ensemble size ( $N$ ) in log-scale on the x-axis, the bandwidth scaling factor  $s_\beta$  from eq. (10) in log-scale on the y-axis, and the colors representing the KL divergence of the GMM expressed therein from the truth. Of the left four, the top left panel represents the KL divergence of the prior GMM of the EnGMF from the true prior, the bottom left panel represents the corresponding posterior GMM KL divergence from the true posterior, the top middle panel represents the ELeNGMF prior GMM KL divergence from the true prior, and the bottom middle represents the KL divergence of the ELeNGMF posterior GMM KL divergence from the true posterior. The right two panels look at the best case scenario for the bandwidth scaling factor  $s_\beta$ , taking the minimal KL divergence for each ensemble size  $N$  for the EnGMF and ELeNGMF. The top right panel represents the best case scenario for the prior GMMs and the bottom right panel represents the best case scenario for the posterior GMMs.

which can be approximated by the following numerically stable, but biased, formula,

$$D_{\text{KL}}(p \parallel q) \approx \frac{1}{M} \sum_{x_i \sim p, i=1}^M \frac{1}{2} (\log q(x) - \log p(x))^2, \quad (26)$$

where  $M$  is a finite number of samples from the candidate posterior. The formulation eq. (26) is taken from [15]. We use the KL divergence to measure the error induced by the GMM estimates of the prior and the posterior versus the truth.

Taking a fixed radius scaling factor eq. (14) of  $s_r = 1$ , and varying both the Silverman bandwidth scaling factor eq. (10)  $s_\beta$  and ensemble size  $N$ , we can compute the KL divergence eq. (25) of the prior and posterior distributions for both the EnGMF and the ELeNGMF. Averaging over 250 samples of possible prior, and using  $M = 25$  quasi-samples to approximate the KL divergence eq. (26), we obtain the KL divergence experiment results which can be seen in fig. 2. Of note is the fact that the ELeNGMF always performs strictly better than the EnGMF, even when both are given their optimal bandwidth scaling factor  $s_\beta$  from eq. (10). The ELeNGMF also remains stable for a larger range of the scaling factor for both the prior and posterior distributions, with the difference in the posterior of particular interest where the difference in the EnGMF and ELeNGMF GMM estimates becomes more pronounced.

Of interest is the fact that a good bandwidth scaling factor for the prior does not necessarily correspond to a good bandwidth scaling factor for the posterior. In fact, while there seems to be a correlation between a good representation of the prior and a good representation of the posterior in the EnGMF and the ELeNGMF algorithms, it is not a direct relationship.

## V. SEQUENTIAL FILTERING WITH THE LORENZ '63 EQUATIONS

The final round of experiments aims to put the ELeNGMF to practical use in a sequential filtering experiment. In sequential filtering our aim is to find an optimal estimate of the truth  $\mathbf{x}^t$  at some time index  $i$ . This is performed by ‘forecasting’ a state estimate to time index  $i$  to obtain a prior and performing inference eq. (3) to obtain a posterior estimate of the state  $\mathbf{x}_i^+$ , then repeating the cycle over and over again *ad infinitum* (or until a desired time).

To this end we use the Lorenz '63 equations [16],

$$\begin{aligned} \dot{x}_1 &= 10(x_2 - x_1), \\ \dot{x}_2 &= x_1(28 - x_3) - x_2, \\ \dot{x}_3 &= x_1x_2 - \frac{8}{3}x_3, \end{aligned} \quad (27)$$

which is one of the foundational problems for particle filters [3].

