

# **Implications of non-Gaussian data assimilation for probabilistic prediction using ensembles**

Benjamin H. Sheppard<sup>a</sup> and Jonathan Poterjoy<sup>a</sup>

<sup>a</sup> *University of Maryland, College Park, College Park, Maryland*

*Corresponding author:* Benjamin H. Sheppard, [benshep@umd.edu](mailto:benshep@umd.edu)

## 1. Introduction

The chaotic nature of Earth's atmosphere merits the use of ensemble prediction systems, such as NOAA's Global Ensemble Forecast System (GEFS), to predict the weather. The purpose of data assimilation (DA) is to use observations to inform the correction of NWP model states, thus creating "analysis" states to serve as initial conditions for generating predictions. In Bayesian statistics, the NWP ensemble members being updated represent independent samples from a "prior" distribution and the analysis ensemble members (end product of DA) represent samples from a "posterior" distribution. Measured quantities such as radiances from a satellite, reflectivity from a radar, or temperature from a thermometer are examples of "observations" that DA uses when correcting model states, which requires a careful consideration of uncertainty as well. DA takes uncertainties in measurements into account for the "likelihood" portion of Bayes and tries to create a mapping between prior members and posterior members based on assumptions that go into the formulation of the prior and likelihood.

For multiple decades, ensemble Kalman filters (EnKFs; Evensen 1994) have been the preferred ensemble DA method, and they remain widely used in NWP systems to this day (Bannister 2016). Though there are many versions of the EnKF (Tippett et al. 2003), the common factor is the assumption that the prior probability density function (PDF) and likelihood are Gaussian; the former assumption dictates that the relationships between state variables must be linear. These assumptions hold true under many atmospheric scenarios, particularly at the synoptic scale, but the nonlinearities evident in mesoscale and sub-mesoscale dynamics present non-Gaussian PDFs that the EnKF cannot adequately represent (Poterjoy 2022a). Nevertheless, Gaussian-based ensemble filters have proven to be skillful for high-dimensional applications, despite the use of rather small ensemble sizes compared to the state dimension; see Houtekamer and Zhang (2016) for a review.

As an alternative to EnKFs, the localized particle filter (LPF) is an ensemble DA method that represents the prior as a sum of delta functions (Poterjoy 2016, 2022a; Poterjoy et al. 2019). It makes no assumptions regarding the prior error distribution, so it is well suited to handle nonlinear relationships needed for applications such as radar reflectivity and precipitation assimilation; see Poterjoy and Anderson (2016) and Poterjoy et al. (2017) for additional examples.

With computational expense being one of the biggest drawbacks of ensemble prediction systems, operational NWP models must rely on a limited number of ensemble members [~100 members] to represent prior and posterior statistics, making it impossible to properly sample all sources of error. Failure to account for the full range of errors causes the ensemble variance to collapse. The resulting over-confidence in the ensemble will cause observational data to be ignored in subsequent assimilation cycles, creating a runaway effect known as “filter divergence.” To prevent filter divergence, virtually all ensemble DA systems employ covariance inflation and localization (e.g., Whitaker and Hamill 2012). These strategies generally do not preserve the convergence properties of Monte Carlo DA techniques, so their effects on the posterior estimate must be studied.

Inflation seeks to prevent filter divergence by artificially increasing the variance of the prior or posterior ensemble. Multiplicative inflation works by multiplying the prior or posterior ensemble perturbations by a given factor (Anderson and Anderson 1999). Additive inflation adds a “perturbation” to each ensemble member, and these perturbations are drawn from a prescribed type of distribution with a mean of zero (Mitchell and Houtekamer 2000). The ideal amount of multiplicative or additive inflation can be approximated via experimentation, but once the parameters are set, the amount of dispersion provided by the inflation is usually fixed for each DA cycle. In NWP applications, different DA cycles may experience vastly different covariance (e.g., Poterjoy and Zhang 2011), thus, a more adaptive approach for inflation is ideal (Anderson 2008). Relaxation-based inflation uses statistics on the prior ensemble to inform the amount of posterior inflation. “Relaxation to prior spread” (RTPS; Whitaker and Hamill 2012) inflates the posterior to match the spread (standard deviation) of the prior; it is also the method used operationally by all NOAA weather models. “Relaxation to prior perturbations” (RTPP; Zhang et al. 2004) inflates the posterior to match the “prior perturbations,” i.e. the difference between individual prior members and their collective mean. RTPP has been hypothesized to improve dynamical balance during inflation but is not widely used for NWP.

Localization seeks to reduce spurious error dependence between points near and far from observations. For the EnKF, this involves tapering ensemble-estimated covariances to zero at a certain distance from observations, which we call a radius of influence (Roi). Localization is used in conjunction with the EnKF for NOAA weather models. For the LPF, localization acts in a similar manner, except it appears in the calculation of importance weights rather

than ensemble covariance. Because the LPF uses relaxed assumptions for the shape of error distributions while relying on a similar localization strategy as EnKFs, it also provides a means of clearly distinguishing the effects of Gaussian assumptions from heuristic measures typically used to cope with sampling errors, which is the primary topic of the current study.

NOAA's next-generation NWP system for hurricanes, the Hurricane Analysis and Forecasting System (HAFS), is set to become operational in summer 2023. The present study examines a series of HAFS experiments, simulating a portion of August 2020 that captures the life cycles of hurricanes Laura and Marco. Comparisons between the EnKF and LPF reveal stark qualitative differences in analyses, with the EnKF producing spuriously large energy at smaller scales than the LPF. This result is best characterized as a "spectral bias," which can be found by examining power spectral density of state variables following DA. Localization is often blamed for geophysical imbalances during EnKF updates, which can contribute to the observed spectral bias. Kepert (2009) showed that localization can disrupt geostrophic balance and cause "excessive levels of divergence." Greybush et al. (2011) demonstrated the same issue with 2 separate implementations of localization on the EnKF: tapering ensemble-estimated covariance to zero at a certain Roi and raising observation error covariance as a function of distance. In general, DA can produce unphysical model solutions whenever assumptions made to generate analyses are not satisfied. The most common problems stem from the application of localization and inflation steps to treat sampling error and (for the EnKF) the assumption that the prior PDF and likelihood are Gaussian. All three items require strict assumptions regarding error dependence across state variables and have the potential to cause dynamical imbalances that would produce the observed bias in HAFS analyses.

To isolate all possible causes of bias induced by ensemble DA, we perform numerical experiments with an idealized 1-dimensional model. We study the effect of Gaussian assumptions via comparing analyses from the EnKF and LPF. We also compute analyses with and without inflation and formulate experiments with the two EnKF relaxation-based inflation methods: RTPS and RTPP. The LPF relies on two alternative approaches to maintain spread in the ensemble, which we explore alongside EnKF experiments using posterior inflation. Lastly, we explore the implications of localization to induce spectral bias by adjusting the Roi parameter and examining its effects on the EnKF and LPF.

## 2. Findings from convective-permitting regional data assimilation experiments

### *a. The HAFS modeling system*

This study uses numerical experiments performed by Kurosawa and Poterjoy (In progress), which use the HAFS stand-alone regional model (HAFS-SAR) with the same configuration introduced by Knisely and Poterjoy (Under revision). The experiments use a domain centered at 25°N and 60°W spanning 109.5° of longitude and 68.2° of latitude, covering the entire North Atlantic hurricane basin. The model uses an approximate 6-km horizontal grid spacing with 81 vertical levels with the FV3 dynamical core. The model physics closely resemble the NCEP Global Forecast System (GFS) version 16, but with a refined boundary layer parameterization scheme that is catered toward tropical cyclones. The initial conditions used to initialize the first set of forecasts, as well as boundary conditions during the experiments, are provided by the 2020 operational GFS. Before studying the output of the HAFS experiments, the edge of the model domain is cropped to avoid contamination from said boundary conditions, and the first 7 days after the initialization of the experiments are discarded to reduce the dependency of model fields on the GFS. The experiments use hybrid DA, where an ensemble-3D-variational (En3DVar) method updates a deterministic state and an EnKF or LPF updates ensemble members. The ensemble is re-centered about the variational analysis each cycle, following the same strategy as the NOAA Global Data Assimilation System (GDAS). The exact effect of the re-centering is beyond the scope of this paper but will be explored in future research. Excluding the model spin up, the experiments run from 00 UTC August 18th to 00 UTC September 22nd, and measurements are assimilated every 6 hours. The experiments assimilate all the same measurements as the GFS, excluding proprietary measurements.

