

RESEARCH ARTICLE

Regularization and tempering for a moment-matching localized particle filter

Jonathan Poterjoy 

Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD

Correspondence

Jonathan Poterjoy, Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD 20742, USA.

Email: poterjoy@umd.edu

Funding information

National Oceanic and Atmospheric Administration, Grant/Award Number: NA20OAR4600281; National Science Foundation, United States, Grant/Award Number: AGS1848363

Abstract

Iterative ensemble filters and smoothers are now commonly used for geophysical models. Some of these methods rely on a factorization of the observation likelihood function to sample from a posterior density through a set of “tempered” transitions to ensemble members. For Gaussian-based data assimilation methods, tangent linear versions of nonlinear operators can be relinearized between iterations, thus leading to a solution that is less biased than a single-step approach. This study adopts similar iterative strategies for a localized particle filter (PF) that relies on the estimation of moments to adjust unobserved variables based on importance weights. This approach builds off a “regularization” of the local PF, which forces weights to be more uniform through heuristic means. The regularization then leads to an adaptive tempering, which can also be combined with filter updates from parametric methods, such as ensemble Kalman filters. The role of iterations is analyzed by deriving the localized posterior probability density assumed by current local PF formulations and then examining how single-step and tempered PFs sample from this density. From experiments performed with a low-dimensional nonlinear system, the iterative and hybrid strategies show the largest benefits in observation-sparse regimes, where only a few particles contain high likelihoods and prior errors are non-Gaussian. This regime mimics specific applications in numerical weather prediction, where small ensemble sizes, unresolved model error, and highly nonlinear dynamics lead to prior uncertainty that is larger than measurement uncertainty.

KEYWORDS

data assimilation, ensemble Kalman filters, localization, particle filters

1 | INTRODUCTION

Particle filters (PFs) are sequential Monte Carlo methods that can solve data assimilation problems characterized by non-Gaussian error distributions for prior model variables

or measurements (Doucet *et al.*, 2001). From a geoscience perspective, PFs contain several theoretical properties that make them attractive for research and environmental prediction. Namely, they preserve dynamical balances during data assimilation update steps; they require no

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Author. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

special treatment for nonlinear measurement operators or non-Gaussian errors; and they provide an elegant solution to the underlying Bayesian filtering problem. Recent efforts applying PFs for geophysical models have resulted in “localized” PFs, which approximate a given data assimilation application as a large set of loosely coupled problems that can be solved independently using relatively small ensembles—an approach long used for ensemble Kalman filters (EnKFs). These efforts resulted in a large variety of filtering and smoothing methods, including those introduced by Bengtsson *et al.* (2003), Poterjoy (2016), Penny and Miyoshi (2016), Poterjoy and Anderson (2016), Lee and Majda (2016), Robert and Künsch (2017), Chustagulprom *et al.* (2016), and Morzfeld *et al.* (2018); see Van Leeuwen *et al.* (2019) for a more exhaustive list of PF approaches designed for high-dimensional applications. Though localization delivers a potentially transformative strategy for implementing PFs for high-dimensional systems, this technique alone is often inadequate for real geophysical applications. For example, the conditions for PF weight collapse identified in past studies (e.g., Bengtsson *et al.*, 2008; Bickel *et al.*, 2008; Snyder *et al.*, 2008) still hold within the neighborhood of observations for localized PFs. Therefore, filter degeneracy is still inevitable for dynamical systems that are observed by many accurate, collocated measurements, or when observation-space prior statistics are heavily biased.

Data assimilation problems in atmospheric science that are hypothesized to benefit from PF methodology—because of strong nonlinearity in model dynamics and measurement operators—also tend to be characterized by the problems already listed herein. For example, data assimilation for convective weather regimes introduces many challenges for Gaussian-based filters, owing to the non-Gaussian errors needed to quantify model and measurement uncertainty and nonlinearity in the underlying system dynamics and observing systems (Posselt *et al.*, 2014; Posselt, 2016). At the same time, prior and posterior statistics are often mischaracterized, owing to the large computational cost of generating many high-resolution ensemble members or the presence of unknown error sources in physical parametrization schemes for turbulence, cloud microphysics, land-surface processes, and so on. Data assimilation for tropical cyclones presents a similar set of challenges, except the most abundant measurements come from radiometers onboard satellites. In general, these characteristics present difficulties for numerous research and prediction problems, including many outside of weather forecasting.

This study introduces new methodology for the local PF described in Poterjoy (2016) and revised in Poterjoy *et al.* (2019), which targets the type of data assimilation problems just discussed. The new approaches rely on a

factorization of particle weights to draw samples from the localized posterior density through successive iterations. It borrows from past studies that introduce iterations for EnKFs and smoothers to cope with mildly nonlinear model dynamics and measurement operators (Emerick and Reynolds, 2012; 2013). In the context of PFs, these iterations serve a different purpose. Each iteration uses a regularization coefficient on particle weights to maintain a threshold effective ensemble size. For filters that update particles to match the first two posterior moments (e.g., Feng *et al.*, 2020) accurate particle updates can be achieved without higher-order statistics. In addition to showing significant benefits over non-iterative formulations of the local PF, the method presents a natural framework for extending the local PF into a hybrid PF–EnKF, as proposed by past studies (Frei and Künsch, 2013; Chustagulprom *et al.*, 2016; Robert and Künsch, 2017; Grooms and Robinson, 2021).

This article is organized in the following manner. Section 2 discusses the mathematical framework for localized PFs, including a derivation of the posterior probability density function (pdf) assumed by the Poterjoy *et al.* (2019) local PF. Calculations of this pdf are needed to illustrate the behavior of the local PF with and without iterations. Section 3 discusses the collapse of localized PFs to provide motivation for regularized, iterative, and hybrid approaches, which are discussed in Section 4. Section 5 shows numerical experiments performed using a bivariate problem and the 40-variable dynamical system of Lorenz (1996). The last section discusses major findings from this study.

2 | LOCALIZED PARTICLE FILTERS

This section introduces a subset of localized PFs, which behave similarly to the filter introduced in Poterjoy (2016). The primary goal of this section is to provide context for the new filtering methodology introduced in Section 4, which requires a clear definition of the posterior density certain localized PFs attempt to sample from. We direct readers to Farchi and Bocquet (2018) for a comprehensive review of strategies recently adopted for this purpose.

For geophysical data assimilation, the state vector representing variables for a dynamical system is treated as a random variable $\mathbf{x} \in \mathbb{R}^{N_x}$ with the underlying goal of estimating various properties of the probability density $p(\mathbf{x})$. The time propagation of \mathbf{x} is approximated through a prediction model $\mathbf{x}_{t+1} = M(\mathbf{x}_t) + \eta_t$, where η_t is an additive model error. The probability density for \mathbf{x} is updated periodically to reflect new information from incomplete and noisy measurements $\mathbf{y} \in \mathbb{R}^{N_y}$, where the mapping between

model variables and observations is given by $\mathbf{y} = H[\mathbf{x}] + \epsilon$. The error ϵ is the combined error from instruments on observing systems, representativeness error, and imperfect estimates of $H()$ itself. When estimating the probabilistic evolution of \mathbf{x} , we are often interested in the filtering density $p(\mathbf{x}_t | \mathbf{y}_{0:t})$, which is the conditional probability of \mathbf{x} at the current time given all current and past measurements. By drawing samples from this density and passing them through $M()$, we can estimate various parts of $p(\mathbf{x}_{t+\tau} | \mathbf{y}_{0:t})$ for some forecast time $\tau > 0$. This framework is used to formulate sequential Monte Carlo filters, such as EnKFs and PFs, thus providing the cornerstone of probabilistic environment prediction systems such as those used for weather forecasting (e.g., Houtekamer and Zhang, 2016).

2.1 | The localized posterior probability density

The local PF attempts to sample from a posterior that assumes a gradual decoupling of marginal state variables displaced spatially, similar to EnKFs used for high-dimensional geophysical applications. Derivations of this probability density are avoided in past studies, primarily because it is not needed to draw samples from approximate forms of this density. Given the nature of this study, we provide a transparent examination of the localized probability density.

Consider the case where we are given a prior density $p(\mathbf{x}_t)$ and need to approximate the conditional density $p(\mathbf{x}_t | \mathbf{y}_t)$. For simplicity, ignore time indices and consider the two-dimensional problem: $\mathbf{x} = [u \ v]^T$, where v is observed directly with measurement y . The sequential Monte Carlo framework requires that we sample from the conditional probability distribution $p(\mathbf{x} | y)$. One way of conceptualizing this problem is to consider the factorization

$$\begin{aligned} p(\mathbf{x} | y) &= p(u, v | y), \\ &= p(u | v, y) p(v | y), \\ &= p(u | v) p(v | y), \end{aligned} \quad (1)$$

where $p(u | v) = p(u | v, y)$ follows from the Markovian memoryless assumption for the observation process. In this form, we can sample from $p(\mathbf{x} | y)$ by drawing from $p(v | y)$ and using the result to sample from $p(u | v)$. For data assimilation purposes, these steps can also be envisioned as a particle ‘‘update’’ for v (or observation-space variable) followed by an update to unobserved model variables. For example, an ensemble square-root filter can be formulated to perform linear regression to update u in a manner that is consistent with changes to v (Anderson, 2003). This step reflects the deterministic coupling between Gaussian random variables; see Reich and Cotter (2015, chapter 2) for a

review. The coupling is not unique for the Gaussian case, which is why multiple types of ensemble square-root filters exist.

Multiple filters that are widely used in public community software packages adopt the serial algorithm described in Anderson and Collins (2007), which exploits the factorization described earlier herein. Examples include the NCAR Data Assimilation Research Testbed (Anderson *et al.*, 2009) and the NOAA Gridpoint Statistical Interpolation (Shao *et al.*, 2016) system used for operational weather prediction in the United States. By construction, the Poterjoy (2016) local PF follows the same algorithmic framework, so it can be implemented easily for a broad selection of geophysical models. Therefore, it is important to examine how the filter solves the bivariate data assimilation problem introduced in this section.

First, consider the delta function approximation of the prior pdf made by PFs:

$$\begin{aligned} p(\mathbf{x}) &\approx \frac{1}{N_e} \sum_{n=1}^{N_e} \delta(\mathbf{x} - \mathbf{x}^n), \\ &= \frac{1}{N_e} \sum_{m=1}^{N_e} \sum_{n=1}^{N_e} \delta(u - u^m) \delta(v - v^n) \delta_{mn}, \end{aligned} \quad (2)$$

where $p(\mathbf{x})$ is expressed as a sum of products between Dirac delta functions for variables u and v , and δ_{mn} is a Kronecker delta function. We can then write the conditional probability density for \mathbf{x} given y using Bayes’ theorem:

$$\begin{aligned} p(\mathbf{x} | y) &= p(u, v | y), \\ &= p(u | v) p(v | y), \\ &\propto p(u | v) p(v) p(y | v), \\ &\approx \sum_{m=1}^{N_e} \sum_{n=1}^{N_e} w^n \delta(u - u^m) \delta(v - v^n) \delta_{mn}, \end{aligned} \quad (3)$$

where $w^n = p(y | v^n) / \sum_{k=1}^{N_e} p(y | v^k)$. We will denote the resulting approximation of the posterior density as $p_{\text{pf}}(u, v | y)$, since it is the common form used for PFs.

For many geophysical problems, it is practical to assume state variables displaced by a critical physical distance have independent errors. For example, we can write

$$\begin{aligned} p(u, v | y) &\approx p_i(u, v | y), \\ &= p(u | v) p(v | y), \\ &= p(u) p(v | y), \\ &= \sum_{m=1}^{N_e} \frac{1}{N_e} \delta(u - u^m) \sum_{n=1}^{N_e} w^n \delta(v - v^n), \\ &= \frac{1}{N_e} \sum_{m=1}^{N_e} \sum_{n=1}^{N_e} w^n \delta(u - u^m) \delta(v - v^n), \end{aligned} \quad (4)$$

where $p_i(\cdot)$ denotes an independent error approximation for all variables in \mathbf{x} . The approximate posterior, $p_i(u, v|y)$, can be factored in the same manner as $p_{\text{pf}}(u, v|y)$ (see Equation (3)) to show that the assumed conditional density for the pair of variables is $p(u^m|v^n) = 1/N_e$ for all combinations of n and m , and zero otherwise, which extends the number of solutions with non-zero probability from N_e to $N_e^{N_x=2}$. This approximation greatly reduces the sample size needed to apply the bootstrap PF for dynamical systems characterized by large sets of loosely coupled systems; for example, see Poterjoy, (2016, figure 3). Lee and Majda (2016) adopt a similar strategy to form a “block” localization of the bootstrap PF, which applies Equation (3) for state variables in the vicinity of observations in state space and Equation (4) for all remaining variables.