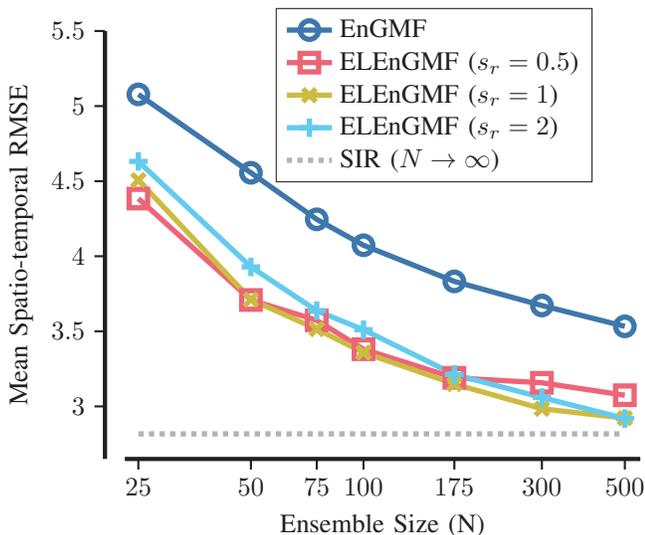


Fig. 3. Sequential filtering results for the Lorenz '63 equations. The x-axis represents the ensemble size  $N$  in log-scale, and the y-axis represents the mean spatio-temporal RMSE eq. (29) of the analysis mean generated by each algorithm. All ensemble Gaussian mixture variants tested herein use the same bandwidth scaling factor of  $s_\beta = 1$ . The dark blue line with circular markers represents the canonical EnGMF, the red line with square markers represents the ELENGMF with a radius scaling factor of  $s_r = 0.5$ , the yellow line with x markers represents the ELENGMF with a radius scaling factor of  $s_r = 1$ , and the sky blue line with plus markers represents the ELENGMF with a radius scaling factor of  $s_r = 2$ . The sequential importance sampling filter is used to compute the best-case spatio-temporal RMSE in the limit of ensemble size  $N \rightarrow \infty$  that should be close to true Bayesian inference, and is represented by a constant dashed grey line at the bottom of the graph.

The following problem setup is taken from [8]. We perform inference every  $\Delta t = 0.5$  time, with 5500 sequential steps taken (the first 500 of which are discarded to account for spinup—discarding errors from initial over- or under-confidence of the algorithm). We take the scalar range observation,

$$h(\mathbf{x}) = \|\mathbf{x} - \mathbf{c}_2\|_2, \quad (28)$$

where  $\mathbf{c}_2 = [6\sqrt{2} \ 6\sqrt{2} \ 27]^T$  is the center of one of the wings of the Lorenz butterfly, with unbiased Gaussian observation error determined by the variance  $\mathbf{R} = 1$ . For this experiment we take the spatio-temporal RMSE of the posterior means of the data with respect to the truth,

$$\text{RMSE}(\mathbf{x}^+) = \sqrt{\frac{1}{|\mathbf{x}|} \sum_{i,j} x_{i,j}^+ - x_{i,j}^t} \quad (29)$$

where  $\mathbf{x}^+$  is the temporal collection of posterior means,  $\mathbf{x}^t$  is the collection of true states of the system,  $i$  is the time index,  $j$  is the state index, and  $|\mathbf{x}|$  is the cardinality of the data. In the following experiments the mean of the RMSE eq. (29) is taken over four independent sequential filtering runs.

For the EnGMF and the ELENGMF we fix the bandwidth scaling factor eq. (10) to  $s_\beta = 1$  assuming a naive Gaussian optimal bandwidth factor on the prior—which is known to be suboptimal. For the ELENGMF we vary the radius scaling factor eq. (14) between  $s_r = 0.5, 1, 2$  in order to observe the

differences in the RMSE at small ensemble sized. Finally we vary the ensemble size from  $N = 25$  to  $N = 500$  to observe both small ensemble and asymptotic behavior. We additionally compute a bootstrap sequential importance resampling filter for  $N = 10000$  particles to attempt to estimate the lowest possible error under the assumptions of conventional Bayesian inference.

The results of this experiment can be seen in fig. 3. Of note is the fact that for  $N = 500$  particles all the ELENGMF variants effectively converge to performing true Bayesian inference, while the EnGMF seems to require significantly more particles. While the influence of the covariance on the algorithm is diminished by the fact that bandwidth parameter eq. (9) tends towards zero [8], the Silverman bandwidth assumption still severely hampers the naive EnGMF. The ELENGMF significantly outperforms the EnGMF for the same error, requiring from two to three times less ensemble members for the same level of accuracy. Additionally, the ELENGMF is fairly robust to changes in the radius scaling factor eq. (14), and had very little noticeable difference for all possible factors tested.