### *b. Results over multi-week HAFS experiments*

For the purpose of this study, we examine qualitative differences in analyses, short-term (6-h) forecasts, and 102-h ensemble forecasts generated over the course of these experiments. The relative forecast skill of HAFS under the DA configurations used for this study is summarized in Kurosawa and Poterjoy (In progress). In summary, Kurosawa and Poterjoy find LPF-Var to produce the most accurate forecasts, owing to the improved representation of

uncertainty provided by the LPF ensemble during DA. The present manuscript explores this finding in further detail.

To begin our analysis, we first examine results in the upper troposphere (250-mb and 300-mb) where infrequent measurements may lead to challenges for Gaussian-based DA methods. Comparing 300-mb absolute vorticity, we find that EnKF analyses are visibly “noisier” than the LPF equivalent (not shown), which matches findings from Poterjoy (2022b). Figure 1 shows this excess noise manifested in the elevated standard deviation of ensemble analyses (forecast hour zero) of the EnKF (solid red line), which decays as the forecast integrates into the future, resulting in a collapse of the ensemble spread for the first 30 forecast hours. Noting that the LPF does not experience a sudden collapse in spread, these experiments point to greater initial condition deficiencies in the EnKF analyses, which we suspect to be caused by either Gaussian approximations, inflation, or localization. We also note that combining the LPF with Var (denoted PF\_Var in Fig. 1) contributes to similar behavior in ensemble spread, but to a much smaller effect as the EnKF and EnKF-Var experiments.

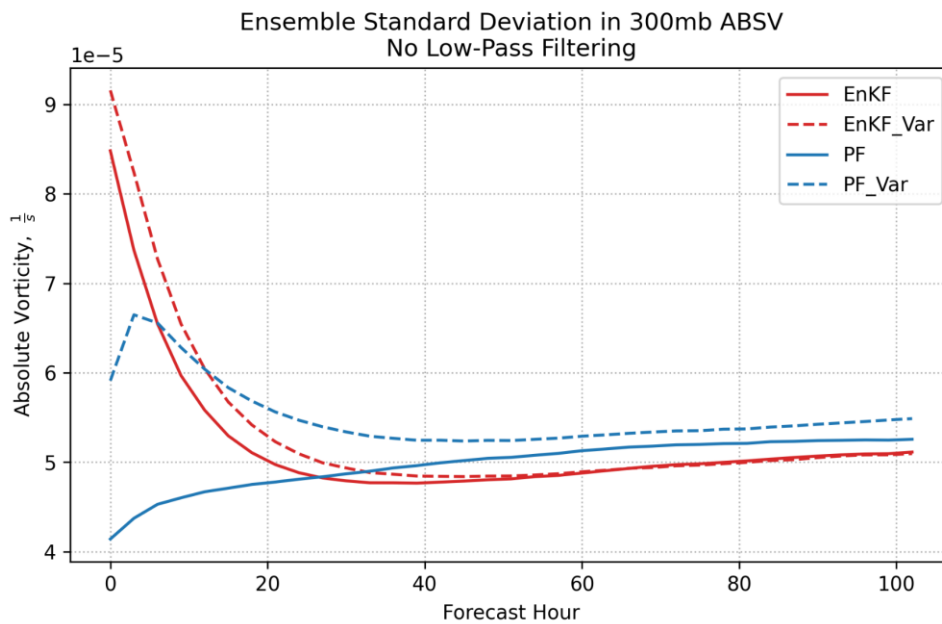


Fig. 1. HAFS ensemble standard deviation in 300mb absolute vorticity, plotted w.r.t. increasing forecast hour. The analysis frame appears as forecast hour zero. Solid lines denote pure EnKF and LPF, and the dashed lines represent variational solutions recentered off of the GFS. To smooth the results for this plot, the full-domain average of the standard deviation values has been taken (excluding the extreme edges which are contaminated by GFS boundary conditions), and that result was then averaged over multiple analysis cycles.

To investigate the findings in Fig. 1 further, we select the first ensemble member from LPF and EnKF experiments and compute meridional averages of zonal power spectral density of kinetic energy (KE). Results are averaged temporally over the entire multi-week HAFS experiments and plotted in Fig. 2 for (dark lines) analyses and (shaded) forecasts. Figure 2 clearly shows that the EnKF update (dark red, 0h) has significantly biased the KE spectrum away from the (dashed black line) standard FV3 climatology, especially at mesoscale wavelengths ( $< 400$  km). As the model advances forward in time, the wind field slowly adjusts so that the KE spectrum returns back to states that are supported by the FV3 model.

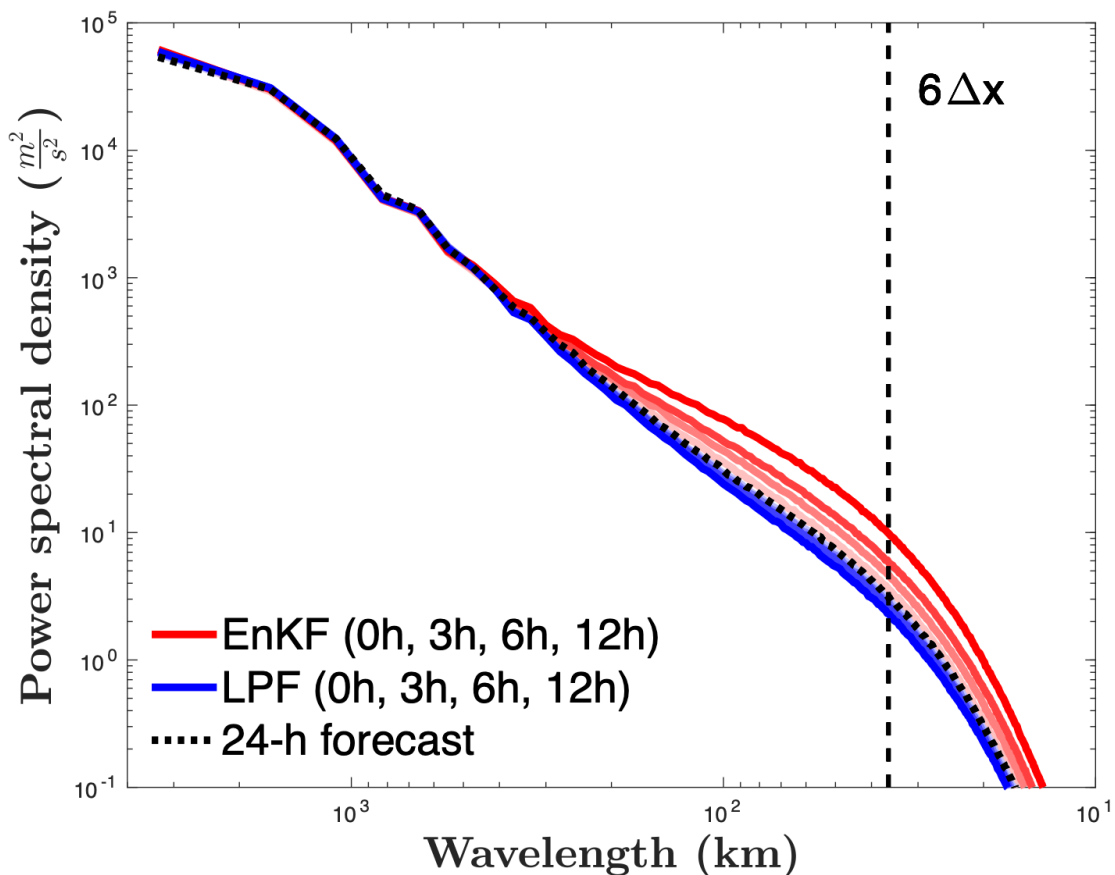


Fig. 2. 250-hPa kinetic energy spectrum, averaged from 2 weeks of single-member EnKF (red) and LPF (blue) forecasts at 0-h, 3-h, 6-h, and 12-h lead times. Shading indicates forecast lead time, with lighter shading corresponding to later lead times. The dotted black line shows a climatological estimate for the FV3 model using 24-h forecasts. The vertical dashed line corresponds to the length scale that is 6 times the grid spacing of the model.

### 3. Analysis from idealized experiments

To identify the source of spectral bias in ensemble analyses, we adopt an idealized model that allows us to separately study the effects of Gaussian assumptions, inflation, and localization through a rigorous exploration of the parameter space (i.e., coefficients controlling inflation and localization). The experiments examine multiple methods for preventing filter divergence, including different strategies for inflation, as well as different settings for the amount of inflation and localization radius of influence (Roi). Performing such a wide variety of configurations for the two DA methods on an operational NWP system like HAFS is too computationally prohibitive to ensure the robustness of findings. Hence, we resort to a one-dimensional (1-D) idealized model for this study.