For data assimilation applications targeted by this research, such as weather prediction, the decoupling must be specified carefully to maintain true cross-variable error dependencies that come from physical processes depicted by the dynamical model. Not doing so can result in artificial dynamical imbalances during model integration (Kepert, 2009; Greybush *et al.*, 2011), which reduce predictive skill. This challenge is complicated by the realization that spatial structure of error correlations tends to be anisotropic and dependent on the underlying flow (e.g., Poterjoy and Zhang, 2011). Furthermore, errors do not often exhibit abrupt spatial changes that would permit the use of Equation (4) alone for applications of this type. As is true for localized EnKFs, large correlations will inevitably need to be modulated by localization to reduce sampling error. Under these circumstances, localization can introduce a bias in the resulting posterior estimate, but the bias is generally smaller than what would be obtained without localization (for small N_e), which is why localization is useful for many applications. Therefore, an appropriate decoupling should reflect a combination of $p_{\text{pf}}(u, v|y)$ and $p_i(u, v|y)$. For example, we can choose a conditional probability density of the form $p(u^m|v^n) = \delta_{mn}\rho_{u,v} + (1/N_e)(1 - \rho_{u,v})$, where $\rho_{u,v}$ is a “localization” coefficient controlling a coupling between the two variables. Note that this choice resorts back to the two limiting cases when $\rho_{u,v} = 0$ and $\rho_{u,v} = 1$. The resulting posterior density for u and v is then approximated using

$$\begin{aligned} p(u, v|y) &\approx p_i(u, v|y), \\ &= \sum_{m=1}^{N_e} \sum_{n=1}^{N_e} w^n \left[\delta_{mn}\rho_{u,v} + \frac{1}{N_e}(1 - \rho_{u,v}) \right] \\ &\quad \times \delta(u - u^m)\delta(v - v^n), \end{aligned} \quad (5)$$

where $p_i(\cdot)$ indicates the localized posterior pdf. For geophysical applications, ρ is typically modeled by an empirical correlation function, which depends only on distance

in physical space; see Lei and Anderson (2014) for a review. This function can be chosen to satisfy a number of properties, such as smoothness and compactness (Gaspari and Cohn, 1999). The latter quality leads to a straightforward parallelization of filters that process observations serially, such as those in the NCAR Data Assimilation Research Testbed and NOAA Gridpoint Statistical Interpolation (Anderson *et al.*, 2009; Shao *et al.*, 2016).

The choice for $p_i(u, v|y)$ introduced in Equation (5) leads to

$$\begin{aligned} p(u|y) &\approx \int_{\mathbb{R}} p_i(u, v|y) dv, \\ &= \sum_{m=1}^{N_e} \sum_{n=1}^{N_e} w^n \left[\delta_{mn}\rho_{u,v} + \frac{1}{N_e}(1 - \rho_{u,v}) \right] \delta(u - u^m), \\ &= \sum_{m=1}^{N_e} \sum_{n=1}^{N_e} w^n \delta_{mn}\rho_{u,v}\delta(u - u^m) \\ &\quad + \sum_{m=1}^{N_e} \sum_{n=1}^{N_e} \frac{w^n}{N_e}(1 - \rho_{u,v})\delta(u - u^m), \\ &= \sum_{m=1}^{N_e} w^m \rho_{u,v}\delta(u - u^m) \\ &\quad + \sum_{m=1}^{N_e} \frac{1}{N_e}(1 - \rho_{u,v})\delta(u - u^m), \\ &= \sum_{m=1}^{N_e} \left[\left(w^m - \frac{1}{N_e} \right) \rho_{u,v} + \frac{1}{N_e} \right] \delta(u - u^m), \end{aligned} \quad (6)$$

which is identical to the marginal posterior density used for the local PF in Poterjoy *et al.* (2019).

To extend the bivariate example to multiple state variables and observations, the Poterjoy (2016) and Poterjoy *et al.* (2019) filters assimilate all observations at a given time serially and use a joint observation–model space localization that is identical to the approach discussed in Anderson and Collins (2007). This strategy assumes observations have independent errors and well-defined physical locations; the latter assumption allows observation-space priors to be treated as augmented state variables and updated alongside all other variables in \mathbf{x} . Appendix A discusses potential ways of extending Equation (5) for $N_y > 1$ and $N_x > 2$.

PFs designed to operate within pre-existing local ensemble transform Kalman filter frameworks (e.g., Penny and Miyoshi, 2016; Potthast *et al.*, 2019) adopt a sliding window localization identical to Hunt *et al.* (2007). This strategy is also effective at providing the desired spatial decoupling, but results in slightly different formulations for the conditional distributions specified earlier herein. The two methods have similar properties for bivariate applications, which will be explored through simple examples in the following sections. We also note

that joint observation–model space localization and sliding window localization are suboptimal for assimilating non-local observations, such as satellite radiance. Readers are encouraged to review Lei and Anderson (2014) for a discussion on the positive and negative aspects of different localization methodologies.

2.2 | Benefits and drawbacks of the localized posterior

We can visualize how PFs approximate $p_1(\mathbf{x}|y)$ for a bivariate application to demonstrate the influence of ρ on the assumed posterior pdf. Consider the example shown in Figure 1a, where 10 particles are sampled from $p(u, v)$ and reweighted based on $p(y|u, v)$, where y measures v directly. The value of each sampled particle is indicated in the u – v plane shown in Figure 1 by black circles. The red markers in this figure indicate particle weights, which are proportional to the posterior joint probability of u and v , with larger markers corresponding to a larger weight. Each axis also shows blue markers, which indicate the marginal posterior probability of both u and v .

Figure 1a shows the sample-estimated $p_1(u, v|y)$ when $\rho_{u,v} = 1$, which is equivalent to $p_{\text{pf}}(u, v|y)$. Note that the marginal posterior probabilities for u and v are the same for particles with like indices—an assumption that becomes relaxed for $\rho_{u,v} < 1$. Figure 1b,c shows how the sample-estimated pdf changes as $\rho_{u,v}$ decreases to 0.8 and 0.0. The marginal weights for the observed variable v remain the same, but they become equal to $1/N_e$ for u to reflect the decoupling. As shown in Figure 1, localization

leads to non-zero probability for all combinations of sampled u and v values while reducing the joint probability of solutions with matching indices. As a result, the number of plausible model states increases from 10 to 100. The new solutions that arise from localization are linear combinations of the original particles. The resulting pdf is less likely to collapse onto the original samples (marked by circles), thus partially solving the filter degeneracy issues discussed in the previous section. At the same time, there is no guarantee that samples drawn from parts of the pdf that represent combinations of particles will satisfy the model equations used to produce particles (Van Leeuwen, 2009; Farchi and Bocquet, 2018). Most notably, forming combinations of particles comprised of geophysical quantities, such as wind, temperature, and pressure, will ultimately yield large discontinuities between grid points, regardless of whether ρ is modeled using a spatially smooth function. Though this problem is inevitable for localized PFs, strategies exist for reducing discontinuities; some are discussed in the following subsection.

2.3 | Sampling from the localized posterior

Drawing samples from $p_1(\mathbf{x}|y)$ becomes non-trivial for even moderately sized problems, which is why several forms of localized PFs have been proposed in recent years. In this section, we discuss two approaches that have been used for high-dimensional geophysical applications. Before discussing these strategies, it is important to note that methodology already exists for drawing

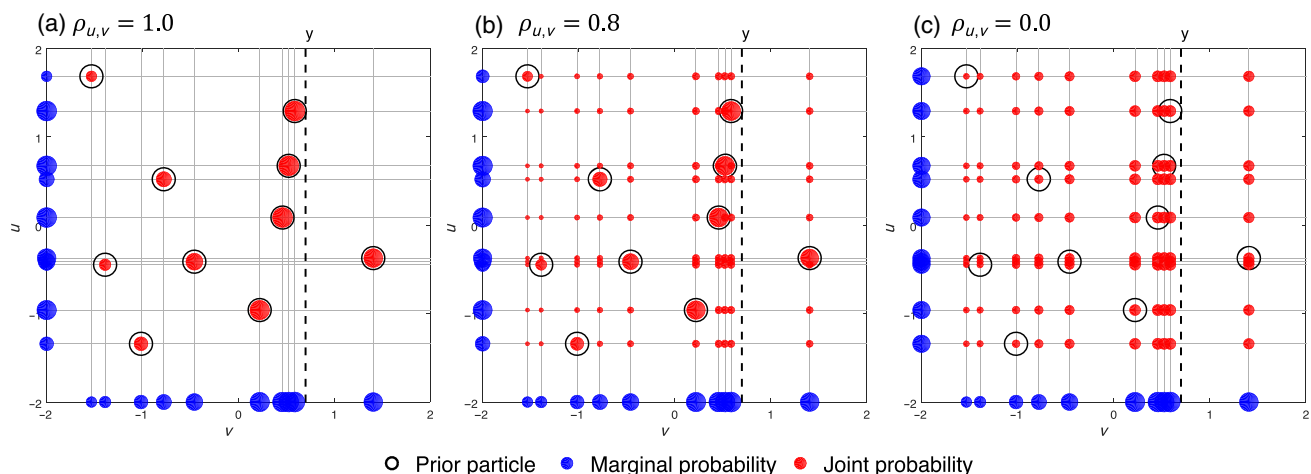


FIGURE 1 A bivariate demonstration of how localization modulates estimates of (blue) marginal and (red) joint probabilities from (black circles) prior particles ($N_e = 10$). Light gray lines indicate the sample value for variables u and v , and the black dashed line shows the value of an observation. Results are shown for (a) $\rho_{u,v} = 1$, (b) $\rho_{u,v} = 0.8$, and (c) $\rho_{u,v} = 0$ [Colour figure can be viewed at wileyonlinelibrary.com]

particles from $p_1(u, v|y)$ via a set of sequential sampling steps for each variable (Metref *et al.*, 2014). This approach can be expanded for $N_x > 2$ and $N_y > 1$, but it is computationally prohibitive for high dimensions. Though spatial discontinuities still remain following this sampling, the dependence of marginal samples on neighboring marginal samples should yield the smallest possible discontinuities.

To sample from $p_1(\mathbf{x}|y)$, Penny and Miyoshi (2016) form a transform matrix for each variable in \mathbf{x} , which resembles an independent sampling at each physical grid point for models with a spatial dimension. Starting with a transform matrix populated by binary coefficients, they apply a smoothing operator that turns the coefficients into continuous values between 0 and 1, thus reducing discontinuities in posterior samples. To achieve stable results for weather applications, Robert *et al.* (2018) and Potthast *et al.* (2019) apply a similar form of localized PF update, but on a coarse model grid, which is then extrapolated to the native domain after data assimilation, thus reducing discontinuities in the updated particles.

Poterjoy (2016) propose a localized PF that uses the strategy outlined in Anderson and Collins (2007) for performing joint observation- and model-space updates in parallel. This approach processes observations serially to perform an observation-space update via bootstrap sampling of particles, followed by a model-space update that preserves the non-localized PF solution where $\rho = 1$ and the prior solution when $\rho = 0$. By performing the observation-space update first, posterior particles can be adjusted in a manner that is consistent with resampled particles in the vicinity of observations. For the remaining variables, the filter uses an approximation to the more general sampling strategy outlined in Metref *et al.* (2014). When $0 < \rho < 1$, this filter only guarantees that the first two moments of marginal quantities in $p_1(\mathbf{x}|y)$ are matched by posterior particles.

For a single observation, the equation used by Poterjoy (2016) and Poterjoy *et al.* (2019) for state updates is given by

$$\mathbf{x}_y^n \leftarrow \bar{\mathbf{x}}_y + \mathbf{r}_1 \circ (\mathbf{x}^{k_n} - \bar{\mathbf{x}}_y) + \mathbf{r}_2 \circ (\mathbf{x}^n - \bar{\mathbf{x}}_y), \quad (7)$$

where $\bar{\mathbf{x}}_y$ is the mean of $p_1(\mathbf{x}|y)$, \circ is an element-wise product, and the vectors \mathbf{r}_1 and \mathbf{r}_2 ensure the resulting particles match the first two moments of the marginals of $p_1(\mathbf{x}|y)$. The derivation of \mathbf{r}_1 and \mathbf{r}_2 also constrains the posterior particles to match particles sampled using marginal weights for $p(v|y)$, which are given by indices $k_1 \dots k_{N_e}$ in Equation (7); see Poterjoy (2016) for details. Equation (7) avoids the need to include a smoothing operator for a dynamical system characterized by highly correlated errors on a spatial grid for two reasons:

- 1 Particle updates consider the localized form of $p(u|v)$ discussed in Section 2; that is, updates to u are a function of the localization coefficient and particle indices obtained by sampling from $p(v|y)$. This property leads to a more consistent representation of $p_1(\mathbf{x}|y)$.
- 2 Solutions are only required to fit the first two moments of $p_1(\mathbf{x}|y)$, which tend to vary smoothly in space.

The update strategy adopted by Poterjoy (2016) and Poterjoy *et al.* (2019) is similar to the rank histogram filter of Anderson (2010), but with Equation (7) replacing a Kalman filter update. We refer readers to Poterjoy *et al.* (2019) for a full algorithmic description of this filter along with an exhaustive list of variable definitions.

The computational costs of the Poterjoy (2016) and Poterjoy *et al.* (2019) filters are much smaller than strategies that can sample perfectly from the localized posterior discussed in Section 2 (Metref *et al.*, 2014), thus making them affordable for high-dimensional problems such as weather prediction (e.g., Poterjoy *et al.*, 2021). Nevertheless, serial filters are more costly than filters that perform an independent sampling step for each variable—but require additional steps to model marginal dependence. The fitting of posterior particles to only the first two moments for unobserved quantities also presents problems for highly non-Gaussian applications (see Section 5), which is why past versions of this filter use an additional probability mapping step after resampling. Poterjoy (2016) and Poterjoy *et al.* (2019) refer to this step as kernel density distribution mapping (KDDM).