## VI. CONCLUSIONS

In this work we have introduced the E-localization methodology and have applied it to the ensemble Gaussian mixture filter (EnGMF), creating the E-localized ensemble Gaussian mixture filter (ELENGMF). We have shown through both qualitative and quantitative means that this methodology addresses one issue—the disparity between local and global notions of covariance—that arises in the application of the EnGMF to the case of a small ensemble size. Additionally, through the application of the methodology to the Lorenz '63 equations, we have shown that the ELENGMF has the potential to be a superior particle filter to the EnGMF in the sequential filtering setup.

While this work has tackled some issues inherent to the EnGMF, it has left many as open problems, and has even introduced some new issues. One new issue is the selection of localization radius  $r_j$  in eq. (12). The methodology presented in this work for tackling this issue clearly has nice empirical behavior, but its justification is purely heuristic. Another related issue is that of the choice of decorrelation function eq. (15). Exploring this choice in future work is desirable. Another more serious issue is the lack of theory for the E-localization procedure. Significant theoretical work has to be performed before E-localization can be applied in a more practical setting.

Future non-theoretic work would involve applying adaptive covariance parameterization techniques [8] to the ELENGMF, such that the choice of bandwidth scaling factor eq. (10) and radius scaling factor eq. (14) could be performed adaptively in the sequential filtering regime. Another future direction would involve applying the ELENGMF to a practical orbit determination problem [7].

## REFERENCES

- [1] W. R. Tobler, "A computer movie simulating urban growth in the detroit region," *Economic geography*, vol. 46, no. sup1, pp. 234–240, 1970.
- [2] M. Asch, M. Bocquet, and M. Nodet, *Data assimilation: methods, algorithms, and applications*. SIAM, 2016.
- [3] S. Reich and C. Cotter, *Probabilistic forecasting and Bayesian data assimilation*. Cambridge University Press, 2015.
- [4] S. Reich, "A nonparametric ensemble transform method for Bayesian inference," *SIAM Journal on Scientific Computing*, vol. 35, no. 4, pp. A2013–A2024, 2013.
- [5] J. L. Anderson and S. L. Anderson, "A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts," *Monthly weather review*, vol. 127, no. 12, pp. 2741–2758, 1999.
- [6] B. Liu, B. Ait-El-Fquih, and I. Hoteit, "Efficient kernel-based ensemble Gaussian mixture filtering," *Monthly Weather Review*, vol. 144, no. 2, pp. 781–800, 2016.
- [7] S. Yun, R. Zanetti, and B. A. Jones, "Kernel-based ensemble Gaussian mixture filtering for orbit determination with sparse data," *Advances in Space Research*, vol. 69, no. 12, pp. 4179–4197, 2022.
- [8] A. A. Popov and R. Zanetti, "An adaptive covariance parameterization technique for the ensemble Gaussian mixture filter," *arXiv preprint arXiv:2212.10323*, 2022.
- [9] I. Gilitschenski, J. Steinbring, U. D. Hanebeck, and M. Simandl, "Deterministic Dirac mixture approximation of Gaussian mixtures," in *17th International Conference on Information Fusion (FUSION)*. IEEE, 2014, pp. 1–7.
- [10] B. W. Silverman, *Density estimation for statistics and data analysis*. Routledge, 2018.
- [11] P. Janssen, J. S. Marron, N. Veraverbeke, and W. Sarle, "Scale measures for bandwidth selection," *Journal of Nonparametric Statistics*, vol. 5, no. 4, pp. 359–380, 1995. [Online]. Available: <https://doi.org/10.1080/10485259508832654>
- [12] A. A. Popov, A. Sandu, E. D. Nino-Ruiz, and G. Evensen, "A stochastic covariance shrinkage approach in ensemble transform Kalman filtering," *arXiv preprint arXiv:2003.00354*, 2020.
- [13] M. H. (<https://math.stackexchange.com/users/11667/michael-hardy>), "Calculation of the covariance of Gaussian mixtures," Mathematics Stack Exchange, <https://math.stackexchange.com/q/195984> (version: 2019-09-15). [Online]. Available: <https://math.stackexchange.com/q/195984>
- [14] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.
- [15] J. Schulman, "Approximating KL divergence," 2020. [Online]. Available: <http://joschu.net/blog/kl-approx.html>
- [16] E. N. Lorenz, "Deterministic nonperiodic flow," *Journal of atmospheric sciences*, vol. 20, no. 2, pp. 130–141, 1963.