*a. 1-D Solitary Wave Experiment Setup*

These experiments use a 1-D solitary wave model, which follows an exponential function with Gaussian perturbations to the position to form a prior ensemble. The same process is used to create a single “truth” dataset from which we can draw synthetic observations. The prior ensemble and observations are then used by the EnKF and LPF to compute analyses. Figure 3 illustrates an example of a 50-member prior ensemble and truth dataset generated with said strategy.

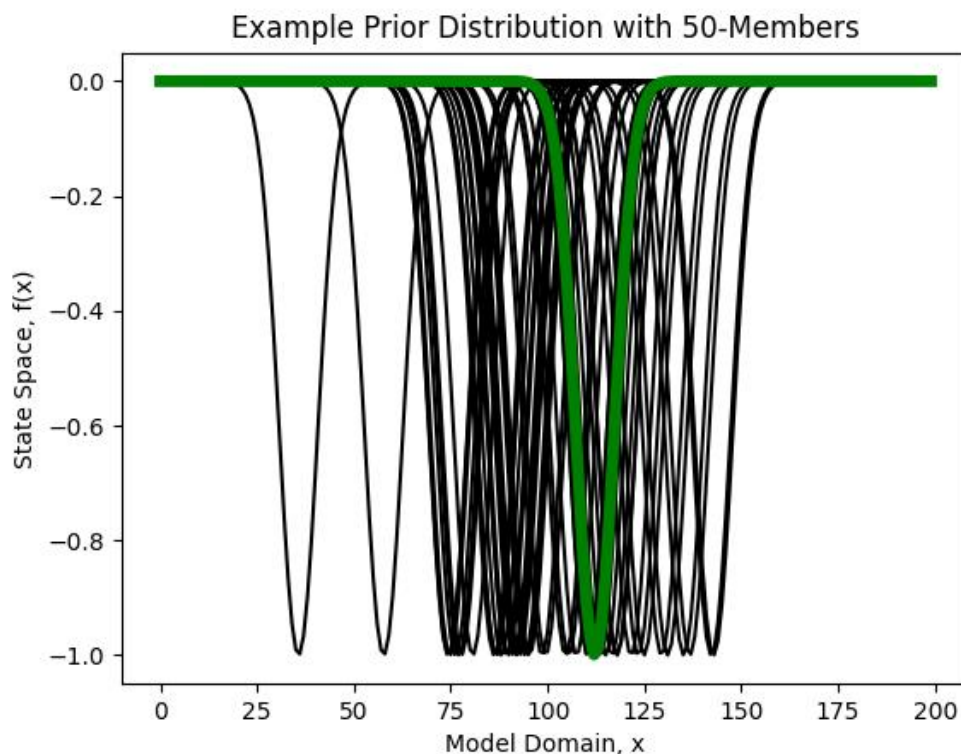




Fig. 3. Solid black lines illustrate a 50-member prior distribution initialized with equations (1) and (2), and an example “truth” dataset plotted in green. The truth is essentially just a random draw from the prior.

Though the wave-shaped exponential function depicted in Fig. 3 has no physical meaning, it is analogous to a cross section through a meteorological feature. For example, Fig. 3 resembles a vertical cross section of pressure taken across the core of a tropical cyclone or a cross section of potential vorticity through a tropopause fold (if the wave were inverted). The Gaussian perturbations to the location of the exponential function act to replicate position uncertainty in meteorological features, which are dominated by nonlinear advection terms in weather models. A common example is track uncertainty visualized in ensemble predictions for hurricanes. Despite the location perturbations being Gaussian, the joint distribution between different grid points on the model domain will be non-Gaussian, an effect that the EnKF will neglect when forming posterior members. Regardless of the shape for the prior distribution, the LPF should be able to correctly represent relationships between variables at different locations on the model domain when provided with a sufficiently large ensemble size—and when no inflation or localization are applied.

For these experiments, we model an exponential function of width  $L = 50$  grid points over a model domain vector,  $\vec{a}$ , comprised of  $N_x = 200$  equally-spaced grid points from 0 to 199 inclusive, with the wave centered approximately in the middle of the domain. We form prior ensembles,  $\vec{x}_i$  for  $i = 1$  to  $N_e$ , by including random perturbations to the center parameter used to model the exponential function:

$$c_x = \frac{N_x}{2} + \epsilon, \quad (1)$$

$$\vec{x}_i = - \exp \left[ \frac{-1}{L} (c_x + \epsilon - \vec{a})^2 \right], \quad (2)$$

where  $\epsilon$  is a random draw from a normal distribution,  $\mathcal{N}(0, \sigma_p^2)$ , with mean zero and standard deviation of  $\sigma_p = 10$ . From (1) – (2) we generate a 120-member prior ensemble, and a random sample designated to be the “truth” for each experiment. Synthetic observations are created over a region that spans the middle of the domain by sampling the true solution at an interval of 1 grid point and adding noise sampled from an assumed Gaussian observation

error distribution of  $\mathcal{N}(0, \sigma_o^2)$ , with a standard deviation of  $\sigma_o = 0.01$ . We then compute posterior members using the same ensemble filters used in HAFS, namely the ensemble square root filter of Whitaker and Hamill (2002) and the Poterjoy (2022a) iterative LPF. Each parameter setting is scrutinized using DA experiments repeated over 1000 trials with different random draws to explore the impacts of localization and inflation parameters on state estimates. We then analyze our results in physical space and spectral space to draw conclusions that are relevant to the findings from HAFS experiments.

### *b. EnKF Configuration*

For EnKF experiments that investigate the role of inflation during DA, we perform separate experiments using RTPS (Whitaker and Hamill 2012) to replicate the common strategy adopted for all NOAA NWP systems, and RTPP (Zhang et al. 2004) to explore an alternative inflation method. An inflation parameter, “gamma,” is used to adjust the strength of RTPS and RTPP inflation. For example, when gamma is set to 1.0, RTPS inflates the posterior to the full spread of the prior, and RTPP inflates the posterior to the perturbations of the prior. When gamma is set to 0.5, these methods only inflate half as much. Our experiments will use gamma values less than or equal to 1.0, as the posterior should be more confident than the prior for Gaussian errors. For experiments that investigate the role of localization, we adjust the Roi parameter. For the EnKF, localization tapers the ensemble-estimated covariances to zero at a certain number (Roi) of grid points away from the observation. Roi will be set to 1, 10, 100, and 1000 grid points to experiment with varying levels of localization, and we will set it to infinity to disable localization. The exact choice of these values was based on experimentation; it was found that Roi values exceeding 1000 grid points produced results nearly identical to Roi = infinity, while Roi = 1 represents the smallest-possible Roi for our model domain. The EnKF only reacts to localization when there is a linear relationship between a given observed and unobserved point on the model domain.

### *c. LPF Configuration*

The LPF for this experiment is based on Poterjoy (2022a), which has evolved from Poterjoy (2016) and Poterjoy and Anderson (2016) following numerous studies investigating non-Gaussian DA challenges for both idealized and high-dimensional models (Poterjoy et al.

2017, 2019). To prevent filter divergence, the LPF is equipped with mixing, localization, and regularization. Our results illustrate the effect of all three. The LPF uses a mixing parameter, “alpha,” instead of relaxation-based inflation, which seeks to increase diversity in posterior members in the presence of potentially non-Gaussian prior or posterior distributions—noting that traditional variance inflation methods used by EnKFs are only appropriate for Gaussian distributions. The alpha parameter “is a scalar between 0 and 1, which forces the LPF to update particles using combinations of the current particles and resampled particles everywhere in state space, including at the location of observations” (Poterjoy et al. 2019). This parameter directly modifies the equation that determines how to adjust state variables based on importance weights but, unlike inflation, it is formulated to preserve the posterior mean and variance that would occur without mixing. Setting alpha to 1.0 will disable mixing, and decreasing it towards zero will increase the mixing and disperse the posterior. We test the LPF using the same Roi parameter values as the EnKF, but note that localization has a different effect on the LPF. Rather than tapering ensemble-estimated covariances, the LPF simply adjusts the “importance weights” of the particles to diminish the DA update at a given distance from the observations, which can have a much different effect than localization applied for EnKFs when the prior is non-Gaussian. The LPF is also equipped with regularization, which dilutes the impact of observations via a “regularization” coefficient on particle weights to maintain a “threshold effective ensemble size” (Poterjoy 2022b). The effective ensemble size parameter is denoted “Neff.” Setting Neff equal to one disables regularization, since the actual effective ensemble size will never go below this value. The role of Neff in controlling spread in posterior members is complex, since this parameter is also used to determine how to perform iterations in the LPF as described in Poterjoy (2022b), and it only serves as a regularization term if the number of iterations reaches a user-specified maximum (set to 3 in HAFS experiments). For the idealized experiments, we cap the number of iterations to 1 and treat Neff exclusively as a term for controlling regularization.