3 | THE COLLAPSE OF LOCALIZED PARTICLE FILTERS

Bengtsson *et al.* (2008), Bickel *et al.* (2008), and Snyder *et al.* (2008) outline conditions for collapse of the standard PF with Gaussian errors. Here, collapse refers to situations where importance weights equal zero for all but one particle, and the standard PF is the bootstrap PF with the prior as the proposal density. To prevent weight collapse, Snyder *et al.* (2008) conclude that N_e must increase exponentially with $\text{Var}[\sum_{j=1}^{N_y} V_j^n]$, where $V_j^n = -\ln [p(y_j|\mathbf{x}^n)]$, and the variance is taken over the proposal density—or prior density in this case. For example, $\text{Var}[\sum_{j=1}^{N_y} V_j^n]$ would be large if a dense network of accurate measurements were assimilated for a dynamical system containing many variables with independent errors.

Localization provides one mechanism for reducing weight collapse. For example, Poterjoy (2016) chooses a marginal probability density for variable x_j in \mathbf{x} that

expands Equation (6) for multiple observations with independent errors:

$$\begin{aligned}
 p_1(x_j|\mathbf{y}) &= A \sum_{n=1}^{N_e} \prod_{i=1}^{N_y} \left[\left(w_i^n - \frac{1}{N_e} \right) \rho_{i,j} + \frac{1}{N_e} \right] \delta(x_j - x_j^n), \\
 &= A \sum_{n=1}^{N_e} \prod_{i=1}^{N_y} \hat{\omega}_{i,j}^n \delta(x_j - x_j^n), \\
 &= A \sum_{n=1}^{N_e} \omega_j^n \delta(x_j - x_j^n), \tag{8}
 \end{aligned}$$

where $w_j^n \propto p(y_i|\mathbf{x}^n)$, A is a normalization factor, ω_j^n is the marginal weight for the n th particle, and $\rho_{i,j}$ is the localization coefficient for the i th observation and j th state variable. The localized contribution of each observation on ω_j^n is denoted by $\hat{\omega}_{i,j}^n$, so we can introduce $V_{i,j}^n \equiv -\ln(\hat{\omega}_{i,j}^n)$. Choosing $\rho_{i,j} < 1$ directly reduces $\text{Var}[\sum_{i=1}^{N_y} V_{i,j}^n]$ by capping the impact of distant observations on local weights. The resulting estimate of $p(x_j|\mathbf{y})$ is biased, since there is no guarantee it will converge to the Bayesian posterior as $N_e \rightarrow \infty$. Its numerical approximation, however, is less prone to weight collapse, thus providing a practical trade-off for small N_e . The same concept holds for all other PFs that enforce local constraints on posterior calculations from measurements.

For geophysical applications with dense networks of observations, $\rho_{i,j}$ can be insufficient for preventing weight collapse. Most noticeably, the arguments put forth by Bengtsson *et al.* (2008), Bickel *et al.* (2008), and Snyder *et al.* (2008) still hold when $\rho_{i,j} = 1$, which is problematic for assimilating very accurate measurements or many collocated measurements using localized PFs. In addition, the dynamical system of interest places a minimum bound on length scale parameters used to characterize $\rho_{i,j}$. This factor is important for global weather and ocean prediction models, where large spatial error correlation structures—commensurate with the Rossby radius of deformation—may dictate the minimum length scale for localization functions. Therefore, alternative strategies are needed to maintain the stability of PFs when $\rho_{i,j}$ provides an insufficient reduction in $\text{Var}[\sum_{i=1}^{N_y} V_{i,j}^n]$.

4 | NEW FILTER METHODOLOGY

This section introduces multiple strategies for overcoming the obstacles discussed in Section 3. Each of these methods builds naturally off one another, following in sequence from regularization to hybridization of PFs with parametric filters. They also resemble strategies that have already been adopted for Gaussian filters and smoothers, as well as different forms of PFs.

Before discussing new methodology, it is important to note that EnKFs suffer the same fate as PFs for small N_e (e.g., Morzfeld *et al.*, 2017), which motivates the use of prior and posterior error variance inflation techniques. Approaches to variance inflation can take multiple forms. Common strategies include multiplying ensemble perturbations by a factor greater than unity (Anderson and Anderson, 1999), adding random noise to posterior samples (Houtekamer and Mitchell, 2005), or relaxing a portion of the posterior update to ensemble perturbations back to the prior (Zhang *et al.*, 2004; Whitaker and Hamill, 2012). These strategies, as well as combining PFs with other filters (e.g., Stordal *et al.*, 2011; Frei and Künsch, 2013), applying tempered transitions and particle flow methodology (e.g., Del Moral *et al.*, 2006; Daum and Huang, 2011), or modifying the transition density between observation times (Van Leeuwen, 2010), are often used to prevent weight collapse for PFs. Likewise, PFs often adopt “pre-” and “post-regularization” schemes (Musso *et al.*, 2001), which are comparable to prior and posterior additive inflation algorithms used in the geophysical modeling community for EnKFs. Readers are encouraged to review Farchi and Bocquet (2018) for comparisons of these methods using low-dimensional numerical experiments with PFs.

More relevant to the current study, Poterjoy *et al.* (2019) discuss mechanisms for artificially increasing posterior error variance in a PF framework—in a manner that does not introduce large deviations from potentially non-Gaussian error approximations by the filter. They propose a strategy that uses observation error inflation to broaden the region of high posterior probability, using “effective sample size” as a metric for determining the amount of inflation. Effective sample size is a heuristic measure of the degrees of freedom in a weighted sample and is defined by $N_{\text{eff}} = [\sum (w^n)^2]^{-1}$ (Liu and Chen, 1998). The observation error inflation acts as regularization strategy, similar to techniques long used in the machine-learning community to prevent the overfitting of data (Moody, 1991). The very nature of this approach means that it does not change the shape of posterior distributions in the same manner as the multiplicative and additive inflation methods listed in the previous paragraph. Poterjoy *et al.* (2019) adopted this form of regularization to maintain filter stability in experiments applying the local PF for a real weather event using $N_e = 36$.

4.1 | A new regularized local particle filter

As an alternative to Poterjoy *et al.* (2019), we propose applying regularization directly to marginal weights,

which take into account the cumulative effect of measurements on posterior density calculations at a given grid point. This strategy has a number of desirable properties, which will be discussed in the following subsections.

First, recall that for independent Gaussian errors, weight collapse at the j th grid point occurs as $\text{Var}[\sum_{i=1}^{N_y} V_{i,j}^n]$ becomes large. The most straightforward means of enforcing an upper bound on this term is to multiply $V_{i,j}^n$ by a coefficient β_j , where $0 \leq \beta_j \leq 1$. This strategy is equivalent to raising particle weights to a power of β_j so that $\omega_j^n \propto \prod_{i=1}^{N_y} (\hat{\omega}_{i,j}^n)^{\beta_j}$. As discussed in Poterjoy *et al.* (2019), a natural choice of β_j is one that changes dynamically as a function of N_{eff} , which requires precomputing particle weights at each grid point and numerically solving for the largest β_j that gives $N_{\text{eff}} \geq N_{\text{eff}}^t$, where N_{eff}^t is a specified threshold value. In the absence of localization, the methods proposed in Poterjoy *et al.* (2019) and the current article are equivalent to inflating the measurement error variance for all observations by a factor of $1/\beta_j$.

To understand how the new approach differs from Poterjoy *et al.* (2019), recall that the observation error inflation strategy of Poterjoy *et al.* (2019) first computes a set of β coefficients $\{\tilde{\beta}_1, \dots, \tilde{\beta}_{N_y}\}$, where each $\tilde{\beta}_i$ is calculated for the observation-space weights $\{w_i^1, \dots, w_i^{N_e}\}$ that go into the calculation of $\hat{\omega}_{i,j}^n$ in Equation (8). Following this step, the final inflation coefficient assigned to the i th observation takes into account $\tilde{\beta}$ from nearby observations via

$$\beta_i = 1 + \sum_{k=1}^{N_y} (\tilde{\beta}_k - 1) \rho_{i,j}. \quad (9)$$

In Equation (9), $\rho_{i,j}$ is the same localization coefficient used for data assimilation. Localization coefficients act as heuristic regression terms that spread the observation error inflation value to nearby observations, thus preventing potential weight collapse caused by assimilating multiple accurate measurements. This approach differs from the regularization method proposed in the current study, which directly modulates the variance in $\{\omega_{i,j}^1, \dots, \omega_{i,j}^{N_e}\}$ for a given grid point, rather than modifying observation likelihoods. The benefits of this approach over Equation (9) are demonstrated via numerical experiments in Section 5. Interested readers are also encouraged to see Appendix B for details on how to reduce round-off error and computational cost for localized particle weight calculations, which are needed prior to estimating each β_j for regularization.

4.2 | An iterative local particle filter

This subsection introduces an adaptive tempering approach that builds off of the regularization discussed

in the previous subsection. As mentioned already, tempering (Neal, 2001) is another mechanism for reducing filter degeneracy for PFs. In general, tempering exploits a factorization of the likelihood to break the particle update step into a sequence of smaller update steps, thus allowing for mutations between iterations; see Van Leeuwen *et al.* (2019) for an extended discussion on the use of tempering for PFs. Iterative approaches of this type are already used extensively to form ensemble filters and smoothers that sample from non-Gaussian posteriors via a set of intermittent linear updates to prior particles (e.g., Zupanski, 2005; Sakov *et al.*, 2012; Emerick and Reynolds, 2012; Emerick and Reynolds, 2013; Bocquet and Sakov, 2014; Stordal and Elsheikh, 2015; Evensen, 2018). These strategies sometimes include a periodic relinearization of dynamical models and measurement operators about more accurate reference solutions, similar to incremental formulations of four-dimensional variational data assimilation methods (Courtier *et al.*, 1994), or generating multiple ensemble predictions each iteration to better estimate temporal covariance.

Dubinkina and Ruchi (2019) recently applied tempering for the ensemble transform PF (Reich, 2013), demonstrating advantages over regularization for problems with non-additive model errors. When applied to the Poterjoy *et al.* (2019) local PF, tempering serves a different purpose, which is to reduce errors introduced by updating unobserved variables based on the first two moments alone. In this section, we discuss two different forms of tempering. The first approach is to apply a factorization of the likelihood, as in past studies. That is, let $p(\mathbf{x}|\mathbf{y}) = p(\mathbf{x})p(\mathbf{y}|\mathbf{x})^{\alpha_1 + \alpha_2 + \dots + \alpha_{N_k}}$, where $\sum_{k=1}^{N_k} \alpha_k = 1$. The tempering introduces an extra recursion to the posterior update; that is,

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}) &\propto p(\mathbf{x})p(\mathbf{y}|\mathbf{x})^{\alpha_1} \dots p(\mathbf{y}|\mathbf{x})^{\alpha_{N_k}}, \\ &= p_{\alpha_1}(\mathbf{x}|\mathbf{y})p(\mathbf{y}|\mathbf{x})^{\alpha_2} \dots p(\mathbf{y}|\mathbf{x})^{\alpha_{N_k}}, \\ &\vdots \\ &= p_{\alpha_{N_k}}(\mathbf{x}|\mathbf{y}), \end{aligned} \quad (10)$$

where $p_{\alpha_k}(\mathbf{x}|\mathbf{y})$ is the posterior probability after the k th iteration of assimilating \mathbf{y} with tempering coefficients $\alpha_1, \dots, \alpha_k$. Note that for Gaussian likelihoods, this strategy is the same as assimilating \mathbf{y} N_k times with the observation error covariance inflated by $1/\alpha_k$ for the k th iteration. When the prior is used as a proposal density and $\rho_{i,j} = 1$, the tempering provides no asymptotic change in the posterior density estimate. But when combined with localization or other forms of particle mutation, the extra recursion adds diversity to posterior particles, which is beneficial for small ensembles.

Because tempering has no benefits when $\rho_{i,j} = 1$ in the local PF, we introduce an additional mixing coefficient for

particles, which increases the contribution of prior perturbations in the particle update equation. This is achieved by multiplying \mathbf{r}_1 by a scalar γ in Equation (7), where γ is specified to be a constant between 0 and 1. The prior weighting vector \mathbf{r}_2 is then recalculated to preserve the marginal posterior error variance for a given γ ; see Appendix C for details. The resulting particles are then recentered about the posterior mean. Typical values for γ range from 0.3 to 1, with γ tending toward 1 (i.e., no mixing) for larger N_e . It is important to mention that particles removed during resampling are the most affected by mixing, since the order determined for selected particles guarantees that surviving particles maintain the same index. The tempering strategy, combined with mixing, also does not completely replace the need for regularization, since choosing too small an N_k can easily lead to local weight collapse.