## 4. Results

### *a. Single-trial experiments*

Figure 4 displays the results from a single-trial of the 1-D experiments. We can see that none of the prior members (Fig. 4a) go above zero in state space, yet some of the EnKF

posterior members (Fig. 4b) do. The prior is a set of single wave-shaped anomalies, but some of the EnKF posterior members have 2 or more waves embedded in them. Both the EnKF and LPF converge to the truth close to the 10 high-accuracy observations, but further away from said observations, we see significant deviations from the truth for EnKF.

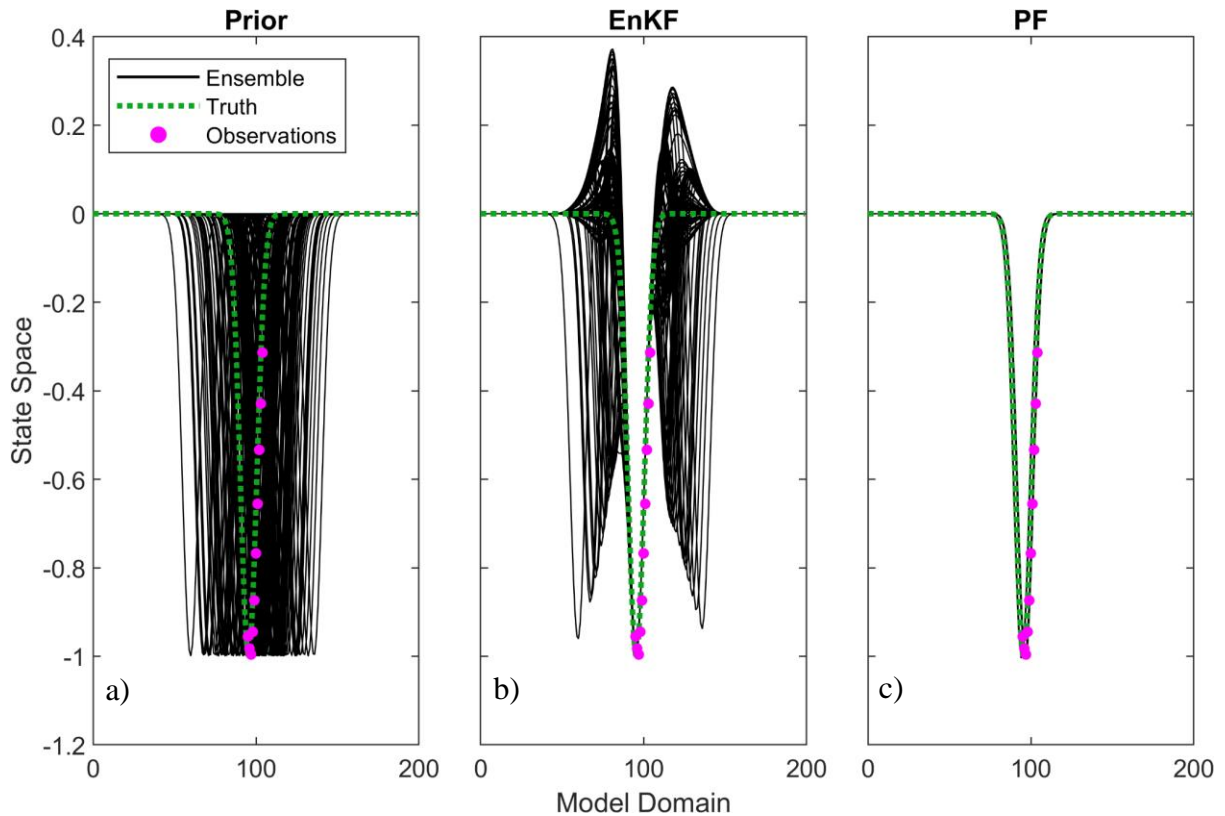


Fig. 4. Line plot of 120-member prior ensemble (left), EnKF posterior (center), and LPF posterior (right). EnKF and LPF assimilated the same prior and the same 10 synthetic observations (magenta dots). Observations were collected from the truth dataset (dashed green line) with a variance of 0.01 grid points. Neither inflation nor localization were implemented for this case study.

To further understand the behavior illustrated in the EnKF in Fig. 4, we examine the joint distribution between an observed point and an unobserved point on the model domain. For this purpose, we perform a single experiment using only 1 synthetic observation of the truth, taken at the 99th grid point on the model domain. This point will be denoted as “Point A” on the model domain, and the unobserved location 10 grid points to its left, the 89th grid point, will be denoted “Point B.” Figure 5 illustrates the setup for this experiment, and Figs. 6-8 show scatter plots for prior and posterior members at points A and B, which represent samples from the joint distribution characterizing this pair of random variables.

### Single-Trial, 1-Observation Study

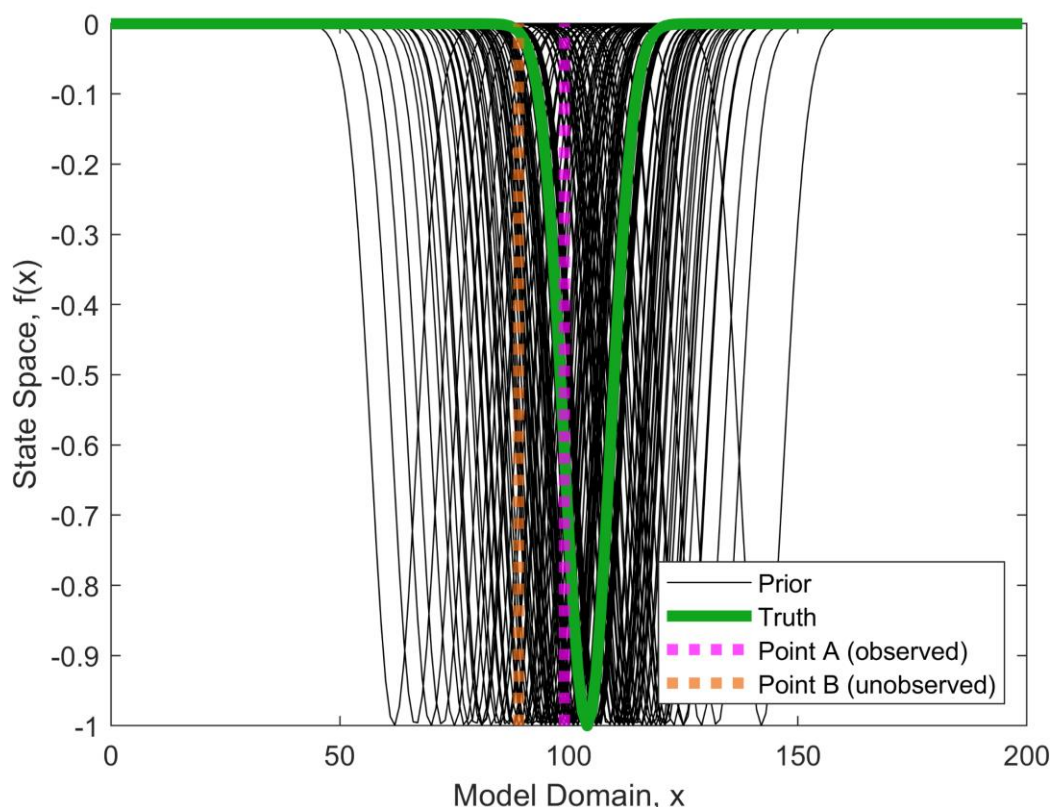


Fig. 5. 120-member prior ensemble (black) and truth dataset (green) with dashed lines denoting the two locations on the model domain used to generate the joint distribution figures (Figs. 6-8). A is the 99th grid point and B is the 89th grid point. Using the prior, the correlation coefficient between  $f(A)$  and  $f(B)$  is 0.0487.

### Joint Distribution of Model State showing EnKF Posterior Inflation

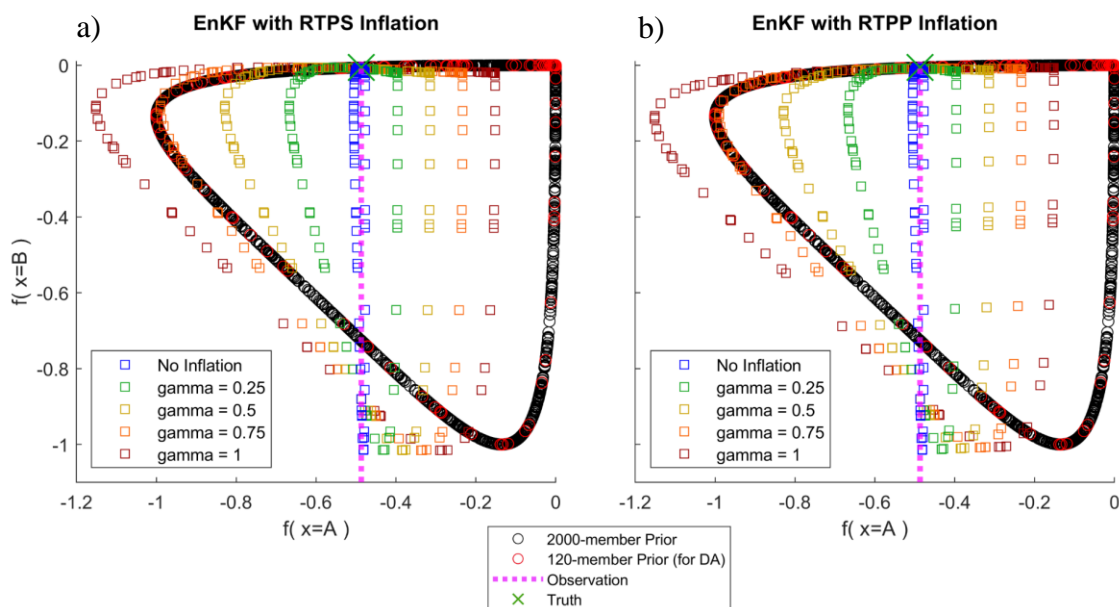


Fig. 6. Joint distribution between an observed point (A) and an unobserved point (B) showing the effect of RTPS (left) and RTPP (right) posterior inflation on the EnKF. Black circles represent a 2000-member example prior, which serves to delineate the manifold. The red circles are a 120-member prior, which along with a single observation at point A (dashed magenta line) were fed into the EnKF. Square markers represent the posterior. The blue squares were generated with no inflation, and subsequent colors denote increasing inflation.

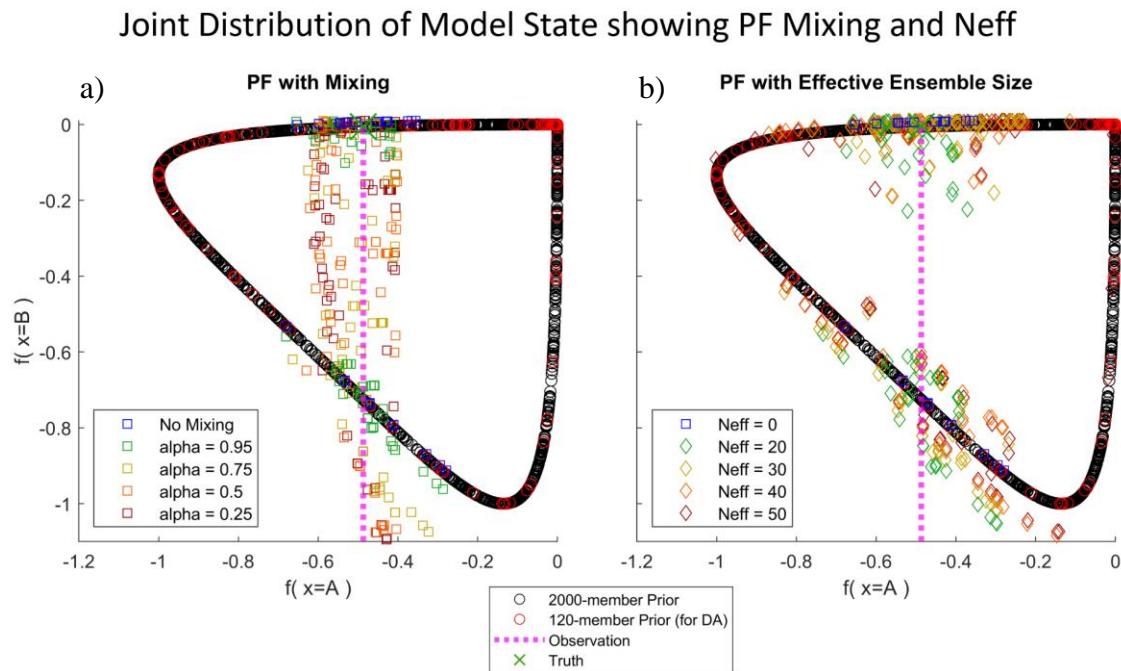


Fig. 7. Same experiment as Figs. 5 and 6, but showing the effect of the mixing parameter ( $\alpha$ , left plot) and effective ensemble size (Neff, right plot) on the LPF. Blue squares represent the raw LPF posterior. Note that setting  $\alpha$  to 1.0 disables mixing, and setting Neff to 0 disables regularization. Subsequent colors denote stronger mixing (left) and higher effective ensemble size (right). All experiments with  $\text{Neff} > 0$  used  $\alpha = 0.9$  for gentle mixing. The right plot uses diamond markers to improve the visibility of the posterior on the manifold.

## Joint Distribution of Model State showing Localization

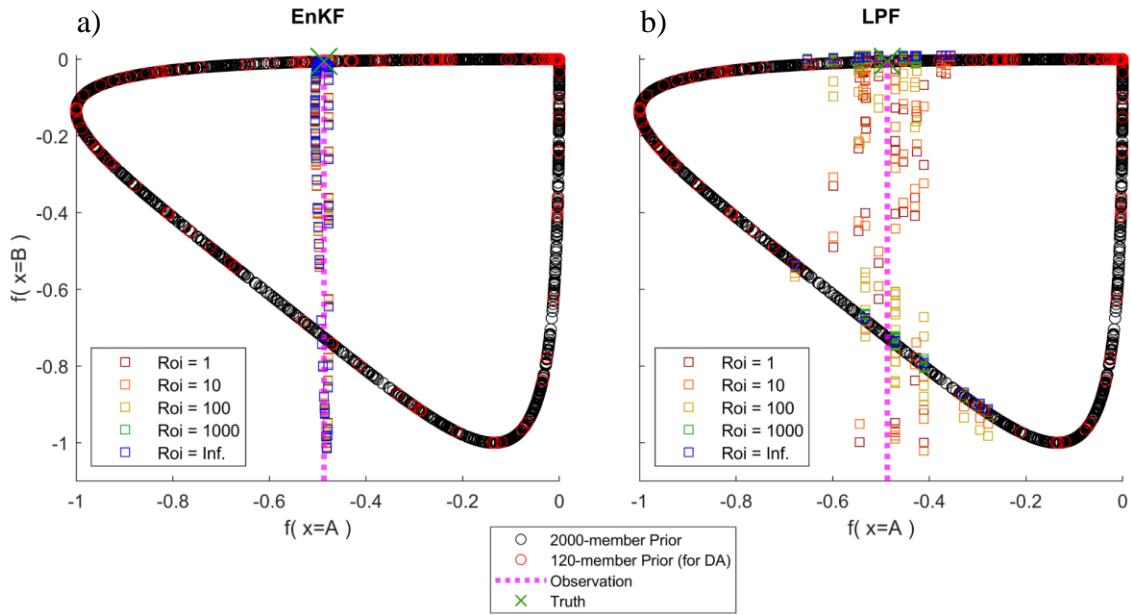


Fig. 8. Same experiment as Figs. 5 and 6, but showing the effect of localization radius of influence (Roi) on the EnKF (left) and LPF (right). Blue squares represent the DA output without localization (Roi = Infinity disables localization), while maroon squares represent the strongest possible localization.