The tempering provides additional benefits for PFs that fit posterior moments, such as the Poterjoy (2016) and Poterjoy *et al.* (2019) filters. Suppose particle weights are regularized, so that N_{eff} is close to N_e , and the bootstrap resampling step leads to one particle being removed and replaced by another. The resulting particles are all assigned equal weights of $1/N_e$ following the sampling. Equivalently, one can assign equal importance weights to all prior particles except the pair being removed or sampled twice; in this case, the removed particle would have a weight of zero and the duplicated particle would have a weight of $2/N_e$. Knowledge of the first moment alone (estimated using importance weights) is then sufficient for identifying how the removed particle should be updated to provide the same solution as the standard bootstrap PF. Likewise, the local PF update can be separated into a sequence of sampling and merging steps, which only fit the first two moments, as in Equation (7). In doing so, the resulting update achieves a set of particles that more closely resembles samples from the localized PF density introduced in Equation (5). This property of tempering is illustrated through numerical simulations in Section 5.

Similar to iterative methods adopted for Gaussian filters, the N_k needed for a given problem depends on a variety of factors, such as how well prior densities are approximated by a Gaussian. For example, we find little to no benefits of tempering for Lorenz (1996) experiments performed using observation networks that are dense in space and time, as discussed in Section 5.2. These experiments yield small prior uncertainty between successive observation times, which is approximated well by a Gaussian. Likewise, each iteration can introduce additional noise in the solution, which may negate some positive benefits of tempering for large N_k .

A more efficient way of introducing iterations is to replace likelihood tempering with a tempering over marginal weights. This strategy expands the particle

weight regularization in Section 4.1 into a posterior tempering strategy. Recall, the regularized local PF uses the marginal weight equation $\omega_j^n \propto \prod_{i=1}^{N_y} (\hat{\omega}_{i,j}^n)^{\beta_j}$, where β_j enforces a minimum N_{eff} for weights. Following the assimilation of \mathbf{y} with regularization, we can repeat the particle updates, calculating $\beta_{j,k}$ for each iteration k based on N_{eff}^t , and stopping the processes when $\sum_{k=1}^{N_k} \beta_{j,k} \geq 1$. For the last iteration, β_{j,N_k} must be set to $1 - \sum_{k=1}^{N_k-1} \beta_{j,k}$ to ensure the weights sum to 1. The resulting method has properties similar to likelihood tempering, except the number of iterations is determined adaptively while enforcing the requirement that each set of marginal particle weights has $N_{\text{eff}} \geq N_{\text{eff}}^t$. Therefore, no additional regularization methods are needed. As in the likelihood tempering approach, expanding regularization to include an adaptive set of iterations (so that $\sum_{k=1}^{N_k} \beta_{j,k} = 1$) does not change the asymptotic behavior of the filter, as localization is relaxed for large sample sizes. It also has the desirable property of iterating only over subsets of variables in \mathbf{x} ; that is, additional iterations may not be needed for portions of a model domain characterized by low variance in particle weights.

4.3 | A hybrid filter

Regularization and tempering strategies provide a natural framework for combining the local PF with alternative filters when appropriate. For example, a recent study by Grooms and Robinson (2021) used an EnKF to update particles following an initial PF sampling step—exploiting the same likelihood factorization of past studies. This idea takes advantage of the fact that the posterior pdf is better approximated by a Gaussian than the prior when observation likelihoods are Gaussian (Morzfeld and Hodyss, 2019). A partial update by a PF can adjust particles to resemble samples from a Gaussian, thus making the EnKF an appropriate choice for the remaining update.

We apply a near identical strategy as Grooms and Robinson (2021) for the iterative local PF, which requires only minor changes to the algorithm. First, the stopping criteria for iterations is altered so that $\sum_{k=1}^{N_k} \beta_{j,k} = \kappa_{\text{max}}$ for $0 \leq \kappa_{\text{max}} \leq 1$. Note that setting $\kappa_{\text{max}} < 1$ results in a regularization of the local PF, similar to the method discussed in Section 4.1. Following the initial set of local PF iterations, the last adjustment is performed using an EnKF with the measurement error variance inflated by the factor $1/1 - \kappa_{\text{max}}$. The coefficient κ_{max} operates as a hybrid parameter, and the scheme approaches the iterative local PF in the limit $\kappa_{\text{max}} \rightarrow 1$ and the EnKF in the limit $\kappa_{\text{max}} \rightarrow 0$.

The state-space regularization, tempering, and hybrid local PF–EnKF methods are summarized in Algorithm 1. Note that because the three methods share many of the

Algorithm 1. Regularized, tempered, or hybrid local PF–EnKF

```

1 Function tempered_filter_cycle ( $\mathbf{x}^n$  for
   $n = 1 \dots N_e$ ,  $\mathbf{y}$ ,  $\mathbf{R}$ ,  $N_{\text{eff}}^t, \kappa_{\text{max}}$ ):
2    $\kappa_j = 0$  for  $j = 1 \dots N_x$ 
3   while  $\text{argmin}(\kappa) < \kappa_{\text{max}}$  do
4     for  $i = 1:N_y$  do
5       for  $j = 1:N_x$  do
6         if  $\kappa_j < \kappa_{\text{max}}$  then
7           Compute  $\rho_{i,j}$  using choice of
             localization function
8           Compute  $\omega_j^n$  in (8) recursively using
              $w_i^n$  and  $\rho_{i,j}$ 
9     for  $j = 1:N_x$  do
10      if  $\kappa_j < \kappa_{\text{max}}$  then
11        Apply bisection method to estimate  $\beta_j$ 
             for each set of marginal weights  $\omega_j^n$ 
             based on  $N_{\text{eff}}^t$ 
12         $\kappa_j \leftarrow \kappa_j + \beta_j$ 
13        if  $\kappa_j > \kappa_{\text{max}}$  then
14           $\beta_j = 1 - \kappa_{\text{max}}$ 
15           $\kappa_j = \kappa_{\text{max}}$ 
16      Update  $\mathbf{x}_n$  for  $n = 1 \dots N_e$  using local PF in
             Poterjoy et al. (2019) with  $\beta_j$  in the
             exponent of  $\omega_j^n$ 
17      switch Filter option do
18        case Regularized local PF do
19          Exit loop
20        case Tempered local PF do
21          Updated particles become prior for
             next iteration
22        case Hybrid local PF-EnKF do
23          if  $\text{argmax}(\kappa) = \kappa_{\text{max}}$  then
24             $\mathbf{R} \leftarrow \frac{1}{1 - \kappa_{\text{max}}} \mathbf{R}$ 
25            Update  $\mathbf{x}_n$  for  $n = 1 \dots N_e$  using EnKF
26            Exit loop
27          else
28            Updated particles become prior for
             next iteration
29      return Updated particles  $\mathbf{x}^n$  for  $n = 1 \dots N_e$ 

```

same calculations, they can be expressed compactly in a single algorithm with switches for determining which operations should be called by each filter. To satisfy an exit criterion, each iteration also requires a running sum of $\beta_{j,k}$, which is represented by κ_j in the algorithmic description.

5 | NUMERICAL EXPERIMENTS

This section explores the behavior of regularization, tempering, and hybrid strategies for the Poterjoy *et al.* (2019) local PF through numerical simulations. For the first set of experiments, we use a simple bivariate problem to illustrate how tempering alters the adjustment of particles compared with a single-step approach. The second set of experiments use the Lorenz (1996) model to compare the performance of regularized, tempered, and hybrid formulations of the local PF to past local PF configurations and the ensemble square-root EnKF of Whitaker and Hamill (2002). These experiments use simulated measurements to target observation collection scenarios with varying degrees of spatial and temporal density.

5.1 | Non-Gaussian bivariate application

We first form a bivariate data assimilation problem to examine the impact of different local PF sampling strategies. Though numerous challenges exist for localizing ensemble filters for high-dimensional problems, the current demonstration focuses more narrowly on a case where marginal dependence exists across a pair of variables. As discussed in Section 2, many geophysical data assimilation problems require a careful decoupling of variables, owing to gradual spatial transitions in error correlations. Therefore, localization must be able to modulate the dependence across variables in a controllable manner. For an application of this size, we can use Equation (5) from Section 2 to calculate and visualize the localized posterior pdf. This sample-estimated pdf is adopted for illustrative purposes, as it approximates the posterior density the candidate methods are trying to sample from.

Continuing the notation from past sections, let $\mathbf{x} = [u \ v]^T$ and consider the case where v is observed directly by measurement y with $\epsilon \sim N(0, 0.1)$. $N()$ indicates the function for a normal distribution, which in this case has a zero mean and variance of 0.1. We will compare four different algorithms for sampling from $p_1(u, v|y)$: (1) single-step Poterjoy *et al.* (2019) filter; (2) single-step Poterjoy *et al.* (2019) filter with KDDM; (3) direct KDDM resampling; and (4) iterative Poterjoy *et al.* (2019) filter. We do not show results for the regularized local PF in these comparisons, since this method represents one iteration of the iterative local PF. Likewise, the tempered transitions only use the local PF update, instead of the EnKF step discussed with regard to the hybrid filter (Section 4.3). When localization is used, the bivariate test problem used here has a multi-modal posterior density, which does not satisfy the underlying assumption of a Gaussian posterior for this method to be beneficial. We also draw special attention

to the direct KDDM sampling algorithm (experiment 3). When multiple observations are present, this is the only algorithm of the four that bypasses serial updates in place of a less costly independent update to u and v . As the kernel density estimates (KDEs) used for the KDDM approach a zero bandwidth, this method becomes identical to performing an independent bootstrap resampling on sorted particles for each variable. Increasing the bandwidth leads to a smoothing of posterior samples in a manner that is similar to the non-serial localized PF sampling strategies discussed in Section 2. For these experiments, we use a bandwidth that is equal to the posterior standard deviation for each variable, which is a robust option for real weather applications (Poterjoy *et al.*, 2019).

For the prior, we draw 1,000 particles from a bimodal distribution by sampling half from $N(\mathbf{x}_1, 0.1\mathbf{I})$ and half from $N(\mathbf{x}_2, 0.1\mathbf{I})$, with $\mathbf{x}_1 = [-1, -1]^T$, $\mathbf{x}_2 = [1, 1]^T$, and \mathbf{I} is an identity error covariance matrix. The large particle size helps illustrate the asymptotic behavior of the localized PF density approximation in Section 2. We acknowledge that localization is not needed for this problem when $N_e = 1,000$ —and even induces bias in the posterior density—but adopt this approximation primarily to examine how various localized filters attempt to sample from the density assumed when localization is introduced. For visualization purposes, it is more practical to calculate the prior and posterior pdfs using Gaussian KDEs instead of a Dirac delta functions, which allows for smooth contours in figures. In doing so, we choose kernels with a standard deviation of 0.2 for all pdfs. For reference, Figure 2a–c shows prior samples and KDEs for $p(u, v)$ and the true (non-localized) $p(u, v|y)$, as well as marginals for

each density. For all subsequent figures in this section, we plot the localized posterior density $p_l(u, v|y)$, which is the posterior density resulting from a decoupling of u and v using $\rho_{u,v} = 0.75$. This posterior differs from the true posterior in that dependence between u and v is modified in a manner typically used to treat sampling error during data assimilation—as such, it is always biased whenever true dependence exists, which is inevitable for geophysical models. Without localization, the resulting marginals are close to Gaussian (Figure 3a), since particles would only retain large weights in the positive mode. The resulting decoupling leads to non-zero posterior probability in the upper left portion of the domain (Figure 4a) as the directly observed variable retains the same posterior marginal, but the unobserved variable retains both modes—albeit with less density in the negative one. We remind readers that the sole purpose of this exercise is to examine how different filters sample from the localized density instead of the true posterior because the true posterior cannot be obtained in practice.

We first examine how the single-step Poterjoy *et al.* (2019) filter draws samples from $p_l(u, v|y)$. By construction, this method properly samples from $p(v|y)$ as indicated by the markers in Figure 3b. For the unobserved variable, the method draws numerous particles between the two modes in $p(u|y)$, thus failing to maintain consistency with the analytical solution. This example shows the limitations of fitting only two moments during the update of unobserved variables, thus motivating the additional probability mapping procedure discussed in Poterjoy (2016). When KDDM is performed following the single-step update (Figure 4), the suboptimal update is corrected; that is, particles that

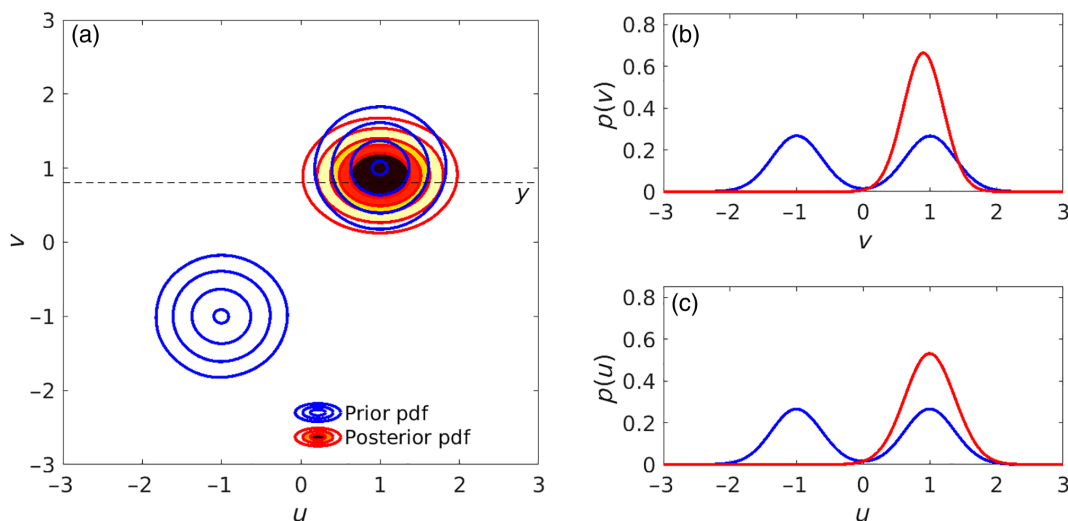


FIGURE 2 (a) Schematic showing the true prior probability density function (pdf; blue contours), true posterior pdf (red contours and shading), and observation for bivariate data assimilation experiments (dashed line). Marginal prior and posterior pdfs are also shown in (b) and (c) for variables v and u , respectively [Colour figure can be viewed at wileyonlinelibrary.com]

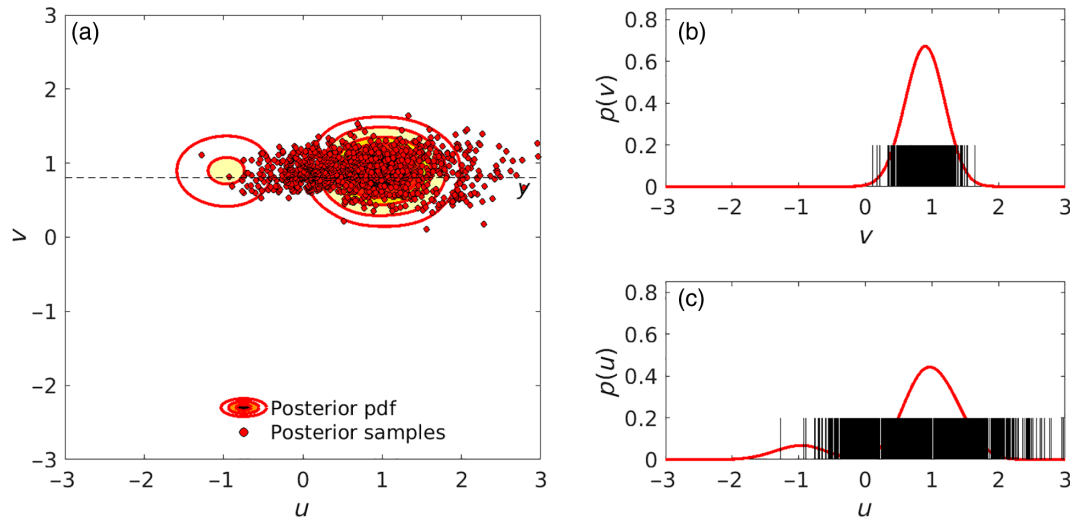


FIGURE 3 (a) Schematic showing the localized posterior probability density function (pdf; red contours and shading), posterior particles (red markers), and observation (dashed line) for an experiment using the Poterjoy *et al.* (2019) local particle filter. Marginal posterior pdfs and particles are also shown in (b) and (c) for variables v and u , respectively [Colour figure can be viewed at wileyonlinelibrary.com]

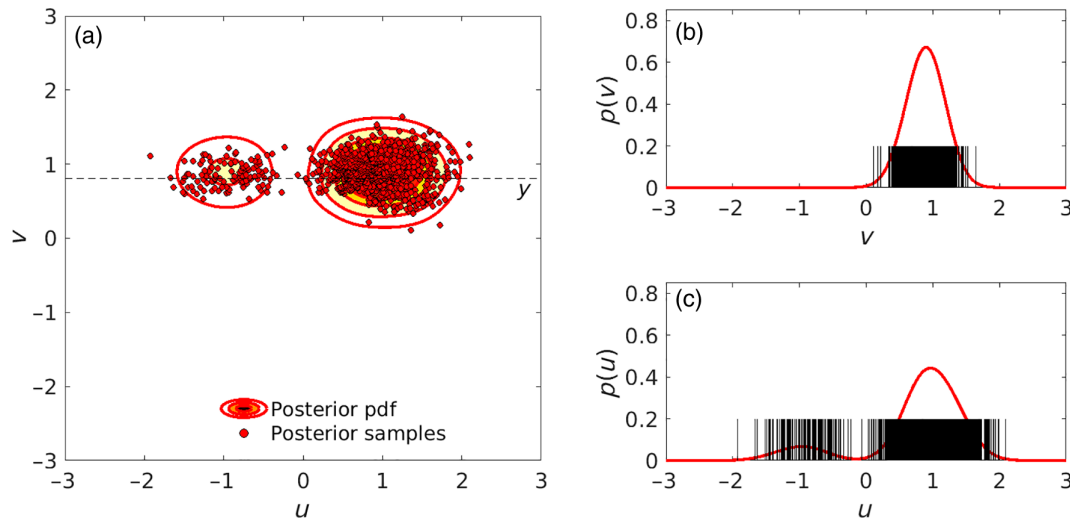


FIGURE 4 Same as Figure 3, but with the kernel density distribution mapping step applied after updates [Colour figure can be viewed at wileyonlinelibrary.com]

fall in the low probability region between the two modes are shifted into the modes. The success of this approach, however, requires applying the sampling and merging process (i.e., Figure 3) prior to the additional adjustment for $p(u|y)$. The sampling step ensures particles are properly ordered before their marginals are mapped into posterior KDEs through KDDM.

Figure 5 demonstrates the outcome of applying KDDM directly to adjust prior particles based on marginal posterior pdfs alone. This strategy is similar to the one-dimensional transport map or anamorphosis technique discussed in Farchi and Bocquet (2018) for updating particles based on marginal quantities alone. They find this approach to give reasonably accurate

solutions for the Lorenz (1996) model, using the observation network examined in their study. Particle updates from KDDM provide a consistent sampling from each marginal (Figure 5b,c). The sampling for unobserved variable u , however, is not explicitly conditioned on the sampling performed for v . Instead, $p(u|v)$ is determined by how particles are sorted prior to sampling and by parameters that enforce smoothness, such as the kernel bandwidth in KDDM. As a result, KDDM and other transform strategies that operate on posterior marginals independently will lead to a suboptimal sampling from the multivariate posterior. As discussed in Poterjoy (2016), KDDM preserves quantiles when adjusting particles, which allows the method to retain

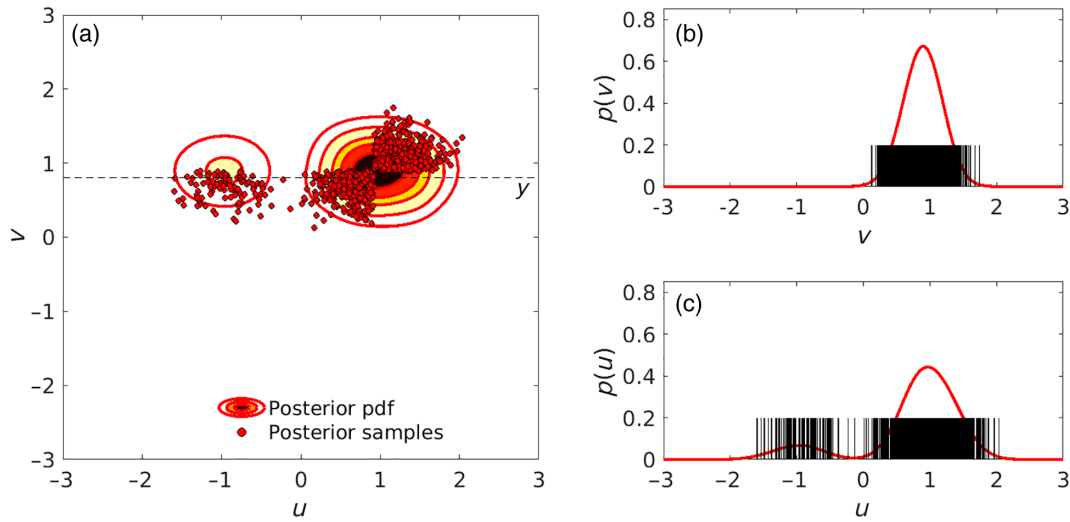


FIGURE 5 Same as Figure 3, but with kernel density distribution mapping used in place of the Poterjoy *et al.* (2019) update step [Colour figure can be viewed at wileyonlinelibrary.com]

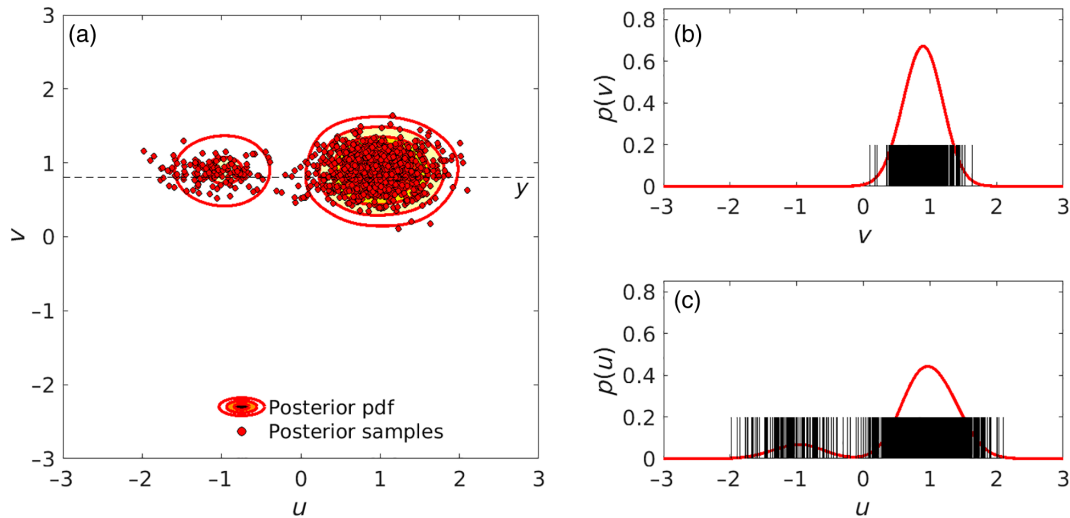


FIGURE 6 Same as Figure 3, but for the iterative local particle filter [Colour figure can be viewed at wileyonlinelibrary.com]

marginal dependence when transforming particles. In this case, dependent structure represented by prior particles persists through the posterior update, which induces artificial bimodal behavior within the first posterior mode (Figure 5a). This behavior is also commonly observed in deterministic ensemble square-root Kalman filters when the prior pdf is multi-modal (e.g., Poterjoy *et al.* 2017, figure 4).

Lastly, Figure 6 shows the results of updating particles through a series of tempered transitions. Particle weights are forced to satisfy $N_{\text{eff}} = 0.8N_e$, which means only a small number of particles are removed when sampling from $p(v|y)$ during each iteration. The update equation in Equation (7) then becomes sufficient for producing particles that satisfy non-Gaussian properties of the localized posterior, despite being derived using the

first two moments. For reasons discussed in Section 4, iterations allow the local PF to provide samples from a non-Gaussian posterior without a need for an additional KDDM step (Figure 6a).

5.2 | The 40-variable dynamical system

For the next set of experiments, we adopt the Lorenz (1996) dynamical model to explore the behavior of the new filters in a sequential data assimilation framework. This model consists of N_x equally spaced variables defined on a periodic domain, which are governed by a set of N_x differential equations:

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \quad (11)$$

where $x_{i+N_x} = x_i$ and $x_{i-N_x} = x_i$. The model is integrated forward in time using the fourth-order Runge–Kutta scheme and a model time step of 0.05. These experiments use the standard configuration of the Lorenz (1996) model, with $N_x = 40$ and $F = 8$, except for one set of experiments that include “model error”; in this case, measurements are simulated from a model trajectory with $F = 9$.

As discussed in Section 1, the current study is motivated by geophysical data assimilation problems characterized by biased prior errors, but accurate measurements, which are known to yield challenges for PFs, including ones that adopt localization. To reproduce this problem, we simulate numerous observation networks for the Lorenz (1996) model, each differing in spatial and temporal density to compare the performance of the new filters. We vary the measurement density by simulating sets of observations that are spaced every of one, two, and four grid points; that is, $N_y = 40$, $N_y = 20$, and $N_y = 10$, respectively. We also vary the forecast integration time (T) between measurements, using $T = 0.05$, $T = 0.10$, $T = 0.20$, and $T = 0.30$. Measurement errors are sampled from $N(0, 0.05)$, thus providing highly accurate observation networks, which can induce particle weight collapse as described in Section 3. For simplicity in notation, we refer to the regularized local PF as “local RPF,” the iterative local PF as “local IPF,” and the hybrid iterative local PF as the “local IPF–EnKF.” We also perform EnKF experiments using the Whitaker and Hamill (2012) square-root filter with Anderson (2007) adaptive state-space prior inflation. This filter and inflation methods are commonly used for geoscience applications, thus providing a suitable benchmark for comparing new methodology.