The prior samples plotted in Fig. 6 show that there is a nonlinear relationship between the state-space value at Point A and Point B, and neither the joint distribution nor its marginals are Gaussian. The correlation coefficient between the two is only 0.0487. Thus, no significant linear relationship exists despite clear dependence across the pair of variables. With a large enough sample size, we can see a tightly constrained manifold with a quasi-triangular shape. It is well known that geophysical relationships like geostrophic and hydrostatic balance can create compact manifolds between state-space variables at different locations, though in high-dimensional hypercomplex systems, the manifolds would be noisier (more diffuse) than what we see with our 1-D exponential function. Nonetheless, our experiments reproduce a well-known result that the joint distribution between variables characterizing common weather variables (e.g., wind, moisture, and pressure) can be non-Gaussian when presented with displacement errors (Chen and Snyder 2007; Hodyss and Reinecke 2013; Poterjoy 2022b).

Referring to the blue squares in Fig. 6 (EnKF) and Fig. 7 (LPF), we can see the posterior results for the two DA methods without any inflation, mixing, or localization. The EnKF effectively updates the prior towards the observation through a step that resembles

multivariate linear regression, which would be a valid result for a Gaussian distribution. For the present application, however, the EnKF drives the posterior members off of the manifold, leading to solutions that are not supported by the 1-D model. The LPF avoids this issue entirely by selecting particles that more closely represent the observation. Thus, like the prior, the LPF posterior lands on the manifold. Applying the mixing step during LPF updates (Fig. 7a) also pushes posterior members off the manifold, but in a manner controlled by the user. This step has an effect that is comparable to using a Gaussian mixture approximation of the posterior density, rather than the typical delta function representation used by particle filters. The ability to shift members off the manifold has advantages for applications where model process error is large but unknown (Poterjoy 2022a), which is often the case for weather models. While holding  $\alpha$  at 0.9, raising the LPF's Neff parameter (Fig. 7b) also allows for greater diversity in posterior members, but with Neff set too high, useful info from the observation is ignored and the posterior will resemble the prior.

Covariance localization has minimal effect on the EnKF's posterior (Fig. 8a) given the lack of a linear relationship in the prior. The LPF is able to consider nonlinear relationships, thus it is sensitive to the Roi parameter (Fig. 8b) despite the small correlation coefficient. Tuning the Roi appropriately can result in greater performance from the LPF. In this case, the lowest RMSE is with localization disabled (Roi = infinity), but in geophysical models with a large state dimension and cycling DA, the LPF requires localization to remove erroneous dependence across state variables and prevent filter divergence. Moving "Point B" closer to "Point A" reveals stronger linear trends in the manifold. The correlation coefficient between A and B rises to 0.6313 when they are 5 grid points apart (Fig. 9a), and 0.9286 when they are 2 grid points apart (Fig. 9b), which then allows the EnKF to be more sensitive to localization.



### Joint Distribution of Model State (Point B closer to A)

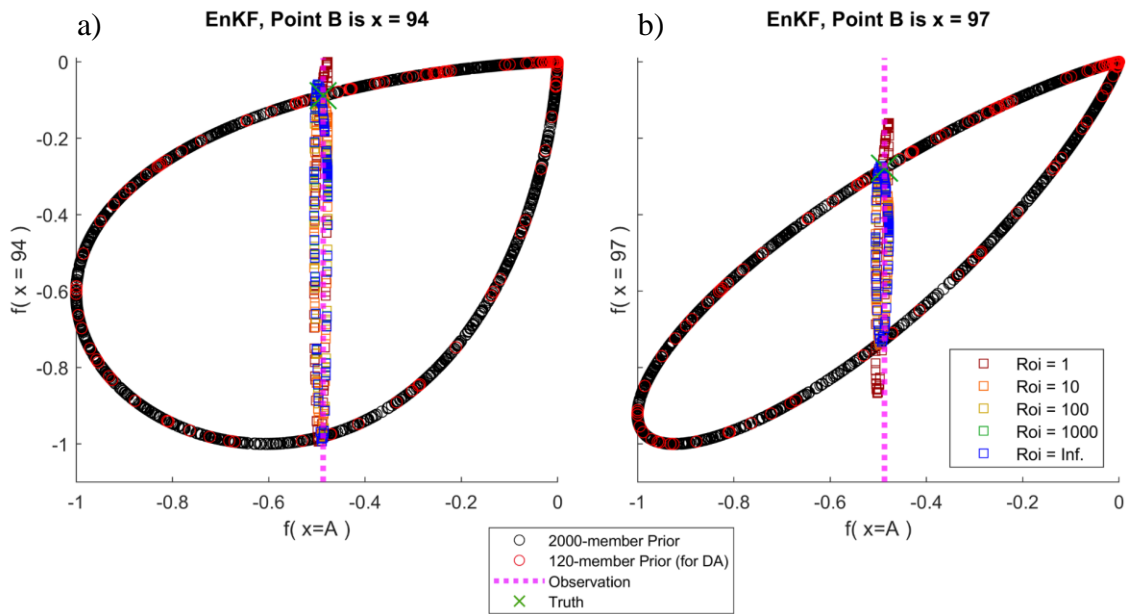


Fig. 9. Same experiment as Figs. 5 and 6, same EnKF output and localization color-coding as Fig. 8a, but with “Point B” moved closer and closer to “Point A” on the model domain. Previous plots used  $B = A$  minus 10 grid points such that the correlation between A and B is 0.0487. The left and right plots here use  $B = A$  minus 5 (correlation of 0.6313) and  $B = A$  minus 2 (correlation of 0.9286) respectively.

#### *b. Analysis of spectral characteristics of posterior members*

To assess which aspect of DA can induce the spectral bias observed in HAFS experiments, we investigate the role of Gaussian assumptions, localization, and inflation through numerous experiments with the 1-D solitary wave model. Each DA configuration is analyzed from posterior metrics accumulated over 1000 trials, which provides a means of quantifying systematic bias. Switching to spectral space (KE vs. wavelength) allows us to determine which aspects of the EnKF update produce noise at small length scales. For this experiment, each trial uses 10 observations of the truth with an observation error covariance of 0.01 (similar to Fig. 4). The LPF was run with a max of 3 iterations and with Neff set to 96 to maintain particle diversity in spite of the 10 high-accuracy observations.

Localization is often blamed for geophysical imbalances during EnKF updates (Kepert 2009; Greybush et al. 2011) so we will study it first. We disable EnKF inflation and PF mixing to isolate the effect of localization, then we apply the same Roi parameter settings

from the single-trial experiment: 1, 10, 100, and 1000 grid points. Figures 10c and 10d demonstrate that the EnKF with broad localization does not have a systematic bias in the KE spectrum, but it does have a large random deviation from the spectrum supported by the prior (red lines). This means that even without localization or inflation the EnKF can cause large deviations from the supported KE spectrum; it's just not consistently in one direction. The LPF with broad localization (Figs. 10g,h) continues to show a positive bias while the EnKF does not. Our HAFS experiments (Figs. 1 and 2) do not exhibit this behavior, thus, localization is likely not the primary source of spectral bias.

Next, we remove localization in the filter updates and examine the sensitivity of each DA method to inflation (using RTPS and RTPP) for the EnKF and mixing for the LPF. For this purpose, we perform trials with gamma values of {0.25, 0.50, 0.75, 1.0} and alpha values of {1.0, 0.75, 0.50, 0.25}. Figures 11a-d make it clear that a combination of EnKF for the choice of DA and RTPS inflation with a high enough gamma parameter will cause the posterior to experience an increase in KE at small wavelengths. This behavior closely matches our findings from the HAFS experiments, which perform EnKF DA with RTPS inflation. The LPF on the other hand (Figs. 11i-l) can perform DA updates without significantly modifying the KE spectrum from what it was in the prior, especially when mixing is minimized or disabled.

Furthermore, our results corroborate Whitaker and Hamill (2012) in that RTPP inflation (Figs. 11e-h) may be a less-optimal adaptive inflation technique for the EnKF compared to RTPS (Figs. 11a-d). Though RTPP fixes the positive KE bias at small wavelengths, the drift of the dashed “median” line off the prior suggests that it might be more disruptive to the KE spectrum than RTPS.

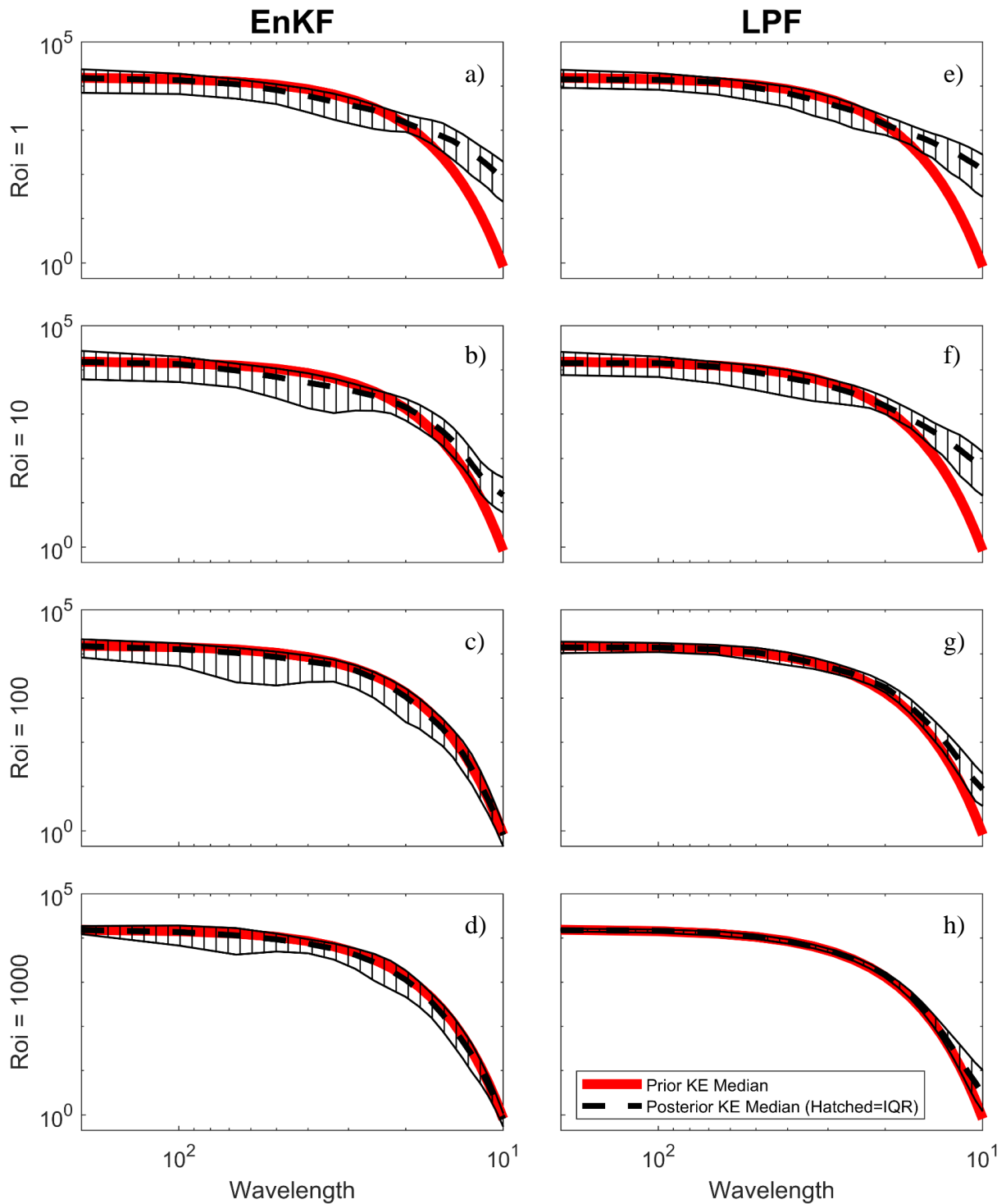


Fig. 10. Median (dashed black line) and IQR (hatched black) posterior kinetic energy spectra based on 1000 trials for the EnKF (left column) and LPF (right column) plotted over the prior KE median (red line). Each trial used a 120-member ensemble and 10 synthetic observations. From top to bottom, the rows illustrate the effect of increasing the localization radius of influence (Roi) labeled at left. Inflation was disabled for this study.

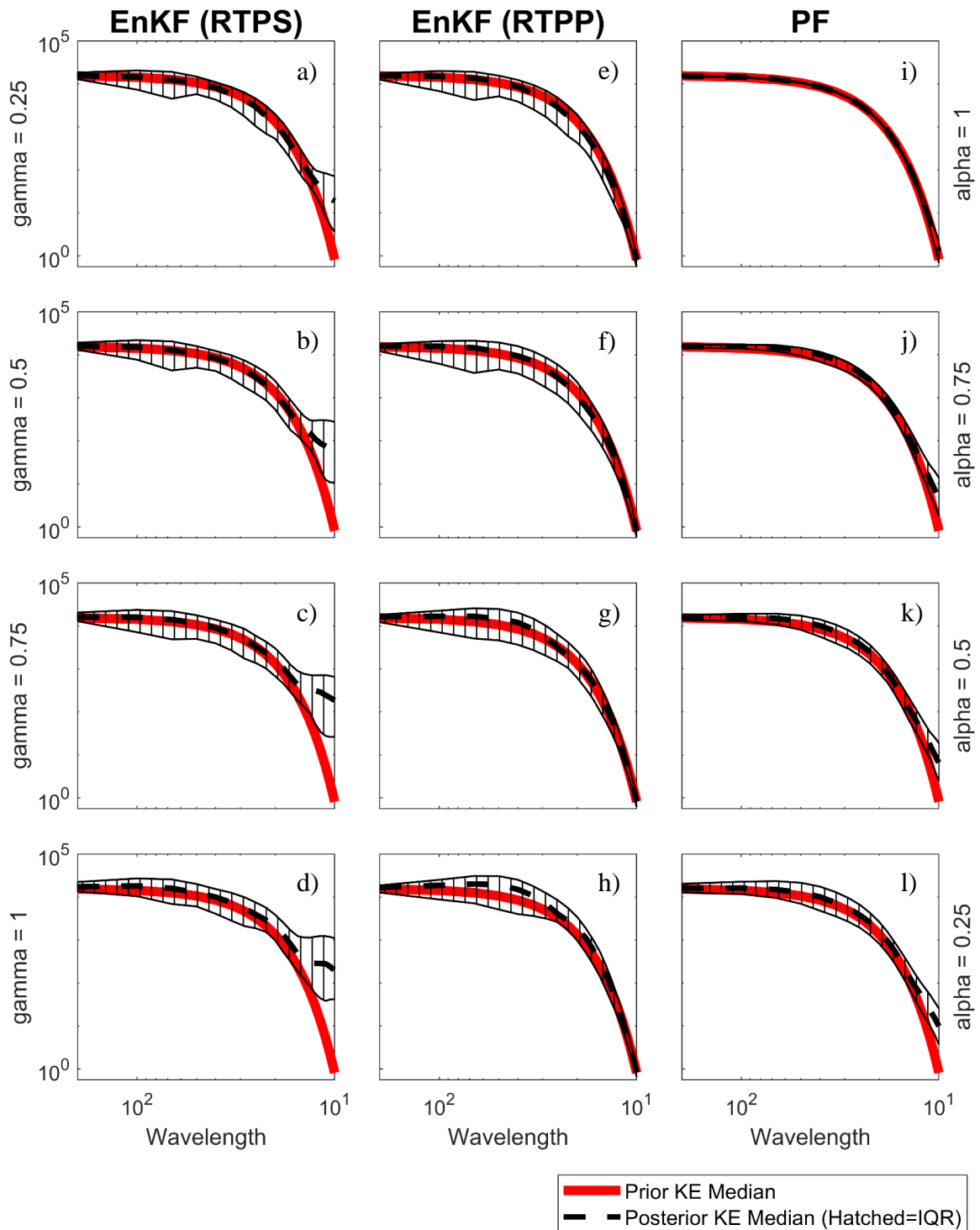


Fig. 11. Median (dashed black line) and IQR (hatched black) posterior kinetic energy spectra based on 1000 trials for the EnKF with RTPS inflation (left column), EnKF with RTPP inflation (center column), and LPF (right column), plotted over the prior KE median (red line). Each trial used a 120-member ensemble and 10 synthetic observations. From top to bottom, the rows illustrate the effect of increasing inflation. The EnKF columns utilize an inflation parameter ( $\gamma$ , row label at left), while the LPF utilizes a mixing parameter ( $\alpha$ , row label at right). Localization was disabled for this study.

## 5. Discussion and conclusions

This study identifies a spectral bias in EnKF analyses in an experimental weather prediction system (NOAA HAFS) and examines the contribution of localization, inflation, and the EnKF's Gaussian assumption to the bias. Multi-week HAFS experiments show that EnKF analyses contain excess KE at mesoscale and sub-mesoscale wavelengths, resulting in a high initial ensemble spread that collapses over the first 30 hours of the forecast. LPF analyses do not exhibit this behavior, which implies that the issue is related to DA, rather than the HAFS model or corrupted measurements feeding into the DA.