We perform two sets of experiments: one using $N_e = 40$ particles and another using $N_e = 120$ particles. Each experiment consists of 11,000 observation times, with the first 1,000 times designated for spin up. For localization, we adopt a Gaussian-shaped localization function with a width controlled by a radius of influence (ROI). We rigorously tune all filter parameters, namely ROI, N_{eff} , γ , and adaptive inflation parameters needed for the EnKF, to provide the smallest root-mean-squared error (RMSE) for a given observation network. To reduce the parameter space, we fix the hybrid parameter κ_{max} at 0.5 for local IPF–EnKF experiments. This choice is motivated by the comparable performance of the local IPF and EnKF for the given observation networks. Filter sensitivity to this parameter will be explored in subsequent studies for real problems. We also refer to Kurosawa and Poterjoy (2021) for an examination of the local IPF–EnKF for nonlinear measurement operators and its comparison with iterative ensemble and variational smoothers.

We summarize numerical results from all Lorenz (1996) experiments in Figures 7 and 8, which shows the RMSE of each method as a function of N_y and T . In general, we find the relative performance of each method to vary substantially with observation density. For observation networks that are dense in both space and time, the EnKF and local IPF–EnKF produce the smallest errors, owing to the prior distributions remaining close to Gaussian over time. The advantage of the EnKF declines when N_y is decreased to 10 but T remains small, as the prior spread increases and becomes less Gaussian. More notably, the regularized local PF configurations show a sharper decline in skill as T increases, regardless of the spatial density of measurements. The decline in performance is less dramatic for experiments performed with $N_e = 120$, owing to the lesser need for regularization. This result is expected, as fewer particles obtain high likelihoods for the accurate measurements simulated in these experiments, which is most detrimental when N_e is small. We also note that Poterjoy *et al.* (2019) show the original non-regularized Poterjoy (2016) filter diverging for sparse observation networks of this type. Regularization helps stabilize these methods by artificially decreasing the impact of potentially accurate measurements. Despite this deficiency, the newly proposed regularization strategy (local RPF) provides comparable RMSEs to the observation error inflation introduced by Poterjoy *et al.* (2019). The new method, however, has the added benefit of being naturally extended to the adaptive tempering strategy used for the local IPF experiments.

When N_y is small and T is large, the local IPF provides improvements over the non-iterative methods and yields RMSEs that are more comparable to the EnKF. The number of iterations, though chosen adaptively through N_{eff}^t , remains low for the experiments performed in this study. The maximum number of iterations for a given marginal does not exceed five for any observation network, with the exception of cycles performed during the “spin-up” period. This algorithm also outperforms the EnKF when model error is introduced (Figure 7d). We hypothesize that the improvements over EnKF for the model error experiments follow mostly from the mixing strategy (controlled by parameter γ), which increases diversity in particles between successive sampling steps. We find further improvements when the local IPF is extended to include an EnKF step during iterations—thus forming the hybrid local IPF–EnKF. This algorithm outperforms all methods tested here and is especially beneficial in experiments with very sparse observations and model error (i.e., Figure 7c,d). We also note that the performance of the hybrid filter does not decline substantially between the $N_e = 120$ and $N_e = 40$ experiments, thus showing greater value for real geophysical data

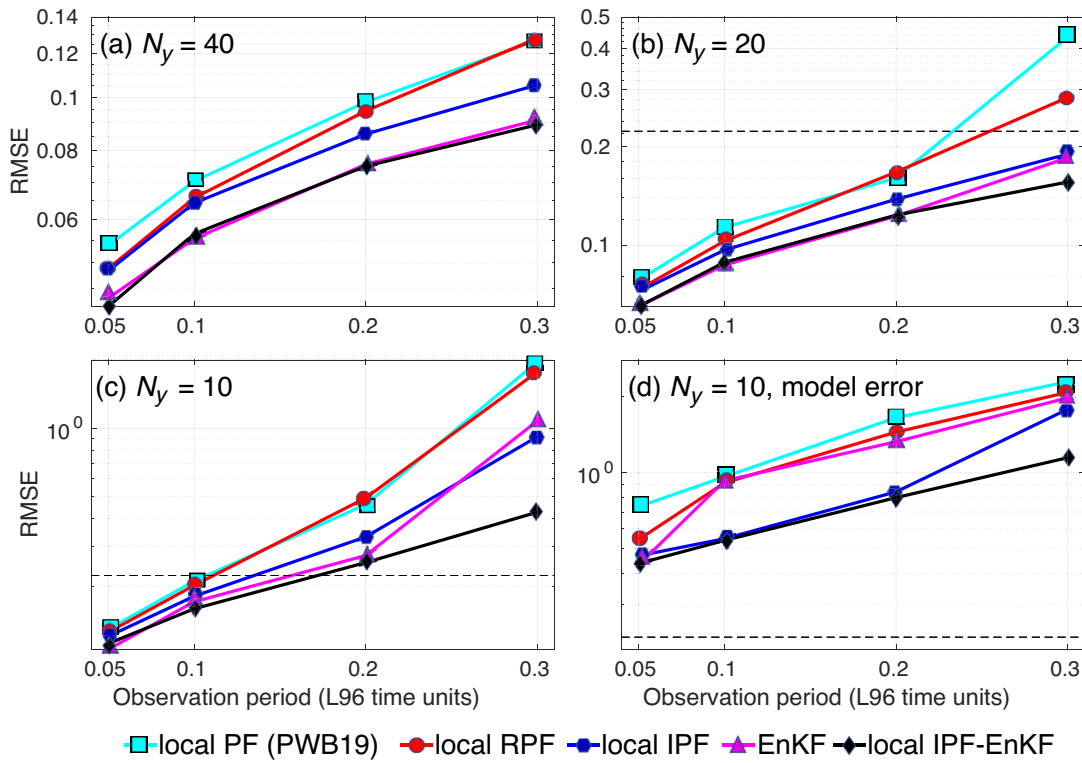


FIGURE 7 Posterior mean root-mean-squared errors (RMSEs) for $N_e = 40$ experiments performed with (a) $N_y = 40$, (b) $N_y = 20$, and (c) $N_y = 10$ equally spaced measurements. RMSEs are shown as a function of observation period, ranging from $T = 0.5$ to $T = 0.30$. (d) Results for $N_y = 10$, but using an imperfect Lorenz (1996) (L96) model for forecast steps. The dashed line in panels indicates the observation error standard deviation. EnKF: ensemble Kalman filter; PF: particle filter; IPF: iterative PF; RPF: regularized PF; PWB19: Poterjoy *et al.* (2019) [Colour figure can be viewed at wileyonlinelibrary.com]

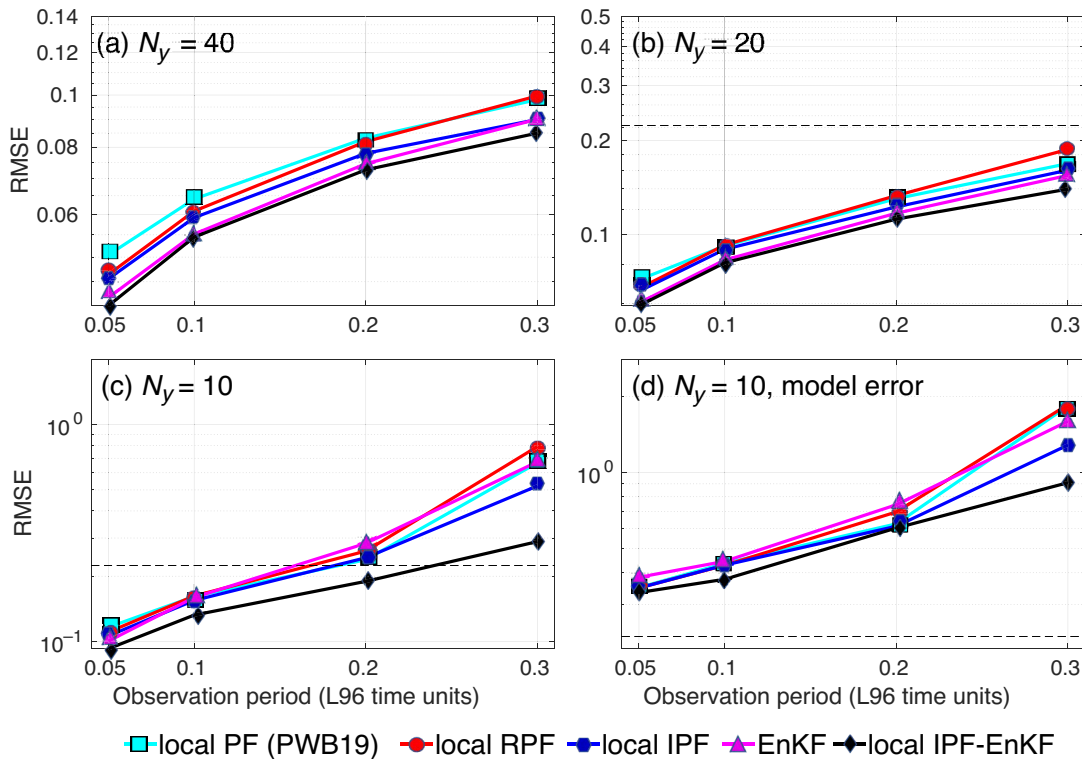


FIGURE 8 Same as Figure 7 but for $N_e = 120$ [Colour figure can be viewed at wileyonlinelibrary.com]

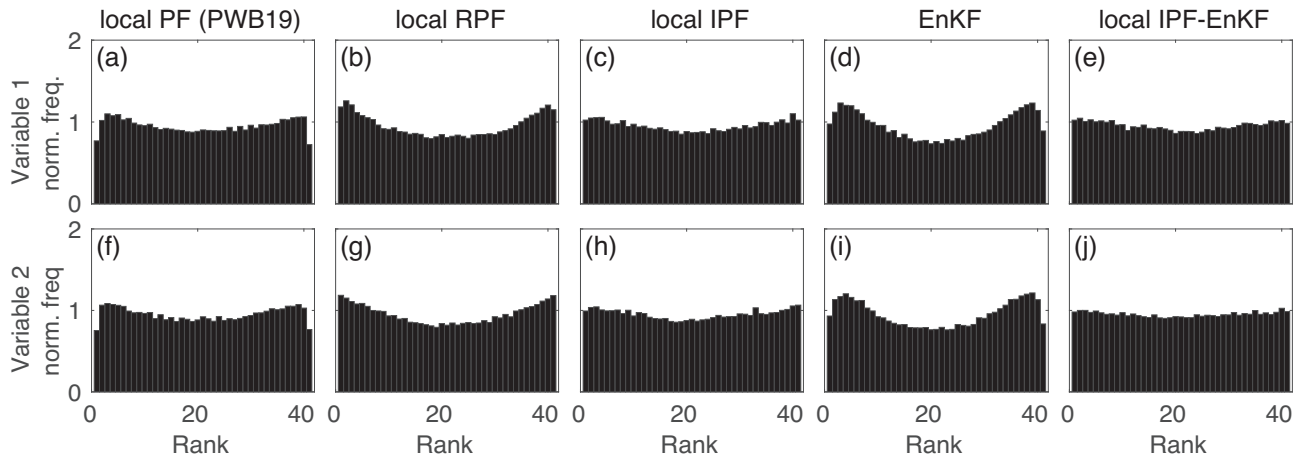


FIGURE 9 Rank histogram verification for experiments performed using $N_y = 20$, $T = 0.2$, and a perfect model: (a–e) variable 1; (f–j) variable 2. The frequency in each bin is normalized to yield values that are close to unity. EnKF: ensemble Kalman filter; PF: particle filter; IPF: iterative PF; RPF: regularized PF; PWB19: Poterjoy *et al.* (2019)

assimilation problems, which use limited numbers of particles.

In addition to comparing posterior RMSEs, we perform a rank histogram verification (Anderson, 1996; Hamill and Colucci, 1996; Talagrand *et al.*, 1997) on posterior particles generated over one set of $N_e = 40$ experiments for a regime where RMSEs are below the measurement uncertainty, but non-Gaussian priors are expected to have a role in interpreting the results. These experiments use a perfect model and an equally spaced observation network defined using $N_y = 20$ and $T = 0.2$ (Figure 7b). Assimilating measurements from this observation network poses challenges because of nonlinearity in the model dynamics and data sparsity, which can be isolated in perfect-model experiments. With $N_e = 40$, either regularization (for PFs) or prior inflation (for EnKF) is needed to retain stable solutions, which are expected to degrade the probabilistic skill of the ensemble filters. Furthermore, we extended these experiments to include 10^6 observation times to increase the reliability of the rank histogram verification.

Figure 9 shows the rank-histogram verification for a pair of observed and unobserved variables, x_1 and x_2 , respectively. In general, the regularized PFs and EnKF produce rank histograms that exhibit larger frequencies near the periphery of the sorted particles, indicating an underdispersion of particles for these methods. This result is expected, as configurations that yield the smallest RMSEs tend to coincide with the smallest regularization needed to prevent filter divergence over the course of the experiments. Though not shown, the EnKF tends to produce flatter histograms for denser observation networks and highly U-shaped histograms for sparse networks. Therefore, we attribute the underdispersion to Gaussian assumptions

made by the EnKF and adaptive inflation. For the regularized PFs, the shape of the histograms tends to remain similar to those shown in Figure 9, which is suboptimal but consistent across experiments.

The local IPF and IPF–EnKF algorithms yield nearly uniform histograms. We also find these results to be less sensitive to choices of N_{eff} , since the two methods use this parameter to dictate the number of tempered transitions, rather than regulate the update. We also draw attention to the ability of the local IPF–EnKF method to maintain a flat histogram for the current observation network. This result indicates that the local PF is effective at shifting particles into an approximate Gaussian before applying the EnKF step for this particular application.

6 | DISCUSSION AND CONCLUSIONS

The theoretical benefits of PFs motivate their application for a large range of numerical weather prediction applications. These applications include high-impact weather events, such as severe convective storms and tropical cyclones, which are guided by highly nonlinear dynamics and observed by remote-sensing platforms on radars and satellites. In addition to posing challenges for Gaussian-based data assimilation techniques, these applications are also characterized by large sampling errors in ensemble estimates for prior probability densities, which limit the direct use of PFs—even with localization.

This study introduces new regularization, tempering, and hybrid approaches for the local PF of Poterjoy (2016) and Poterjoy *et al.* (2019), which aim to improve the effectiveness of this filter for high sampling-error regimes.

Regularization reduces the impact of observations on particle weights, which directly prevents weight collapse when few or no particles contain high likelihoods. Because regularization is achieved by raising particle weights to a power less than one, this strategy permits the assimilation of observations multiple times in the same manner as iterative data assimilation strategies that rely on a factorization of the likelihood function. Likewise, the use of an alternative—possibly parametric—filter during the last iteration of the tempered local PF is trivial. Following past studies (e.g., Frei and Künsch, 2013; Chustagulprom *et al.*, 2016; Grooms and Robinson, 2021), we use this factorization to form a local PF–EnKF hybrid.

To examine the behavior of iterations for the local PF, we derive the posterior density that results from localization and form a bivariate data assimilation problem that produces a multivariate bimodal posterior. Through numerical simulation, we examine how different methods draw samples from a multivariate non-Gaussian density, which is a known challenge for filters currently used for high-dimensional applications. The serial local PF introduced in Poterjoy (2016) solves multivariate problems of this type by first drawing samples for the marginal posterior density in observation space and then updating unobserved model variables using a mix of sampled and prior particles. Because the update of unobserved variables only considers the first two moments, we find this strategy to be suboptimal for the bimodal application. Likewise, a sampling strategy that draws quasi-independently from marginals to approximate samples from the full multivariate posterior density is also found to be suboptimal. For the iterative local PF, the single-step update is replaced with a series of smaller updates, which we find to force appropriate transitions of particles between modes, despite only fitting two moments. This benefit follows from the limited degrees of freedom that exist when only a small number of particles need to be adjusted during each iteration.

We compare the filtering performance of the regularized, iterative, and hybrid local PF with a past implementation of the local PF (Poterjoy *et al.*, 2019) and a square-root EnKF (Whitaker and Hamill, 2012) using the 40-variable model of Lorenz (1996). From these experiments, the iterative filter yields the largest improvements over regularized local PF strategies when observations are sparse in space and time. The number of prior particles with large likelihoods decreases considerably when prior error variance becomes much larger than measurement error variance—thus producing the low observation error regime that motivates the current study. Furthermore, we find the local IPF–EnKF hybrid to outperform both the iterative local PF and EnKF for nearly all measurement networks examined. The performance gains come in the form of reduced posterior RMSEs, as well as more uniform

rank histograms, which demonstrates the benefits of this approach for applications characterized by non-Gaussian prior densities by approximately Gaussian posterior densities.

Though the iterative strategy introduced in this study brings considerable performance gains to the local PF, it also adds to the computational complexity. The number of iterations is determined adaptively in the algorithm through an effective ensemble size parameter, making it difficult to estimate how many additional calculations are required. This aspect of the algorithm is further complicated by allowing the number of iterations to vary across model variables. Therefore, data assimilation problems characterized by biased priors in significant portions of the state space will require more iterations—and more computational resources—than those with low error. Nevertheless, savings can be achieved by batching observations geographically or ignoring observations that yield very small changes to particle weights. These approaches have already been applied for weather forecasting applications of the local PF and will be discussed in a future study.

ACKNOWLEDGEMENTS

Funding for this work was provided by a U.S. National Science Foundation CAREER Award #AGS1848363 and NOAA grant #NA20OAR4600281. The author is thankful for comments and edits provided by two anonymous reviewers.

ORCID

Jonathan Poterjoy  <https://orcid.org/0000-0003-3927-4642>

REFERENCES

- Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R. and Avellano, A. (2009) The Data Assimilation Research Testbed: A community facility. *Bulletin of the American Meteorological Society*, 90(9), 1283–1296.
- Anderson, J.L. (1996) A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, 9, 1518–1530.
- Anderson, J.L. (2003) A local least squares framework for ensemble filtering. *Monthly Weather Review*, 131, 634–642.
- Anderson, J.L. (2007) An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A: Dynamic Meteorology and Oceanography*, 59, 210–224.
- Anderson, J.L. (2010) A non-Gaussian ensemble filter update for data assimilation. *Monthly Weather Review*, 138(11), 4186–4198.
- Anderson, J.L. and Anderson, S.L. (1999) A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127(12), 2741–2758.
- Anderson, J.L. and Collins, N. (2007) Scalable implementations of ensemble filter algorithms for data assimilation. *Journal of Atmospheric and Oceanic Technology*, 24, 1452–1463.

- Bengtsson, T., Bickel, P. and Li, B. (2008). Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. In: D. Nolan and T. Speed (Eds.), *Probability and statistics: Essays in honor of David A. Freedman*, Institute of Mathematical Statistics Collections (Vol. 2 pp. 316–334). Beachwood, OH: Institute of Mathematical Statistics.
- Bengtsson, T., Snyder, C. and Nychka, D. (2003) Toward a nonlinear ensemble filter for high-dimensional systems. *Journal of Geophysical Research: Atmospheres*, 108(D4), 8775.
- Bickel, P., Li, B. and Bengtsson, T. (2008). Sharp failure rates for the bootstrap particle filter in high dimensions. In: B. Clarke and S. Ghosal (Eds.), *Pushing the limits of contemporary statistics: Contributions in honor of Jayanta K. Ghosh*, Institute of Mathematical Statistics Collections (Vol. 3 pp. 318–329). Beachwood, OH: Institute of Mathematical Statistics.
- Bocquet, M. and Sakov, P. (2014) An iterative ensemble Kalman smoother. *Quarterly Journal of the Royal Meteorological Society*, 140(682), 1521–1535.
- Chustagulprom, N., Reich, S. and Reinhardt, M. (2016) A hybrid ensemble transform particle filter for nonlinear and spatially extended dynamical systems. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1), 592–608.
- Courtier, P., Thepáut, J.-N. and Hollingsworth, A. (1994) A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519), 1367–1387.
- Daum, F. and Huang, J. (2011). Particle flow for nonlinear filters, *2011 IEEE international conference on acoustics, speech, and signal processing: Proceedings* (pp. 5920–5923). Piscataway, NJ: IEEE.
- Del Moral, P., Doucet, A. and Jasra, J. (2006) Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 68(3), 411–436.
- Doucet, A., de Freitas, N. and Gordon, N. (2001). An introduction to sequential Monte Carlo methods. In A. Doucet, N. de Freitas, and N. Gordon (Eds.), *Sequential Monte Carlo methods in practice* (pp. 3–14). New York, NY: Springer.
- Dubinkina, S. and Ruchi, S. (2019). Accounting for model error in tempered ensemble transform particle filter and its application to non-additive model error. <https://doi.org/10.48550/arXiv.1901.03144>.
- Emerick, A.A. and Reynolds, A.C. (2012) History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. *Computational Geosciences*, 16, 639–659.
- Emerick, A.A. and Reynolds, A.C. (2013) Ensemble smoother with multiple data assimilation. *Computational Geosciences*, 55, 3–15.
- Evensen, G. (2018) Analysis of iterative ensemble smoothers for solving inverse problems. *Computational Geosciences*, 22(3), 885–908.
- Farchi, A. and Bocquet, M. (2018) Review article: Comparison of local particle filters and new implementations. *Nonlinear Processes in Geophysics*, 25, 765–807.
- Feng, J., Wang, X. and Poterjoy, J. (2020) A comparison of two local moment-matching nonlinear filters: Local particle filter (LPF) and local nonlinear ensemble transform filter (LNETF). *Monthly Weather Review*, 148(11), 4377–4395.
- Frei, M. and Künsch, H.R. (2013) Bridging the ensemble Kalman and particle filters. *Biometrika*, 100(4), 781–800.
- Gaspari, G. and Cohn, S.E. (1999) Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554), 723–757.
- Greybush, S.J., Kalnay, E., Miyoshi, T., Ide, K. and Hunt, B. (2011) Balance and ensemble Kalman filter localization techniques. *Monthly Weather Review*, 139(2), 511–522.
- Grooms, I. and Robinson, G. (2021) A hybrid particle–ensemble Kalman filter for problems with medium nonlinearity. *PLoS ONE*, 16(3), e0248266.
- Hamill, T.M. and Colucci, S.J. (1996). Random and systematic error in NMC’s short-range Eta ensembles, *Preprints. 13th conference on probability and statistics in the atmospheric sciences* (pp. 51–56). Boston, MA: American Meteorological Society.
- Houtekamer, P.L. and Mitchell, H.L. (2005) Ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 131(613), 3269–3289.
- Houtekamer, P.L. and Zhang, F. (2016) Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 144, 4489–4532.
- Hunt, B.R., Kostelich, E.J. and Szunyogh, I. (2007) Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, 230(1–2), 112–126.
- Keper, J.D. (2009) Covariance localisation and balance in an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 135(642), 1157–1176.
- Kurosawa, K. and Poterjoy, J. (2021) Data assimilation challenges posed by nonlinear measurement operators: A comparative study of ensemble and variational filters and smoothers. *Monthly Weather Review*, 149, 2369–2389.
- Lee, Y. and Majda, A.J. (2016) State estimation and prediction using clustered particle filters. *Proceedings of the National Academy of Sciences of the United States of America*, 113(51), 14609–14614.
- Lei, L. and Anderson, J.L. (2014) Comparisons of empirical localization techniques for serial ensemble Kalman filters in a simple atmospheric general circulation model. *Monthly Weather Review*, 142, 739–754.
- Liu, J.S. and Chen, R. (1998) Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93(443), 1032–1044.
- Lorenz, E.N. (1996). Predictability—A problem partly solved. Paper presented at Seminar on Predictability, 4–8 September 1995. ECMWF, Reading, UK. <https://www.ecmwf.int/node/10829>.
- Metref, S., Cosme, E., Snyder, C. and Brasseur, P. (2014) A non-Gaussian analysis scheme using rank histograms for ensemble data assimilation. *Nonlinear Processes in Geophysics*, 21, 869–885.
- Moody, J.E. (1991). The effective number of parameters: An analysis of generalization and regularization in nonlinear learning systems. In: J.E. Moody, S.J. Hanson, and R.P. Lippmann (Eds.), *NIPS’91: Proceedings of the 4th international conference on neural information processing systems* (pp. 847–854). San Francisco, CA: Morgan Kaufmann.
- Morzfeld, M. and Hodyss, D. (2019) Gaussian approximations in filters and smoothers for data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, 71(1), 1600344.
- Morzfeld, M., Hodyss, D. and Poterjoy, J. (2018) Variational particle smoothers and their localization. *Quarterly Journal of the Royal Meteorological Society*, 144(712), 806–825.
- Morzfeld, M., Hodyss, D. and Snyder, C. (2017) What the collapse of the ensemble Kalman filter tells us about particle filters. *Tellus A: Dynamic Meteorology and Oceanography*, 69(1), 1283809.
- Musso, C., Oudjane, N. and Le Gland, F. (2001). Improving regularised particle filters. In: A. Doucet, N. De Freitas, and N. Gordon