1-D exponential function experiments demonstrate that meteorological features with Gaussian position uncertainty have non-Gaussian relationships between different locations on the model domain. The EnKF neglects the non-Gaussian relationships and pulls ensemble members in unphysical directions, resulting in a posterior that disobeys geophysical balances. Previous studies blame localization for disrupting geophysical balances (Kepert 2009; Greybush et al. 2011). This study agrees that localization can disrupt balance (Fig. 10), but if this were to blame, the EnKF and LPF would exhibit a similar spectral bias, and the HAFS experiments disagree (Fig. 2). This study concludes that the EnKF's Gaussian assumption and RTPS inflation likely cause the majority of the spectral bias in DA applications for weather models, as shown in Figs. 11a-d. Given the widespread use of EnKFs and RTPS inflation in modern NWP, it is likely that such issues could be identified in analyses from other NWP systems, which highlights a major challenge for ensemble prediction systems that aim to accurately depict forecast uncertainty stemming from initial condition errors.

Replacing RTPS with RTPP in the EnKF can reduce spectral bias at short wavelengths, but we find it to introduce more significant biases at larger wavelengths (Fig. 11e-h), which is likely more problematic for ensemble forecasting. The elevated spectral bias at long wavelengths is consistent with findings reached by Whitaker and Hamill (2012) who noted that posterior spread induced by RTPP tends to target a small number of fast-growing modes of error as depicted by analysis-error covariance singular vectors. Their calculations were performed from DA experiments using a two-level primitive equation model for Earth's atmosphere, but reveal qualitatively similar behavior as the 1-D experiments performed in this study. The LPF, which makes no assumptions about the shape of the prior and posterior PDFs, avoids the issues discovered in RTPP and RTPS configurations of the EnKF, provided that the mixing parameter is minimal and Roi and Neff are tuned appropriately.

The effects of the EnKF's Gaussian assumption and RTPS inflation on the posterior are well documented in this paper, but future research will focus on the effect of said spectral bias on real-life NWP forecasts initialized from HAFS analyses. Analysis-error covariance singular vectors (Hamill et al. 2003) provide one candidate method for studying this behavior in HAFS experiments. The biased KE spectrum of the EnKF posterior members likely has significant impacts on HAFS forecast skill beyond the observed collapse in spread, and the results from this study are likely widely applicable to a variety of operational NWP systems.

## REFERENCES

- Anderson, J. L., 2008: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **61A**, 72-83, <https://doi.org/10.1111/j.1600-0870.2008.00361.x>.
- Anderson, J. L., and S. L. Anderson, 1999: A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts. *Mon. Wea. Rev.*, **127**, 2741-2758, [https://doi.org/10.1175/1520-0493\(1999\)127<2741:AMCIOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2).
- Bannister, R. N., 2016: A review of operational methods of variational and ensemble-variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **143**, 607-633, <https://doi.org/10.1002/qj.2982>.
- Chen, Y., and C. Snyder, 2007: Assimilating Vortex Position with an Ensemble Kalman Filter. *Mon. Wea. Rev.*, **135**, 1828-1845, <https://doi.org/10.1175/MWR3351.1>.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, **99**, 10143-10162, <https://doi.org/10.1029/94JC00572>.
- Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt, 2011: Balance and Ensemble Kalman Filter Localization Techniques. *Mon. Wea. Rev.*, **139**, 511-522, <https://doi.org/10.1175/2010MWR3328.1>.
- Hamill, T. M., C. Snyder, and J. S. Whitaker, 2003: Ensemble Forecasts and the Properties of Flow-Dependent Analysis-Error Covariance Singular Vectors. *Mon. Wea. Rev.*, **131**, 1741-1758, <https://doi.org/10.1175//2559.1>.

- Hodyss, D., and A. Reinecke, 2013: Skewness of the Prior Through Position Errors and Its Impact on Data Assimilation. *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II)*, S. K. Park, L. Xu, Ed., Springer Berlin Heidelberg, 147-175, [https://doi.org/10.1007/978-3-642-35088-7\\_7](https://doi.org/10.1007/978-3-642-35088-7_7).
- Houtekamer, P. L., and F. Zhang, 2016: Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation. *Mon. Wea. Rev.*, **144**, 4489-4532, <https://doi.org/10.1175/MWR-D-15-0440.1>.
- Keper, J. D., 2009: Covariance localisation and balance in an Ensemble Kalman Filter. *Quart. J. Roy. Meteor. Soc.*, **135**, 1157-1176, <https://doi.org/10.1002/qj.443>.
- Knisely, J., and J. Poterjoy: Implications of Self-Contained Radiance Bias Correction for Data Assimilation within the Hurricane Analysis and Forecasting System (HAFS). *Mon. Wea. Rev.*, under revision.
- Kurosawa, K., and J. Poterjoy, 2023: An augmented background error covariance approach to mitigate sampling errors: Experiments for a regional application of the NOAA Unified Forecast System, In progress.
- Mitchell, H. L., and P. L. Houtekamer, 2000: An Adaptive Ensemble Kalman Filter. *Mon. Wea. Rev.*, **128**, 416-433, [https://doi.org/10.1175/1520-0493\(2000\)128<0416:AAEKF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<0416:AAEKF>2.0.CO;2).
- Poterjoy, J., 2016: A Localized Particle Filter for High-Dimensional Nonlinear Systems. *Mon. Wea. Rev.*, **144**, 59-76, <https://doi.org/10.1175/MWR-D-15-0163.1>.
- Poterjoy, J., 2022a: Implications of Multivariate Non-Gaussian Data Assimilation for Multiscale Weather Prediction. *Mon. Wea. Rev.*, **150**, 1475-1493, <https://doi.org/10.1175/MWR-D-21-0228.1>.
- Poterjoy, J., 2022b: Regularization and tempering for a moment-matching localized particle filter. *Quart. J. Roy. Meteor. Soc.*, **148**, 2631-2651, <https://doi.org/10.1002/qj.4328>.
- Poterjoy, J., and F. Zhang, 2011: Dynamics and Structure of Forecast Error Covariance in the Core of a Developing Hurricane. *J. Atmos. Sci.*, **68**, 1586-1606, <https://doi.org/10.1175/2011JAS3681.1>.

- Poterjoy, J., and J. L. Anderson, 2016: Efficient Assimilation of Simulated Observations in a High-Dimensional Geophysical System Using a Localized Particle Filter. *Mon. Wea. Rev.*, **144**, 2007-2020, <https://doi.org/10.1175/MWR-D-15-0322.1>.
- Poterjoy, J., R. A. Sobash, and J. L. Anderson, 2017: Convective-Scale Data Assimilation for the Weather Research and Forecasting Model Using the Local Particle Filter. *Mon. Wea. Rev.*, **145**, 1897-1918, <https://doi.org/10.1175/MWR-D-16-0298.1>.
- Poterjoy, J., L. Wicker, M. Buehner, 2019: Progress toward the Application of a Localized Particle Filter for Numerical Weather Prediction. *Mon. Wea. Rev.*, **147**, 1107-1126, <https://doi.org/10.1175/MWR-D-17-0344.1>.
- Poterjoy, J., G. J. Alaka Jr., H. R. Winterbottom, 2021: The Irreplaceable Utility of Sequential Data Assimilation for Numerical Weather Prediction System Development: Lessons Learned from an Experimental HWRF System. *Wea. Forecasting*, **36**, 661-677, <https://doi.org/10.1175/WAF-D-20-0204.1>.
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, J. S. Whitaker, 2003: Ensemble Square Root Filters. *Mon. Wea. Rev.*, **131**, 1485-1490, [https://doi.org/10.1175/1520-0493\(2003\)131<1485:ESRF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2).
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble Data Assimilation without Perturbed Observations. *Mon. Wea. Rev.*, **130**, 1913-1924, [https://doi.org/10.1175/1520-0493\(2002\)130<1913:EDAWPO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2).
- Whitaker, J. S., and T. M. Hamill, 2012: Evaluating Methods to Account for System Errors in Ensemble Data Assimilation. *Mon. Wea. Rev.*, **140**, 3078-3089, <https://doi.org/10.1175/MWR-D-11-00276.1>.
- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of Initial Estimate and Observation Availability on Convective-Scale Data Assimilation with an Ensemble Kalman Filter. *Mon. Wea. Rev.*, **132**, 1238-1253, [https://doi.org/10.1175/1520-0493\(2004\)132<1238:IOIEAO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2).