- (Eds.), *Sequential Monte Carlo Methods in Practice* (pp. 247–271). New York, NY: Springer-Verlag.
- Neal, R.M. (2001) Annealed importance sampling. *Statistics and Computing*, 11, 125–139.
- Penny, S.G. and Miyoshi, T. (2016) A local particle filter for high dimensional geophysical systems. *Nonlinear Processes in Geophysics*, 23, 391–405.
- Posselt, D.J. (2016) A Bayesian examination of deep convective squall line sensitivity to changes in cloud microphysical parameters. *Journal of the Atmospheric Sciences*, 73(2), 637–665.
- Posselt, D.J., Hodyss, D. and Bishop, C.H. (2014) Errors in ensemble Kalman smoother estimates of cloud microphysical parameters. *Monthly Weather Review*, 142, 1631–1654.
- Poterjoy, J. (2016) A localized particle filter for high-dimensional nonlinear systems. *Monthly Weather Review*, 144, 59–76.
- Poterjoy, J., Alaka, G. and Winterbottom, H. (2021) The irreplaceable utility of sequential data assimilation for model development: Lessons learned from an experimental HWRP system. *Weather and Forecasting*, 36(2), 661–677.
- Poterjoy, J. and Anderson, J.L. (2016) Efficient assimilation of simulated observations in a high-dimensional geophysical system using a localized particle filter. *Monthly Weather Review*, 144, 2007–2020.
- Poterjoy, J., Sobash, R.A. and Anderson, J.L. (2017) Convective-scale data assimilation for the weather research and forecasting model using the local particle filter. *Monthly Weather Review*, 145, 1897–1918.
- Poterjoy, J., Wicker, L. and Buehner, M. (2019) Progress toward the application of a localized particle filter for numerical weather prediction. *Monthly Weather Review*, 147, 1107–1126.
- Poterjoy, J. and Zhang, F. (2011) Dynamics and structure of forecast error covariance in the core of a developing hurricane. *Journal of the Atmospheric Sciences*, 68(8), 1586–1606.
- Potthast, R., Walter, A. and Rhodin, A. (2019) A localized adaptive particle filter within an operational NWP framework. *Monthly Weather Review*, 147, 345–362.
- Reich, S. (2013) A nonparametric ensemble transform method for Bayesian inference. *SIAM Journal on Scientific Computing*, 35(4), A2013–A2024.
- Reich, S. and Cotter, C. (2015) *Probabilistic forecasting and Bayesian data assimilation*, (pp. 33–64). Cambridge, UK: Cambridge University Press.
- Robert, S. and Künsch, H.R. (2017) Localizing the ensemble Kalman particle filter. *Tellus A: Dynamic Meteorology and Oceanography*, 69(1), 1282016.
- Robert, S., Leuenberger, D. and Künsch, H.R. (2018) A local ensemble transform Kalman particle filter for convective-scale data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144(713), 1279–1296.
- Sakov, P., Oliver, D.S. and Bertino, L. (2012) An iterative EnKF for strongly nonlinear systems. *Monthly Weather Review*, 140, 1988–2004.
- Shao, H., Derber, J., Huang, X.-Y., Hu, M., Newman, K., Stark, D., Lueken, M., Zhou, C., Nance, L., Kuo, Y.-H. and Brown, B. (2016) Bridging research to operations transitions: Status and plans of community GSI. *Bulletin of the American Meteorological Society*, 97(8), 1427–1440.
- Snyder, C., Bengtsson, T., Bickel, P. and Anderson, J. (2008) Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, 136, 4629–4640.
- Stordal, A. and Elsheikh, A.H. (2015) Iterative ensemble smoothers in the annealed importance sampling framework. *Advances in Water Resources*, 86, 231–239.
- Stordal, A.S., Karlsen, H.A., Nævdal, G., Skaug, H.J. and Valles, B. (2011) Bridging the ensemble Kalman filter and particle filters: The adaptive Gaussian mixture filter. *Computational Geosciences*, 15, 293–305.
- Talagrand, O., Vautard, R. and Strauss, B. (1997). Evaluation of probabilistic prediction systems, *Proceedings, ECMWF Workshop on Predictability*, pp. 1–25: ECMWF.
- Van Leeuwen, P.J. (2009) Particle filtering in geophysical systems. *Monthly Weather Review*, 137, 4089–4114.
- Van Leeuwen, P.J. (2010) Nonlinear data assimilation in geosciences: An extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, 136(653), 1991–1999.
- Van Leeuwen, P.J., Künsch, H.R., Nerger, L., Potthast, R. and Reich, S. (2019) Particle filters for high-dimensional geoscience applications: A review. *Quarterly Journal of the Royal Meteorological Society*, 145(723), 2335–2365.
- Whitaker, J.S. and Hamill, T.M. (2002) Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130, 1913–1924.
- Whitaker, J.S. and Hamill, T.M. (2012) Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, 140, 3078–3089.
- Zhang, F., Snyder, C. and Sun, J. (2004) Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Monthly Weather Review*, 132, 1238–1253.
- Zupanski, M. (2005) Maximum likelihood ensemble filter: Theoretical aspects. *Monthly Weather Review*, 133, 1710–1726.

How to cite this article: Poterjoy, J. (2022) Regularization and tempering for a moment-matching localized particle filter. *Quarterly Journal of the Royal Meteorological Society*, 148(747), 2631–2651. Available from: <https://doi.org/10.1002/qj.4328>

APPENDIX A

Calculations of the localized posterior density are not tractable for large N_x , but further insight into the role of localization for higher dimensions can be obtained by considering the case where the variables in \mathbf{x} are conditionally independent. This assumption is not as restrictive as those made by filters that sample independently from posterior marginals, but it is likely not appropriate for complicated posterior pdfs, such as those with multiple modes. For this exercise, we will extend Equation (5) for $N_y > 1$ and $N_x > 2$. It is sufficient to consider the case of $N_y = 1$ and $N_x = 3$, as the posterior pdf calculation for successive observations would acquire the weighted delta function approximation of the posterior marginals obtained from the previous observations in sequence.

First, let $\mathbf{x} = [s \ u \ v]^T$, where v is directly observed by y and s and u are unobserved state variables. Specify localization coefficients for random variable pairs $\{s, v\}$ and $\{u, v\}$ to be $\rho_{s,v}$ and $\rho_{u,v}$, respectively, and assume conditional independence between s and u given v . The posterior pdf for the three-variable problem can then be expressed as a product of three marginals:

$$\begin{aligned} p(\mathbf{x}|y) &= p(s, u, v|y), \\ &= p(s|u, v)p(u|v)p(v|y), \\ &= p(s|v)p(u|v)p(v|y). \end{aligned} \quad (\text{A1})$$

Note that the third line of Equation (A1) holds for the standard PF, because knowledge of how u is sampled provides no new information for s when provided with knowledge of how v is sampled. By adopting a PF approximation for $p(v|y)$ and using the strategy introduced in Section 2 for localizing the remaining conditional distributions in Equation (A1), we can write

$$\begin{aligned} p(\mathbf{x}|y) &\approx p_l(\mathbf{x}|y), \\ &= \sum_{l=1}^{N_e} \sum_{m=1}^{N_e} \sum_{n=1}^{N_e} w^n \left[\delta_{ln} \rho_{s,v} + \frac{1}{N_e} (1 - \rho_{s,v}) \right] \\ &\quad \times \left[\delta_{mn} \rho_{u,v} + \frac{1}{N_e} (1 - \rho_{u,v}) \right] \\ &\quad \times \delta(s - s^l) \delta(u - u^m) \delta(v - v^n). \end{aligned} \quad (\text{A2})$$

Equation (A2) can be integrated with respect to s and v or u and v to arrive at the same form of posterior marginal as in Equation (6) for each unobserved variable. Note that this form requires that localization coefficients only need to be specified between an observation-space variable v and unobserved model variables—as is the case for other ensemble filters that perform a joint observation–model space localization.

We can further show that $p(s|u)$ adopts the same formulation as other localized conditionals, but with $\rho_{s,u} = \rho_{s,v} \rho_{u,v}$. Therefore, dependence across variables is modulated by products of localization coefficients under assumptions of conditional independence. To begin, start with the localized PF approximation for $p(\mathbf{x})$:

$$\begin{aligned} p(\mathbf{x}) &\approx p(s|v)p(u|v)p(v), \\ &\approx \sum_{l=1}^{N_e} \sum_{m=1}^{N_e} \sum_{n=1}^{N_e} \frac{1}{N_e} \left[\delta_{ln} \rho_{s,v} + \frac{1}{N_e} (1 - \rho_{s,v}) \right] \\ &\quad \times \left[\delta_{mn} \rho_{u,v} + \frac{1}{N_e} (1 - \rho_{u,v}) \right] \\ &\quad \times \delta(s - s^l) \delta(u - u^m) \delta(v - v^n). \end{aligned} \quad (\text{A3})$$

We can then integrate Equation (A3) with respect to v to get the following expression for the localized joint pdf for s and u :

$$\begin{aligned} p_l(s, u) &= \int_{\mathbb{R}} \sum_{l=1}^{N_e} \sum_{m=1}^{N_e} \sum_{n=1}^{N_e} \frac{1}{N_e} \left[\delta_{ln} \rho_{s,v} + \frac{1}{N_e} (1 - \rho_{s,v}) \right] \\ &\quad \times \left[\delta_{mn} \rho_{u,v} + \frac{1}{N_e} (1 - \rho_{u,v}) \right] \\ &\quad \times \delta(s - s^l) \delta(u - u^m) \delta(v - v^n) dv, \\ &= \sum_{l=1}^{N_e} \sum_{m=1}^{N_e} \sum_{n=1}^{N_e} \frac{1}{N_e} \left[\delta_{ln} \rho_{s,v} + \frac{1}{N_e} (1 - \rho_{s,v}) \right] \\ &\quad \times \left[\delta_{mn} \rho_{u,v} + \frac{1}{N_e} (1 - \rho_{u,v}) \right] \delta(s - s^l) \delta(u - u^m), \\ &= \sum_{l=1}^{N_e} \sum_{m=1}^{N_e} \sum_{n=1}^{N_e} \left[\frac{1}{N_e} \delta_{ln} \delta_{mn} \rho_{s,v} \rho_{u,v} + \frac{1}{N_e^2} \delta_{ln} \rho_{s,v} (1 - \rho_{u,v}) \right. \\ &\quad \left. + \frac{1}{N_e^2} \delta_{mn} \rho_{u,v} (1 - \rho_{s,v}) + \frac{1}{N_e^3} (1 - \rho_{s,v}) (1 - \rho_{u,v}) \right] \\ &\quad \times \delta(s - s^l) \delta(u - u^m), \\ &= \sum_{l=1}^{N_e} \sum_{m=1}^{N_e} \left[\frac{1}{N_e} \delta_{lm} \rho_{s,v} \rho_{u,v} + \frac{1}{N_e} \rho_{s,v} (1 - \rho_{u,v}) \right. \\ &\quad \left. + \frac{1}{N_e} \rho_{u,v} (1 - \rho_{s,v}) + \frac{1}{N_e^2} (1 - \rho_{s,v}) (1 - \rho_{u,v}) \right] \\ &\quad \times \delta(s - s^l) \delta(u - u^m), \\ &= \sum_{l=1}^{N_e} \sum_{m=1}^{N_e} \frac{1}{N_e} \left[\delta_{lm} \rho_{s,v} \rho_{u,v} + \frac{1}{N_e} (1 - \rho_{s,v} \rho_{u,v}) \right] \\ &\quad \times \delta(s - s^l) \delta(u - u^m). \end{aligned} \quad (\text{A4})$$

These steps result in the same equation that would be obtained if one were to modulate the conditional density $p(s|u)$ using a localization coefficient of $\rho_{s,u} = \rho_{s,v} \rho_{u,v}$. This exercise has practical importance, as it can be used to compute estimates of the covariance matrix for the localized posterior, which is needed for some filters (e.g., Morzfeld *et al.*, 2018).

APPENDIX B

For large applications, where calculations of $\{\omega_j^1, \dots, \omega_j^{N_e}\}$ may require $O(10^3)$ – $O(10^5)$ products of small numbers, special care is needed to avoid underflow errors. In practice, the log form described in previous sections for V_{ij}^n should be used for weights. We also recommend pre-multiplying $\ln(\hat{\omega}_j^n)$ by a first-guess β_j before transforming back through an exponential function to estimate a final β_j . Following these steps, calculations of ω_j^n require evaluating $\sum_{i=1}^{N_y} \ln(\hat{\omega}_{i,j}^n)$ at each grid point.

To reduce computation cost, we recommend multiplying $\hat{\omega}_{i,j}^n$ by N_e and applying a first-order Taylor series approximation about $N_e \hat{\omega}_{i,j}^n - 1$:

$$\begin{aligned} \ln(N_e \hat{\omega}_{i,j}^n) &= \ln[(N_e w_i^n - 1)\rho_{i,j} + 1], \\ &\approx (N_e w_i^n - 1)\rho_{i,j}, \end{aligned} \tag{B1}$$

which is sufficiently accurate when $|(N_e w_i^n - 1)\rho_{i,j}| \leq 0.3$. This approximation greatly reduces the number of log calculations needed for computing ω_j^n , thus reducing the cost of regularization and the subsequent strategies that rely on this method.

APPENDIX C

The local PF adopts an additional mixing strategy in Equation (7) for increasing particle diversity between iterations. This approach operates by multiplying \mathbf{r}_1 by a user-specified parameter γ , where $0 \leq \gamma \leq 1$, and re-calculating \mathbf{r}_2 to fit the posterior variance for each variable. Ignoring indices for state variables and considering a single marginal variable x , we start by setting

$$\sigma^2 = \frac{1}{N_e - 1} \sum_{n=1}^{N_e} (\gamma r_1 x'^{k_n} + r_2 x'^n)^2, \tag{C1}$$

where prime terms indicate the posterior mean has been removed and σ^2 is the posterior variance estimated from importance weights. One solution for r_2 that satisfies this equation is the positive root of

$$t_3 r_2^2 + t_2 r_2 + t_1 = 0, \tag{C2}$$

where

$$t_1 = (\gamma r_1)^2 \sum_{n=1}^{N_e} (x'^{k_n})^2 - \sigma^2 (N_e - 1), \tag{C3}$$

$$t_2 = 2\gamma r_1 \sum_{n=1}^{N_e} x'^{k_n} x'^n, \tag{C4}$$

$$t_3 = \sum_{n=1}^{N_e} (x'^n)^2. \tag{C5}$$

The r_2 found from this equation is used in the particle update step and the posterior particles are recentering on the posterior mean estimated from importance weights.